Results of CoALL 07-03 study childhood ALL based on combined risk assessment by in vivo and in vitro pharmacosensitivity

Franziska Schramm,¹ Udo zur Stadt,¹ Martin Zimmermann,² Norbert Jorch,³ Arnulf Pekrun,⁴ Arndt Borkhardt,⁵ Thomas Imschiefer,⁶ Holger Christiansen,⁷ Jörg Faber,⁸ Irene Schmid,⁹ Tobias Feuchtinger,⁹ Gerhard Beron,¹⁰ Monique L. den Boer,¹¹ Rob Pieters,¹¹ Martin A. Horstmann,¹,¹² Gritta E. Janka-Schaub,¹ and Gabriele Escherich¹

¹Clinic of Pediatric Hematology and Oncology, University Medical Centre, Hamburg-Eppendorf Hamburg, Germany; ²Department of Pediatric Hematology and Oncology, Medical School Hannover, Hannover, Germany; ³Department of Pediatric Hematology and Oncology, Hospital Bielefeld, Bielefeld, Germany; ⁴Department of Pediatric Hematology and Oncology, Hospital Bremen-Mitte, Bremen, Germany; ⁵Department of Pediatric Oncology, Hematology and Clinical Immunology, Medical Faculty Duesseldorf, Duesseldorf, Germany; ⁶Department of Pediatric Hematology and Oncology, Helios Hospital, Krefeld, Germany; ⁷Clinic of Pediatric Hematology and Oncology, University Medical Centre Leipzig, Leipzig, Germany; ⁸Department of Pediatric Hematology/Oncology, University Hospital Mainz, Mainz, Germany; ⁹Dr. von Hauner Children’s Hospital, Ludwig Maximilian University, Munich, Germany; ¹⁰Department of Pediatric Hematology and Oncology, Helios Hospital, Wiesbaden, Germany; ¹¹Princess Maxima Center for Pediatric Oncology, Utrecht, The Netherlands; and ¹²Research Institute Children’s Cancer Centre, Hamburg, Germany

We conducted a clinical trial and report the long-term outcome of 773 children with acute lymphoblastic leukemia upon risk-adapted therapy accrued in trial CoALL 07-03 (from the Cooperative Study Group for Childhood Acute Lymphoblastic Leukemia). In a 2-step stratification, patients were allocated to receive either low- or high-risk treatment, based on initial white blood cell count, age, and immunophenotype. A second stratification was performed according to the results of in vitro pharmacosensitivity toward prednisolone, vincristine, and asparaginase (PVA score) and in vivo response after induction therapy (minimal residual disease [MRD]). Therapy was reduced for both risk groups in patients with a low PVA score or negative MRD result, and intensified in patients with a high PVA score. Overall outcome improved significantly compared with the predecessor CoALL 06-97 trial, with identical therapy backbone despite treatment reduction in 15.8% of patients (10-year probability of event-free survival, 83.5% vs 73.9%; overall survival, 90.7% vs 83.8%). Outcome for patients in the reduced treatment arms was superior to that of patients in the standard arms, associated with a profound reduction in frequency and severity of infectious complications. Importantly, we observed a lack of correlation between in vitro and in vivo drug response, as well as a lower predictive value of in vitro drug testing. Reflecting an intrinsic limitation of this methodology that prevents its use for treatment stratification in future trials. In conclusion, it might be possible to reduce chemotherapy in children with acute lymphoblastic leukemia selected by stringent in vivo measurement of MRD without jeopardizing overall outcome.

Introduction

Over the past decades, the prognosis for children with acute lymphoblastic leukemia (ALL) has improved substantially. Given steadily increasing survival rates, treatment-related side effects and mortality (TRM), as well as quality of life, of long-term survivors have moved into focus. The outcome for long-time survivors is often offset by the burden of toxic effects,¹ and further improvement will require a reduction in TRM.² Hence, a risk-adapted treatment approach has been firmly implemented in the Cooperative...
Study Group for Childhood Acute Lymphoblastic Leukemia (CoALL) trials, allocating patients to a low-risk (LR) or high-risk (HR) treatment arm with different treatment intensities, based on the traditional risk factors white blood cell count (WBC), age, immunophenotype, and chromosomal rearrangements. Most patients who had received intensive chemotherapy subsequently experienced at least one serious, sometimes life-threatening, infectious complication. Moreover, it has been shown that the risk of cardiovascular events is substantially increased in patients with severe infectious complications. Among the individual treat-ment elements of the CoALL backbone, infectious complications predominantly occur during or after completion of late intensification. Therefore, a de-escalation of late intensification might offer the opportunity to reduce both acute and long-term toxicities. To this end, LR and HR patients with any known risk factors and excellent in vitro/in vivo treatment response at the end of induction underwent a substantial treatment reduction in late intensification.

High-throughput in vitro drug testing has been proposed in “precision medicine” approaches based on comprehensive (epi-)genomic pro-filing, particularly in refractory or relapsed malignancies. To the best of our knowledge, a detailed comparison between both in vitro and in vivo methodologies assessing drug response has not been performed in a prospective clinical trial of uniformly treated patients. Here, we compared in vitro drug sensitivity and in vivo treatment response of ALL in more detail.

Methods

Patients

From October 2003 until September 2010, a total of 854 patients aged <18 years were enrolled in the multicenter trial CoALL 07-03, which recruited at 15 cooperating treatment centers in Germany (Figure 1A). The study was approved by the ethics committee of the city of Hamburg and the institutional review boards of individual trial centers. Written informed consent was obtained from parents or guardians in each case.

Overall, 773 patients with confirmed diagnosis of acute B- and T-precursor leukemia were eligible for this study, including 18 patients with Down syndrome. Eighty-one patients were excluded for different reasons (Figure 1A).

Assessment of in vitro pharmacosensitivity

At diagnosis, leukemia samples from bone marrow and/or peripheral blood were tested for in vitro cytotoxicity against PVA, based on the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide assay that has been described elsewhere in more detail. In brief, marked sensitivity toward a drug was scored as 1, intermediate sensitivity was scored as 2, and resistance was scored as 3. An individual PVA score, varying between 3 (sensitive) and 9 (resistant to all 3 drugs), resulted from the combination of the scores. For LR patients, the PVA score stratification in CoALL 06-97 seemed feasible, whereas results in HR patients were not satisfying. Hence, we decided not to continue a treatment reduction based on in vitro pharmacosensitivity in the HR group.

Analysis of minimal residual disease

Real-time quantitative polymerase chain reaction studies for minimal residual disease (MRD) analysis were performed with immunoglobulin heavy chain (IGH) and T-cell receptor (TCR) gene rearrange-ments as targets, and interpreted according to the guidelines developed within the European Study Group for MRD detection in ALL (ESG-MRD ALL). Ideally, 2 targets were used per patient, one with a quantitative range of $10^{-4}$ (0.01%) and the second with a quantitative range of at least $5 \times 10^{-4}$ (0.05%). A sample was evaluated as MRD negative if no positive signal was detected in any of three 500-ng patient DNA replicate samples for a specific target, with a quantitative range of 0.01%. The data of the prognostic relevance of the in vivo response after induction (MRD) were not available at the time the CoALL 07-03 trial was initiated but were implemented in 2006 when these results became available.

From 2006 onward, MRD measurement at day 15 and day 29 of induction was used for treatment stratification.

Stratification

All patients were stratified at diagnosis according to conventional risk criteria, allocating patients aged ≥10 years, T- or pro-B-cell immunophenotype (CD10 negative), or with a WBC ≥25/μL to the LR arm and all others to the LR arm. A second stratification was applied at the end of induction treatment (EOI), which included the state of cytomorphological remission and the results of the in vivo (MRD) and in vitro (PVA) pharmacological response, as well as molecular-/cytogenetics. Patients without remission at EOI, with a mixed-lineage-leukemia rearrangement or with a hypodiploid karyotype, were allocated to the HR arm. Three LR groups and two HR groups were defined based on the in vitro and, starting in 2006, in vivo drug response (MRD) (Figure 1A-B).

LR patients with a PVA score of 3 and 4 or negative MRD at days 15 and 29 were stratified to the LR reduced (LR-R) arm; those with a PVA score of 5 and 6 and any positive MRD result at day 15 and/or day 29 were stratified to the LR standard (LR-S) arm; and those with a PVA score of 7 to 9 were stratified to the LR intensified (LR-I) arm. In patients with discordant PVA and MRD results, the PVA score was decisive.

HR patients with T-precursor ALL and negative MRD results on days 15 and 29 and T-cell ALL (T-ALL patients with MRD days 15 and 29 < $10^{-3}$) were stratified to the HR reduced arm (HR-R). The remaining HR patients were allocated to the HR standard arm (HR-S). HR patients with a PVA score of 8 or 9 were eligible for stem cell transplantation (SCT) in their first complete remission (CR), provided that a fully matched donor was available. Patients whose PVA score and MRD result could not be determined received standard therapy.

Treatment

Details of treatment are summarized in Figure 2. There were no major modifications in the treatment backbone compared with the predecessor CoALL 06-97 trial, including the treatment reductions in patients with excellent in vitro response. However,
Figure 1. Consort and stratification diagrams of patients with ALL according to CoALL 07-03. (A) Overview of treatment and stratification of patients with ALL according to CoALL 07-03. (B) Stratification in CoALL 07-03 from 2006 onward. NHL, non-Hodgkin lymphoma; T-NHL, T-cell non-Hodgkin lymphoma; T-ALL, T-cell ALL; WBC, white blood cell count.
the cumulative amount of intrathecal methotrexate (IT-MTX) was reduced in CoALL 07-03 from 18 to 14 doses, and pegylated asparaginase replaced the native \textit{Escherichia coli}–derived asparaginase, not only in case of allergic reactions toward asparaginase but for all patients from the third dose onward. All patients received the same 3-drug induction (supplemental Figure 1), including a randomized prephase comparing the kinetics of blast counts after a single dose of daunorubicin vs doxorubicin.\textsuperscript{21} Consolidation and delayed intensification phases differed substantially according to risk stratification.

**LR patients**

All LR patients received identical consolidation therapy consisting of 4 therapy blocks (Figure 2A). The consolidation therapy was followed by 6-mercaptopurine (6-MP) (50 mg/m\(^2\)) daily for 4 weeks and 3 doses of IT-MTX and late intensification. For LR-R patients, doses of doxorubicin and dexamethasone were halved.

**HR and LR-I patients**

HR-S and LR-I patients received the same treatment (Figure 2B), including 6 blocks of consolidation therapy followed by daily 6-MP for 4 weeks and 3 doses of IT-MTX. In late intensification, doses of doxorubicin and dexamethasone were halved for the HR-R patients analogous to LR-R patients.

Patients who failed to achieve remission by day 29 (induction failure), and HR patients with a t(4;11) translocation or a PVA score

---

**Figure 2.** Treatment diagrams for all study patients of CoALL 07-03 treated according to the low- and high-risk arms. (A) Treatment overview for low-risk therapy in CoALL 07-03. (B) Treatment overview for high-risk therapy in CoALL 07-03. ADR, doxorubicin (30 mg/m\(^2\)); C1, methotrexate (1 g/m\(^2\))/\textit{Escherichia coli}–derived asparaginase (45 000 IE/m\(^2\))/6-MP (100 mg/m\(^2\) per day)/IT-MTX (age-dependent dose); C1a, methotrexate (1 g/m\(^2\))/pegylated asparaginase (2500 IE/m\(^2\))/6-MP (100 mg/m\(^2\) per day)/IT-MTX (age-dependent dose); C2, methotrexate (1 g/m\(^2\))/teniposide (VM26) (165 mg/m\(^2\))/cytarabine (300 mg/m\(^2\))/6-thioguanine (100 mg/m\(^2\) per day)/IT-MTX (age-dependent dose); C2a, methotrexate (1 g/m\(^2\))/teniposide (VM26) (165 mg/m\(^2\))/cytarabine (300 mg/m\(^2\))/6-MP (100 mg/m\(^2\) per day)/IT-MTX (age-dependent dose); C3, high-dose cytarabine (4 \times 3 g/m\(^2\))/pegylated asparaginase (2500 IE/m\(^2\))/C4, cyclophosphamide (900 mg/m\(^2\))/methotrexate (1 g/m\(^2\))/\textit{Escherichia coli}–derived asparaginase (45 000 IE/m\(^2\))/IT-MTX (age-dependent dose); C4a, cyclophosphamide (900 mg/m\(^2\))/methotrexate (1 g/m\(^2\))/\textit{Escherichia coli}–derived asparaginase (45 000 IE/m\(^2\))/6-MP (100 mg/m\(^2\) per day)/IT-MTX (age-dependent dose); Dex, dexamethasone (10 mg/m\(^2\) per day); I, induction; PEG-ASP, pegylated asparaginase (2500 IE/m\(^2\)) R, randomization; R1, cyclophosphamide (900 mg/m\(^2\))/cytarabine (4 \times 90 mg/m\(^2\))/6-thioguanine (100 mg/m\(^2\) per day)/IT-MTX (age-dependent dose); R1a, cyclophosphamide (900 mg/m\(^2\))/cytarabine (4 \times 90 mg/m\(^2\))/6-thioguanine (100 mg/m\(^2\) per day); VCR, vincristine (1.5 mg/m\(^2\)).
of 8 or 9, were eligible for SCT in 1. CR, provided that a fully matched donor was available.

**Central nervous system-directed therapy**

Patients with primary central nervous system (CNS) involvement (defined as nontraumatic lumbar puncture with leukocytes ≥5/μL and blasts in cerebrospinal fluid (CSF)) received twice-weekly IT-MTX (age-dependent dose) until blasts in the CSF were cleared and were eligible for therapeutic cranial radiotherapy, with an age-dependent dose (age 1-2 years, 12 Gy; age >2 years, 18 Gy). Patients with B-precursor ALL and an initial WBC count of >200/nL or 100 to 200/nL and >1 × 10^9/L leukemic blasts in the peripheral blood at day 8, and patients with T-ALL with a WBC count of ≥50/nL, were eligible for prophylactic cranial radiotherapy, with a total dose of 12 Gy. For all other patients, cranial radiotherapy was omitted. Patients with cranial radiotherapy received 8 IT doses of MTX until completion of the CNS phase, whereas patients without cranial irradiation received a total of 14 IT doses of MTX.

**Maintenance therapy**

Maintenance therapy was administered until 104 weeks after diagnosis and was identical for all patients. It comprised daily doses of oral 6-MP (50 mg/m^2) and weekly methotrexate (20 mg/m^2) adjusted to the WBC. Patients without cranial irradiation received three IT doses of MTX 3, 6, and 9 months after the start of maintenance therapy.
The probability of event-free survival (pEFS) and overall survival (OS) was estimated by using the Kaplan-Meier method, and compared between subgroups by using the log-rank test.22 pEFS was the time from diagnosis to the first event, including remission failure (presence of leukemic blasts ≥5% until day 56 of treatment), induction death, relapse, death by any cause in remission, second malignancy, or censoring at last follow-up. OS was defined as the time from diagnosis to death by any cause or censoring at last follow-up. Cumulative incidence functions of isolated CNS or any (isolated and combined) CNS relapse, as well as testicular relapse, treatment-related second malignancies, and toxicity-related death, were calculated by using the Kalbfleisch and Prentice method and compared by using Gray’s test.23,24 The χ² test, Fisher’s exact test, and Spearman’s rank correlation analyses were applied to compare distribution of parameters between subgroups and correlation between parameters. The status of patients was monitored annually on protocol-specific forms reviewed by the trial data center before data entry. A recent update of the database (1 June 2018) was used for this analysis.

### Statistical analysis

The probability of event-free survival (pEFS) and overall survival (OS) was estimated by using the Kaplan-Meier method, and compared between subgroups by using the log-rank test.22 pEFS was the time from diagnosis to the first event, including remission failure (presence of leukemic blasts ≥5% until day 56 of treatment), induction death, relapse, death by any cause in remission, second malignancy, or censoring at last follow-up. OS was defined as the time from diagnosis to death by any cause or censoring at last follow-up. Cumulative incidence functions of isolated CNS or any (isolated and combined) CNS relapse, as well as testicular relapse, treatment-related second malignancies, and toxicity-related death, were calculated by using the Kalbfleisch and Prentice method and compared by using Gray’s test.23,24 The χ² test, Fisher’s exact test, and Spearman’s rank correlation analyses were applied to compare distribution of parameters between subgroups and correlation between parameters. The status of patients was monitored annually on protocol-specific forms reviewed by the trial data center before data entry. A recent update of the database (1 June 2018) was used for this analysis.

### Toxicity reporting

Local trial centers conducted mandatory documentation of toxicities for each treatment block, based on the National Cancer Institute’s Common Toxicity Criteria, version 2.0, modified by the German Society of Pediatric Oncology and Hematology on a patient-specific therapy flowsheet and an extra documentation sheet for each single treatment element. In addition, treatment-related toxicities were captured by medical reports. Treatment-related toxicities were assigned to a specific treatment element, only if they had occurred during the application of chemotherapy or before the next treatment element was started.

### Results

**Patient characteristics and treatment results according to presenting features and risk groups**

In 773 evaluable patients, the 10-year pEFS was 83.5% (standard error [SE], 1.4%) and OS was 90.7% (SE, 1.1%), which were significantly improved compared with those of the predecessor trial CoALL 06-97, which had a pEFS of 73.9% (SE, 1.7%; P < .0001) and an OS of 83.8% (SE, 1.5%; P ≤ .0001) (Figure 3A-B). Ten patients (1.3%) died in CR, 7 patients died due to infectious complications, 1 patient died due to multiple organ failure caused by methotrexate toxicity, and 2 patients died at home with unexplained cause of death. All patients with lethal infectious complications were treated according to the HR protocol, except 1 LR-S patient. Those severe infections occurred throughout all treatment phases, with no obvious preference. No treatment-related death was
observed in those patients who underwent treatment de-escalation (Table 1). The 10-year cumulative incidence of relapses for all patients was 12% (SE, 1%), and the incidence of isolated CNS relapses was 1.4% (SE, 0%).

Table 2 presents an overview of patient characteristics, outcome data, and the univariate survival analysis in 773 patients. In all CoALL studies, age $\leq$ 10 years, initial WBC $\geq$ 25 $\times 10^9$/L, and pro-B and T-ALL immunophenotype were of significant prognostic relevance despite allocation to HR treatment. T-lineage ALL had a comparable 10-year pEFS but a significantly lower OS compared with B-lineage ALL (10-year pEFS, 80.8% vs 83.9% [P = .26]; 10-year OS, 84.1% vs 91.7% [P = .007]). CNS involvement was of significant prognostic impact, despite therapeutic cranial radiation (10-year pEFS 84.3% for patients without vs 60% for patients with CNS involvement [P = .001]; 10-year OS, 91.8% vs 59.2% [P = .000]).

In vitro pharmacosensitivity testing could be performed in 471 of 773 patients. In 302 patients, no PVA score was available, either because no initial material was available (n = 204 [26%]) or the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide; 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide assay was not evaluable (n = 98 [12.6%]). A total of 445 patients had available day 29 MRD results, and 404 patients had available day 15 and day 29 MRD results. In the whole cohort of 773 patients, 241 (31.1%) were stratified to the LR-S, 103 (13.3%) to the LR-R, 36 (4.7%) to the LR-I, 374 (48.4%) to the HR-S, and 19 (2.5%) patients to the HR-R treatment arm.

In 236 of 773 patients, mostly recruited before the introduction of MRD in 2006, only a PVA score was available for stratification: 43 (18.2%) were LR-S, 57 (24.2%) were LR-R, and 14 (5.9%) were LR-I. In the 121 (51.3%) HR patients, the PVA score was not used for stratification as indicated earlier; all patients were treated according to the HR-S arm.

Of the 445 patients with MRD results, 235 patients also had an available PVA score, both of which were used for stratification. Among those 445 patients, 152 (34.1%) were LR-S, 44 (10%) were LR-R, 22 (4.9%) were LR-I, 208 (47%) were HR-S, and 19 (4%) were HR-R.

Among the 44 patients stratified to the LR-R arm, 19 patients were MRD-negative on day 29 showing no relapse. However, among 25 MRD-positive but according to PVA score 3 and 4 LR-R stratified patients, 5 relapse events and 1 TRM were observed; this resulted in a statistically significant difference in pEFS (P = .025) (supplemental Figure 4A).

Of those 19 HR-R patients with negative MRD at EOI, 11 patients had a corresponding PVA score of 3 or 4. Two patients had a PVA

| Table 2. Treatment results of all CoALL 07-03 patients according to presenting features |
|-----------------------------------|----------|----------|---------|---------|---------|---------|---------|---------|
| | pEFS | Survival |
| | Year 5 | Year 10 | Year 5 | Year 10 |
| | Prob. | SE | Prob. | SE | Prob. | SE | Prob. | SE |
| Immunophenotype | | | | | | | | |
| Non-T NCI standard | 474 | 87.9 | 1.5 | 87.1 | 1.6 | .000 | 95.8 | 0.9 | 94.9 | 1.0 | .000 |
| Non-T NCI high | 189 | 77.2 | 3.1 | 75.9 | 3.1 | .650 | 85.8 | 4.0 | 84.4 | 4.1 | .990 |
| T-ALL NCI standard | 32 | 78.1 | 7.3 | 78.1 | 7.3 | .260 | 93.0 | 1.0 | 91.7 | 1.1 | .007 |
| T-ALL NCI high | 78 | 81.9 | 4.4 | 81.9 | 4.4 | .260 | 93.0 | 1.0 | 91.7 | 1.1 | .007 |
| Non-T | 663 | 84.8 | 1.4 | 83.9 | 1.4 | .260 | 93.0 | 1.0 | 91.7 | 1.1 | .007 |
| T-ALL | 110 | 80.8 | 3.8 | 80.8 | 3.8 | .260 | 93.0 | 1.0 | 91.7 | 1.1 | .007 |
| Sex | | | | | | | | | |
| Male | 429 | 82.4 | 1.8 | 81.7 | 1.9 | .100 | 91.6 | 1.3 | 89.7 | 1.5 | .290 |
| Female | 344 | 86.6 | 1.8 | 85.7 | 1.9 | .100 | 92.7 | 1.4 | 91.9 | 1.5 | .290 |
| Age at diagnosis | | | | | | | | | |
| Age $\leq$ 10 y | 601 | 85.6 | 1.4 | 85.0 | 1.5 | .021 | 93.8 | 1.0 | 92.7 | 1.1 | .000 |
| Age $\geq$ 10 y | 172 | 79.5 | 3.1 | 78.0 | 3.2 | .080 | 85.9 | 2.7 | 83.2 | 3.1 | .002 |
| WBC $\times 10^9$/L | | | | | | | | | |
| WBC $<25$ | 525 | 86.0 | 1.5 | 84.3 | 1.6 | .800 | 94.6 | 1.0 | 92.9 | 1.1 | .002 |
| WBC $\geq25$ | 248 | 80.6 | 2.5 | 80.0 | 2.6 | .800 | 86.7 | 2.2 | 86.2 | 2.2 | .000 |
| CNS status | | | | | | | | | |
| CNS negative | 743 | 85.1 | 1.3 | 84.3 | 1.4 | .001 | 93.1 | 0.9 | 91.8 | 1.0 | .000 |
| CNS positive | 20 | 60.0 | 11.0 | 60.0 | 11.0 | .001 | 64.6 | 10.8 | 59.2 | 11.1 | .000 |
| t(12;21) | | | | | | | | | |
| ETV6-RUNX1 negative | 517 | 80.0 | 1.8 | 78.8 | 1.8 | .000 | 89.7 | 1.3 | 87.8 | 1.5 | .000 |
| Positive | 170 | 92.4 | 2.0 | 92.4 | 2.0 | .000 | 98.2 | 1.0 | 97.6 | 1.2 | .000 |

NCI, National Cancer Institute; Prob., probability.
score of 5. In 6 patients, a PVA score was not available. By definition, HR-R patients were all MRD-negative.

The distribution of the in vitro pharmacosensitivity score differed according to immunophenotype and TEL/AML1 status. Patients with T-ALL were more likely to present with a resistant PVA score of 7 to 9 compared with B-precursor patients ($p < 0.0001$). TEL/AML1-rearranged ALL exhibited an association with a sensitive PVA score of 3 + 4 ($p < 0.001$) (supplemental Table 5).

Long-term treatment results after 5 and 10 years for treatment-reduced patients with excellent in vitro or in vivo drug response exhibited a trend toward improved outcome compared with patients in standard arms (Table 2; Figure 4). With a median follow-up time of 120 months (range, 9-178 months), LR-R patients achieved a 10-year pEFS of 88.9% (SE, 3.2%; 5-year pEFS, 91.1% [SE, 2.8%]) and a 10-year OS of 96.0% (SE, 1.9%; 5-year OS, 98.0% [SE, 1.4%]). For HR-R patients, the 10-year pEFS was 87.4% (SE, 8.4%; 5-year pEFS, 94.7% [SE, 5.1%]), with a 10-year OS of 91.7% (SE, 8.0%; 5-year OS, 100% [SE, 0.0%]).

Furthermore, the results in the reduced treatment arms showed a trend toward improved outcome, compared with the preceding CoALL 06-97 study. LR-R patients in CoALL 06-97 had a 10-year pEFS of 86.9% (SE, 3.5%; $p = .63$) with a 10-year OS of 94.5% (SE, 2.4%; $p = .61$), whereas HR-R patients reported a 10-year pEFS of 69.4% (SE, 5.7%; $p = .088$) and a 10-year OS of 79.5% (SE, 4.9%; $p = .12$).

Regarding the standard risk groups, the results in LR-S as well as HR-S were significantly better in CoALL 07-03 compared with CoALL 06-97. LR-S patients in CoALL 06-97 achieved a 10-year pEFS of 79.5% (SE, 3.1%; $p = .052$) and a 10-year OS of 91.2% (SE, 2.2%; $p = .081$). HR-S patients in CoALL 06-97 exhibited a 10-year pEFS of 65.9% (SE, 2.9%; $p = .00030$) and a 10-year OS of 74.6% (SE, 2.7%; $p = .00090$).

Although treatment intensity for the HR patients was much higher, pEFS and OS were lower compared with the LR group, except for the LR-I patients. This patient group was stratified to LR by using traditional criteria, but according to the second stratum based on PVA score and MRD, patients were assigned to LR-I receiving HR standard treatment. LR-I patients had a 10-year pEFS of 83% (SE, 6.2%) and a 10-year OS of 94.4% (SE, 3.8%) (Table 2; Figure 4). In the LR-I arm, only 1 of 22 patients was stratified without an available PVA score because of MRD EOI $>10^{-3}$. All other patients ($n = 21$)
had a PVA score of 7 to 9 with sometimes nonconcurrent MRD results

Induction failure was diagnosed in 50 patients, 6 of whom presented with persisting leukemic blasts at day 56 (remission failure). Three TRM, 1 second malignant neoplasm (SMN), and 13 relapses were registered, resulting in an OS and pEFS of 72.2% (SE, 6.8%) and 63.0% (SE, 7.4%), respectively.

Among the 24 HR patients with a PVA score of 8 and 9, no patient received a SCT in 1. CR from a matched sibling donor solely based on the PVA score. Six patients received a matched unrelated donor SCT because of MRD persistence, and the remaining patients were treated according to the HR-S protocol. No TRM occurred, but 1 relapse occurred in these 8 patients after SCT. In the whole cohort of 24 HR patients with PVA 8 and 9, four relapses and a single SMN were observed.

**Lack of correlation between in vitro PVA score and in vivo assessment of MRD**

No correlations were observed in either the whole cohort (n = 235) or in the B-precursor ALL patient subcohort (n = 200) with available PVA scores and MRD results between the in vitro pharmacosensitivity results and in vivo response after induction therapy (Spearman’s rank correlation).

**Figure 6. Comparisons of cumulative incidence probabilities at 10 years.** (A) Probability of the cumulative incidence of relapse at 10 years for CoALL 07-03 patients according to MRD at the end of induction. (B) Probability of the cumulative incidence of relapse at 10 years for CoALL 07-03 patients according to the PVA score.
correlation coefficient, 0.24 for the whole cohort and 0.15 for the B-precursor ALL patients, respectively) (Figure 5A-B). Thus, the MRD level at EOI treatment could not be predicted by initial in vitro drug testing before the start of chemotherapy. A rather weak correlation (Spearman’s rank correlation coefficient, 0.64) (Figure 5C) was found only in patients with T-ALL (n = 35). MRD negativity at the end of induction was associated with an excellent EFS, irrespective of the PVA score indicative of a greater predictive value of MRD compared with in vitro drug testing (Figure 6A-B; supplemental Figures 4B and 6A-D).

Incidence of treatment-related toxicities according to risk groups

Detailed toxicity data were available from 482 patients (66 LR-R, 148 LR-S, 11 HR-R, and 257 LR-I/HR-S patients). In the LR arm, incidence of severe infections (grade 2-4 according to the Common Terminology Criteria for Adverse Events grading) differed substantially between the reduced and standard treatment arms (Figure 7; supplemental Tables 2 and 3). Because toxicity data were available for only 11 of 19 patients in the HR-R arm, a comparison between the HR-R and HR-S arms was not meaningful. Patients in the LR-R arm reported significantly fewer infectious complications during the first and second part of delayed intensification (DI) than those in the LR-S arm (P = .01). However, there was no significant difference between the LR-R and LR-S arms in mucositis within the first and second part of DI. Severe infectious complications grade 3 or higher occurred in 3 patients in the LR-S arm and in 8 patients in the LR-I/HR-S arm, but no patient in the reduced treatment arms experienced an infection grade 3 or higher within the first part of DI. A reduction of infectious complications, including mucositis, was not only seen in the first part of DI, in which treatment reduction was applied for the reduced treatment arm, but also in the second part of DI, which was identical for all treatment arms. No differences were found regarding neurotoxicity, hepatotoxicity, thrombosis, or pancreatitis.

Discussion

In the past decades, several study groups for childhood ALL achieved a tremendous improvement in EFS and OS, resulting in relapse rates approximating TRM. The overall results of the CoALL 07-03 trial are comparable to those of other international study groups, with excellent long-term results with a 10-year pEFS of 83.5% and an OS of 90.7% in the entire patient population, including remission failures. TRM rate (1.7%) was low compared with that of the predecessor trial CoALL 06-97 (3.2%). We identified at least 2 potential variables accounting for the improvement in outcome: the implementation of polymerase chain reaction–based MRD measurements and the introduction of pegylated asparaginase could have contributed to better pEFS and OS, as previously implied in a report on study UKALL 2003.25

Several study groups have aimed at treatment reduction in patients with the most favorable prognosis.25-28 However, divergent results were generated by the different study groups in terms of the feasibility of treatment reduction. The BFM-ALL and the UKALL study groups randomly assigned patients with standard-risk ALL to either reduced intensity or standard delayed intensification. In the BFM-ALL 2000 study, there was a significantly increased event rate in the reduced treatment arm,28 whereas in UK-ALL 2003, treatment reduction was not associated with reduced efficacy according to study criteria.25 Likewise, in the nonrandomized Dutch Childhood Oncology Group (DCOG) ALL10 trial, therapy reduction during delayed intensification did not deteriorate outcome for patients in the standard group.27

One reason for the opposing data could be a varying definition of MRD negativity. The BFM and DCOG groups defined standard-risk...
patients by MRD less than $10^{-4}$ EOI and at week 11, with MRD marker sensitivity of at least $10^{-4}$.\textsuperscript{27,28} In the UKALL study, randomized patients classified as LR MRD were defined by undetectable MRD at day 29, or detectable MRD (less than $10^{-4}$ at day 29 and undetectable by week 11).\textsuperscript{25} These varying MRD definitions resulted in a different proportion of patients who were eligible for treatment reduction, ranging from 34% (BFM) to 24.9% (DCOG).\textsuperscript{25,27,28} In the CoALL 07-03 trial, truly negative, undetectable MRD at day 15 and at day 29 EOI were required for treatment reduction. This very strict definition led to a very low proportion of only 5% MRD-negative HR patients. Another reason for the discrepancy could be the different criteria used to assess noninferiority. In summary, our long-term results illustrate that it might be possible to reduce chemotherapy without impairing an excellent outcome, which is in line with the results of UKALL and DCOG. In the past years, more refined genetic risk profiles combined with treatment response evaluation have been evaluated, and it remains to be shown whether a more advanced risk assessment can be successfully translated into a more precise allocation of patients to different treatment intensities or modalities.

In our CoALL 03-07 trial presented here, patients in the LR-R treatment arm had significantly fewer infectious complications during late intensification. This outcome can be attributed to a lesser treatment intensity, given that the length of treatment was the same for the standard and reduced arms. Moreover, the reduction in infectious complications was not associated with a significant reduction in mucositis, often suggested as a predisposing factor for infectious complications. The NOPHO ALL 2008 (Nordic Society for Pediatric Hematology and Oncology) trial reported an overall relapse of 7.7% and a TRM rate including SMN of 4.5% for pediatric patients aged 1 to 17 years.\textsuperscript{29} Similar results were published by the DCOG for T-ALL patients whose toxic death rate approximated the relapse rate.\textsuperscript{27} The results of these and other study groups clearly illustrate that the upper limit of treatment intensity has been reached.

The lack of a correlation between in vitro and in vivo drug response behavior in our clinical trial may point toward the limited clinical value of in vitro drug testing. We believe that the complexity of host-specific factors, such as tumor microenvironment, individual drug metabolism, various resistance mechanisms, and the interplay in modern multi-agent chemotherapy, are inadequately reflected by in vitro drug sensitivity testing, at least in children with B-precursor ALL. Consequently, the PVA drug testing has been replaced by in vivo MRD assessment in the stratification procedure of the subsequent CoALL 08-09 trial. The data presented here may also serve as a cautionary note for ongoing ex vivo drug screening approaches to targets in other tumor entities identified by (epi-) genomic profiling.\textsuperscript{30} To prove the feasibility of treatment reduction in ALL, a randomized trial including advanced clinical diagnostics of ALL has been conceived as a joint effort of CoALL, DCOG, UKALL, NOPHO, and other study groups.

Acknowledgments

The authors are grateful to the Fördergemeinschaft Kinderkrebs-Zentrum for financial support. They also thank all individuals who collaborated in the designated CoALL trials in the past.

Authorship

Contribution: G.E.J.-S. designed and supervised the research as study chair; M.L.d.B. and R.P. conceived and performed the in vitro pharmacosensitivity testing; U.Z.S. analyzed in vivo pharmacosensitivity; M.Z. performed statistical analyses; N.J., A.P., A.B., T.I., H.C., J.F., I.S., T.F., G.B., and G.E. were involved in the clinical care of the patients and provided patient data; F.S., M.A.H., and G.E. analyzed data; F.S., M.A.H., and G.E. wrote the manuscript; and all authors approved the final version of the manuscript.

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

A complete list of the members of the Cooperative Study Group for Childhood Acute Lymphoblastic Leukemia (CoALL) appears in the supplemental Appendix.

ORCID profile: T.F., 0000-0002-8517-9681.

Correspondence: Franziska Schramm, University Medical Centre Eppendorf, Clinic of Pediatric Hematology and Oncology, Martinistr 52, 20246 Hamburg, Germany; e-mail: fr.schramm@uke.de; or Gabriele Escherich, University Medical Centre Eppendorf, Clinic of Pediatric Hematology and Oncology, Martinistr 52, 20246 Hamburg, Germany; e-mail: escherich@uke.de.

References

1. Schmiegelow K, Attarbaschi A, Barzilai S, et al; Ponte di Legno toxicity working group. Consensus definitions of 14 severe acute toxic effects for childhood lymphoblastic leukaemia treatment: a Delphi consensus. Lancet Oncol. 2016;17(6):e231-e239.
2. O’Connor D, Bate J, Wade R, et al. Infection-related mortality in children with acute lymphoblastic leukemia: an analysis of infectious deaths on UKALL2003. Blood. 2014;124(7):1056-1061.
3. 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.
4. Escherich G, Tröger A, Göbel U, et al; CoALL study group, Hamburg, Germany. The long-term impact of in vitro drug sensitivity on risk stratification and treatment outcome in acute lymphoblastic leukemia of childhood (CoALL 06-97). Haematologica. 2011;96(6):854-862.
5. Graubner UB, Porzig S, Jorch N, et al. Impact of reduction of therapy on infectious complications in childhood acute lymphoblastic leukemia. Pediatr Blood Cancer. 2008;50(2):256-263.
6. Yende S, Linde-Zwirble W, May F, Weissfeld LA, Reis S, Angus DC. Risk of cardiovascular events in survivors of severe sepsis. Am J Respir Crit Care Med. 2014;189(9):1065-1074.
7. Jafarzadeh SR, Thomas BS, Warren DK, Gill J, Fraser VJ. Longitudinal study of the effects of bacteremia and sepsis on 5-year risk of cardiovascular events. Clin Infect Dis. 2016;63(4):495-500.
8. Frismanitas V, Dobay MP, Rinaldi A, et al. Ex vivo drug response profiling detects recurrent sensitivity patterns in drug-resistant acute lymphoblastic leukemia. *Blood*. 2017;129(11):e26-e37.

9. Swords RT, Azzam D, Al-Ali H, et al. Ex-vivo sensitivity profiling to guide clinical decision making in acute myeloid leukemia: a pilot study. *Leuk Res*. 2018;64:34-41.

10. Sanguinetti M, Posteraro B. New approaches for antifungal susceptibility testing. *Clin Microbiol Infect*. 2017;23(12):931-934.

11. van Dongen JJ, Seriu T, Panzer-Grümmayer ER, et al. Prognostic value of minimal residual disease in acute lymphoblastic leukaemia in childhood. *Lancet*. 1998;352(9142):1731-1738.

12. 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 leukaemia: results in 3184 patients of the AIEOP-BFM ALL 2000 study. *Blood*. 2010;115(16):3206-3214.

13. Schrappe M, Valsecchi MG, Bartram CR, 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):2077-2084.

14. Càve H, van der Werff ten Bosch J, Suciu S, et al; European Organization for Research and Treatment of Cancer—Childhood Leukemia Cooperative Group. Clinical significance of minimal residual disease in acute lymphoblastic leukemia. European Organization for Research and Treatment of Cancer—Childhood Leukemia Cooperative Group. *N Engl J Med*. 1998;339(9):591-598.

15. Coustan-Smith E, Behm FG, Sanchez J, et al. Immunological detection of minimal residual disease in children with acute lymphoblastic leukemia. *Lancet*. 1998;351(9102):550-554.

16. Den Boer ML, Harms DO, Pieters R, et al. Patient stratification based on prednisolone-vincristine-asparaginase resistance profiles in children with acute lymphoblastic leukemia. *J Clin Oncol*. 2003;21(17):3262-3268.

17. van der Velden VH, Cazzaniga G, Schrauder A, et al; European Study Group on MRD detection in ALL (ESG-MRD-ALL). Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. *Leukemia*. 2007;21(4):604-611.

18. Janka GE, Winkler K, Juergens H, Goebel U. Early intensification therapy in high-risk childhood acute lymphocytic leukemia: lack of benefit from high-dose methotrexate. *Haematol Blood Transfus*. 1987;30:456-460.

19. Harms DO, Janka-Schaub GE. Co-operative Study Group for Childhood Acute Lymphoblastic Leukemia (COALL): long-term follow-up of trials 82, 85, 89 and 92. *Leukemia*. 2000;14(12):2234-2239.

20. Jürgens H, Janka G, Ibrahim M, Tonert C, Winkler K, Göbel U. Prognostic significance of exposure to intermediate-dose methotrexate in children with standard risk ALL: the COALL 82/85 Experience. In: Hiddemann W, Büchner T, Wörmann B, Plunkett W, Keating M, Andreeff M, eds. Berlin, Heidelberg: Springer; 1992:338-342.

21. Escherich G, Zimmermann M, Janka-Schaub G; CoALL study group. Doxorubicin or daunorubicin given upfront in a therapeutic window are equally effective in children with newly diagnosed acute lymphoblastic leukemia. A randomized comparison in trial CoALL 07-03. *Pediatr Blood Cancer*. 2013;60(2):254-257.

22. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481.

23. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. 2nd ed.. New York, NY: Wiley; 2011.

24. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann Stat*. 1988;16(3):1141-1154.

25. Vora A, Goulden N, Wade R, et al. Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): a randomised controlled trial. *Lancet Oncol*. 2013;14(3):199-209.

26. 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.

27. Pieters R, de Groot-Kruseman H, Van der Velden V, et al. Successful Therapy Reduction and Intensification for Childhood Acute Lymphoblastic Leukemia Based on Minimal Residual Disease Monitoring: Study ALL10 From the Dutch Childhood Oncology Group. *J Clin Oncol*. 2016;34(22):2591-2601.

28. Schrappe M, Bleckmann K, Zimmermann M, et al. Reduced-intensity delayed intensification in standard-risk pediatric acute lymphoblastic leukemia defined by undetectable minimal residual disease: results of an international randomized trial (AIEOP-BFM ALL 2000). *J Clin Oncol*. 2018;36(3):244-253.

29. Toft N, Birgens H, Abrahamsson J, et al. Results of NOPHO ALL2008 treatment for patients aged 1-45 years with acute lymphoblastic leukemia. *Leukemia*. 2018;32(3):606-615.

30. Worst BC, van Tilburg CM, Balasubramanian GP, et al. Next-generation personalised medicine for high-risk paediatric cancer patients—the INFORM pilot study. *Eur J Cancer*. 2016;65:91-101.