IntReALL SR 2010

International Study for Treatment of Standard Risk Childhood Relapsed ALL 2010

A randomized Phase III Study Conducted by the Resistant Disease Committee of the International BFM Study Group

Statistical Analysis Plan
Version 3.2 – 12 July 2012

Based on protocol IntReALL SR 2010
version 1.7 (dated 01.07.2012)

Name	Role	Signature	Date
Dr. med. Arend von Stackelberg	Principal investigator		5 Sept 2012
Prof. Vaskar Saha	Co-chair investigator		
Ms Sharon Love	Trial Statistician	S Love	4 Sept 2012
Mr Cristian Ciria	Trial Statistician		4 Sep 2012

Centre for Statistics in Medicine

CANCER RESEARCH UK

Contents

1. INTRODUCTION

This document details the proposed presentation and analysis for the main paper(s) reporting results from the International Randomized phase III study for treatment of Standard Risk Childhood Relapsed Acute Lymphoblastic Leukemia (ALL) (IntReALL2010). The results reported in these papers should follow the strategy set out here. Subsequent analyses of a more exploratory nature will not be bound by this strategy, though they are expected to follow the broad principles laid down here. The principles are not intended to curtail exploratory analysis (for example, to decide cut-points for categorisation of continuous variables), nor to prohibit accepted practices (for example, data transformation prior to analysis), but they are intended to establish the rules that will be followed, as closely as possible, when analysing and reporting the trial.

The analysis strategy will be available on request when the principal papers are submitted for publication in a journal. Suggestions for subsequent analyses by journal editors or referees, will be considered carefully, and carried out as far as possible in line with the principles of this analysis strategy; if reported, the source of the suggestion will be acknowledged.

Any deviations from the statistical analysis plan will be described and justified in the final report of the trial. The analysis should be carried out by an identified, appropriately qualified and experienced statistician, who should ensure the integrity of the data during their processing. Examples of such procedures include quality control and evaluation procedures.

1.1 Trial Statistician(s)

Ms Sharon Love
Centre for Statistics in Medicine
University of Oxford
Wolfson College Annexe
Linton Road
Oxford OX2 6UD

Tel: +44 (0)1865 284414
Fax: +44 (0)1865 284424
Email: sharon.love@csm.ox.ac.uk

Mr Cristian Ciria
Centre for Statistics in Medicine
University of Oxford
Wolfson College Annexe
Linton Road
Oxford OX2 6UD

Tel : +44 (0)1865 284 437
Fax: +44 (0)1865 284 424
Email: cristian.ciria@csm.ox.ac.uk

1.2 Principal Investigator

Dr. med. Arend von Stackelberg
Pädiatrische Onkologie / Hämatologie
Charité, OHC
Augustenburger Platz, 13353 Berlin
Heerstr. 111a
14055 Berlin

Tel: +4930 61621381
Mobil: +49(0)1716953754
Email: Arend.Stackelberg@charite.de

1.3 Co-chair Investigator

Prof. Vaskar Saha
Pediatric and Adolescent Oncology,
Manchester Academic Health Sciences Centre,
Central Manchester and Manchester
Children's Hospitals NHS Trust,
Christie Hospital, Wilmslow Road,
Manchester, M20 4BX, UK

Tel: +44 161 446 3023
Fax: +44 161 446 3092
Email: vaskar.saha@manchester.ac.uk

1.4 Trial Manager

Corinna Meier-Windhorst
Charité Clinical Trials Centre (KKS)
Augustenburger Patz 1, 13353
Berlin, Germany

Tel: +49 30 450 553894
Fax: +49 30 450 553937
Email: corinna.meier-windhorst@charite.de

1.5 Data Manager

Dr Andrea Barth
Universitätsklinikum Heidelberg
Pediatric Oncology/Hematology
Marvin Database of the GPOH
Im Neuenheimer Feld 430
60210 Heidelberg

Tel: +49 6221 568354
Support Phone: +49 511 5326717
Email: Andrea.Barth@med.uniheidelberg.de

2. BACKGROUND INFORMATION

Though survival of children with acute lymphoblastic leukemia (ALL) has considerably improved over the past few decades, relapsed ALL remains a leading cause of mortality in children with cancer. Given the rarity of the disease, prospective clinical trials need to be coordinated within an international cooperative group such as the International BFM Study Group (I-BFM-SG). Within the group, over the last few years two different treatment protocols, ALL-REZ BFM 2002 and ALL R3 have been used by most study groups for treatment of relapsed ALL. Both trials have produced comparable results. The trials risk stratified patients based on duration of first remission, immunophenotype, site of relapse and post induction minimal residual disease (MRD) levels to identify patients who should be transplanted. For non-HR or standard risk (SR) patients both ALL-REZ BFM 2002 and ALL R3 have achieved better results than previous trials. Both protocols have however been primarily used in patients relapsing off different frontline protocols. Thus there is need for a prospective randomized controlled comparison across the study groups (randomization 1), before a uniform backbone for further trials can be developed. In SR patients, survival may be improved by modifying the consolidation therapy using targeted non-myelotoxic drugs. As ideal candidate, epratuzumab (humanised chimeric anti CD22 antibody) will be randomly tested in combination with conventional chemotherapy (randomization 2). CD22 is well expressed in all B-cell precursor ALL cells. Epratuzumab has been developed in combination phase I and II trials in childhood relapsed ALL and has shown a favorable toxicity profile and moderate antileukemic activity.

2.1 Aims of the Trial

The main goal of this study is to improve the prognosis of children and adolescents with standard risk relapsed acute lymphoblastic leukemia. Furthermore, goal is to set up a large international study group platform allowing for optimization of standard treatment strategies and integration of new agents.

2.1.1 Primary objectives

- Improvement of event-free survival (EFS1) with arm B (ALLR3) compared to arm A (ALL-REZ BFM 2002) in SR patients
- Improvement of EFS2 after consolidation with versus without epratuzumab in SR patients

2.1.2 Secondary objectives

- Improvement of overall survival (OS) with arm B (ALLR3) compared to arm A (ALL-REZ BFM 2002) in SR patients
- Improvement of OS after consolidation with versus without epratuzumab in SR patients
- Rate of second complete remission rate (CR2) of arm B versus arm A in SR Patients
- Rate of stem-cell transplantation (SCT) performed in arm B versus arm A in SR patients
- Improvement of MRD reduction during consolidation with versus without epratuzumab
- Rate of MRD negativity prior to SCT with arm B versus arm A in SR Patients
- Rate of MRD negativity prior to SCT after consolidation with versus without epratuzumab in SR patients
- Toxicity of randomised arm B versus arm A in SR patients
- Toxicity of consolidation with versus without epratuzumab in SR patients
- Pharmacokinetic of epratuzumab in context with Arm A and B in SR patients

2.1.3 Exploratory analyses

Look at the difference in IntReALL SR 2010 treatment results between the differing countries.

2.2 Study Design

The IntReALL SR 2010 trial is an inter-group, international multi-centre, treatment optimization trial. IntReALL SR 2010 contains the following two branches:

SR1 - Randomization: Prospective, randomized, open label phase III trial comparing the efficacy and toxicity of standard protocols ALL-REZ BFM 2002 (Arm SR-A) and ALLR3 (Arm SR-B).

SR2 - Randomization: Prospective, randomized, open label phase III trial comparing the efficacy and toxicity of the respective consolidation therapy within the arm selected with randomization 1 plus epratuzumab versus standard consolidation alone.

Recruitment for SR1 and SR2 is expected to take four years and patients will be followed up for three years after the last patient accrual. Despite defining the following expected dates, there could be a potential slippage and these would then change.

Date of expected start of recruitment:	Oct 2012[a]
Date expected of first DSMC:	Oct 2013[b]
Date expected of first TSC:	Jan 2014[c]
Date expected of first safety report:	Oct 2013[d]
Date of expected end of recruitment:	Sep 2016
Date of expected start of interim analysis:	Oct 2016[e]
Date of expected end of follow-up:	Sep 2019
Date of expected start of analysis:	Oct 2019

[a] Target number of subjects SR1 per year: 100 in the first and second year and 200 per year thereafter
Target number of subjects SR2 per year: 75 in the first and second year and 150 per year thereafter
[b] First DSMC meeting at 1 year recruitment or 100 patients recruited whichever first; expected DSMC meetings every 6 months during first 4 year recruitment and annually thereafter
[c] First TSC meetings three months after DSMC; TSC meetings every 6 months during first 4 year recruitment and annually thereafter
[d] Annual safety report from first year recruitment to end follow-up
[e] Note that SR2 randomisation is 5/6 weeks behind SR1, so at interim SR2 data will be less mature

Participating centres: International
Number of countries: 20; Participating countries:

1. Australia
2. Austria
3. Belgium
4. Czech Republic
5. Denmark
6. Finland
7. France
8. Germany
9. Ireland
10. Israel
11. Italy
12. Japan
13. Netherlands
14. New Zealand
15. Norway
16. Poland
17. Portugal
18. Sweden
19. Switzerland
20. UK

2.3 Risk Group Stratification

2.3.1 Stratification

The risk stratification is based on the clinical parameters time-point to first relapse, site of first relapse and immunophenotype of first relapse which are defined as follows: The time-point of relapse is defined in relation to the date of primary diagnosis and cessation of frontline therapy (Table 1). In the rare case of inadequately short frontline therapy (f.e. according to BNHL/ALL protocols) the interval after completion of primary therapy is the decisive one for time-point classification. The site of relapse is classified on the basis of conventional light microscopy using the FAB criteria. In morphologically undetermined situations such as bone marrow involvement of extramedullary relapse around 5%, minimal residual disease quantification may be regarded as secondary criteria to determine the extent of bone marrow involvement. The immunophenotype is defined according to EGIL criteria.

Table 1 Definition of time point of relapse

Time-point	After primary diagnosis		After completion of primary therapy
Very early	< 18 months	and	< 6 months
Early	≥ 18 months	and	< 6 months
Late			≥ 6 months

Table 2 Definition of site of relapse

Bone marrow		M1 (< 5% blasts)	M2 (≥ 5% and < 25% blasts)	M3 (≥ 25% blasts)
extramedullary relapse	No	**No ALL relapse**	**Requires follow up control**	**Isolated bone marrow relapse**
	Yes	**Isolated extramedullary relapse**	**Combined bone marrow / extramedullary relapse**	

2.3.1 Definition of the SR group

Standard risk group is defined as late or early isolated extramedullary relapse of BCP or T-ALL; late or early combined bone marrow / extramedullary relapse, late isolated bone marrow relapse of BCP ALL.

Table 3 Definition of IntReALL SR 2010 risk group

	Immunophenotype: B-cell precursor			Immunophenotype: (pre) T		
\ Site Time-point \	Extramed. Isolated	Bone marrow combined	Bone marrow isolated	Extramed. isolated	Bone marrow combined	Bone marrow isolated
Very early	**HR**	**HR**	**HR**	**HR**	**HR**	**HR**
Early	**SR**	**SR**	**HR**	**SR**	**HR**	**HR**
Late	**SR**	**SR**	**SR**	**SR**	**HR**	**HR**

HR: High Risk; SR: Standard Risk

2.4 Study Population

Eligible for the study are principally children with relapsed acute lymphoblastic leukemia.

2.4.1 Inclusion criteria

- Morphologically confirmed diagnosis of relapsed precursor B-cell or T-cell ALL
- Children up to 18 years of age at inclusion
- Meeting SR criteria: late isolated or late/early combined BCP BM relapse, any late/early isolated extramedullary relapse
- Patient enrolled in a participating centre
- Written informed consent
- Start of treatment falling into the study period
- No participation in other clinical trials 30 day prior to study enrolment except trials for primary ALL

Inclusion criteria specific for the epratuzumab randomization

- Precursor B-cell immunophenotype of ALL. A CD22 expression level is not considered as inclusion criterium
- M1 or M2 bone marrow status after induction

2.4.2 Exclusion criteria

- BCR-ABL / t(9;22) positive ALL
- Pregnancy or positive pregnancy test (urine sample positive for β-HCG > 10 U/l)
- Sexually active adolescents not willing to use highly effective contraceptive method (pearl index <1) until 2 years after end of antileukemic therapy
- Breast feeding
- Relapse post allogeneic SCT
- The whole protocol or essential parts are declined either by patient himself/herself or the respective legal guardian
- No consent is given for saving and propagation of pseudonymized medical data for study reasons
- Severe concomitant disease that does not allow treatment according to the protocol at the investigator's discretion (e.g. malformation syndromes, cardiac malformations, metabolic disorders)
- Karnovsky / Lansky score < 50%
- Subjects unwilling or unable to comply with the study procedures
- Subjects who are legally detained in an official institute

Allowed systemic diseases and concomitant medication

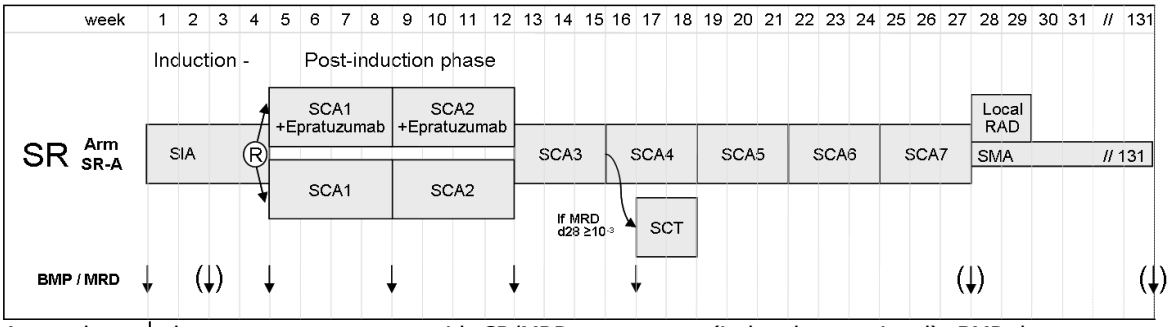
- Patients with systemic diseases such as Down syndrome, cystic fibrosis or diabetes mellitus are eligible for enrolment in this study. Due to the anticipated increase of toxicity dose reductions are suggested after discussion with the national study centre
- Any kind of concomitant medication given due to medical reasons is allowed except antileukemic therapy and investigational drugs other than scheduled in the protocol

2.5 Treatment Interventions

2.5.1 Arm SR-A (ALL-REZ BFM)

Patients randomized to the SR-A Arm receive induction, consolidation and maintenance therapy according to a modified protocol ALL-REZ BFM 2002 with Protocol II-IDA as 1st consolidation element. Induction consists of the course SIA (Standard Risk Induction Arm A), at day 29 the cytological remission and the MRD status are quantified. Post induction, the SCA1 course (Standard Risk Consolidation Arm A1) is given as first consolidation. Patients are randomized to receive or not to receive epratuzumab in addition to SCA1 and SCA2. Patients with MRD ≥ 10^{-3} (more than 1 leukemic cell out of 1000 mononucleated cells) at week 5 and those with isolated extramedullary relapse and high-risk features are allocated to allogeneic SCT after the SCA3 course whereas the other patients continue until SCA7 course and with a maintenance therapy of 24 months duration and a total treatment duration of 131 weeks. At the beginning of maintenance therapy, patients without CNS irradiation receive triple intrathecal therapy once every 4 weeks for a total of 6 times. Patients with CNS involvement receive cranial irradiation at a dose of 18 Gray (Gy) and not intrathecal therapy during maintenance treatment. Boys with testicular relapse receive local treatment according to the national guidelines and as described in protocol –chapter 11.8.2 and 11.9.2. MRD is monitored before every treatment element until SCT or until SCA4 course (week 16) as additional response parameter but not communicated to the treating centres.

Figure 1 Treatment schedule for SR-A arm

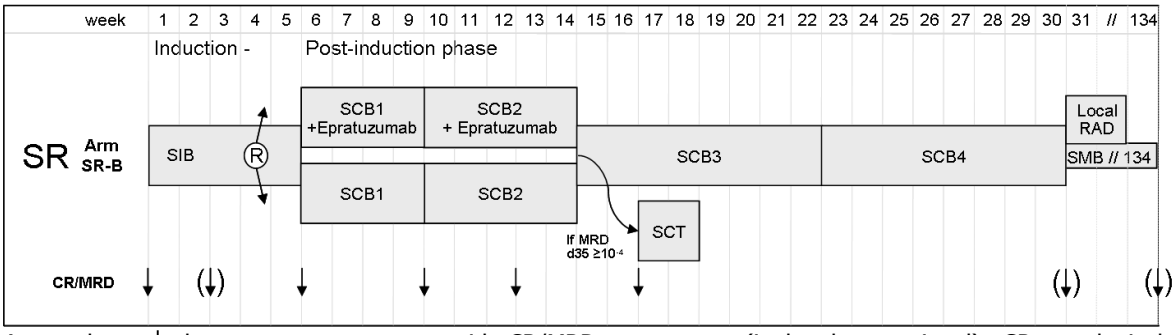


Arrow down ↓, bone marrow puncture with CR/MRD assessment (in brackets optional); BMP, bone marrow puncture; CR, cytological remission; MRD, minimal residual disease; ®, randomization; RAD, irradiation; SCA 1-7, SR consolidation 1-7 arm A; SCT, stem-cell transplantation; SIA, SR induction arm A; SMA, SR maintenance arm A; SR, standard risk group

2.5.1 Arm SR-B (R3)

Patients randomized to the SR-B Arm receive induction, post-induction and maintenance therapy according to the protocol ALL-R3. Induction consists of the courses SIB (former R3-Induction I, II, and III). After regeneration of blood count at day 35 (start of week 6) the cytological remission and the MRD status are quantified. Patients are randomized to receive or not to receive epratuzumab in addition to the R3-Consolidation (SCB). Patients with MRD ≥ 10^{-4} (more than 1 leukemic cell out of 10000 mononucleated cells) at day 35 and those with early isolated extramedullary relapse are allocated to allogeneic SCT after the SCB 2 (former R3-Intensification) course whereas the other patients continue with 2 SCB cycles (SCB3, SCB4) and with a maintenance therapy of 24 months duration and a total treatment duration of 134 weeks. At the beginning of maintenance therapy (week 31), patients with CNS involvement receive cranial irradiation at a dose of 18 Gray (Gy). Boys with testicular relapse receive local treatment according to the national guidelines and as described in protocol - appendix chapter 11.8.2 and 11.9.2. MRD is closely monitored before every treatment element until SCT or the 1st interim maintenance cycle (week 15) as additional response parameter but not communicated to the treating centres.

Figure 2 Treatment schedule for SR-B arm



Arrow down ↓, bone marrow puncture with CR/MRD assessment (in brackets optional); CR, cytological remission; MRD, minimal residual disease; ®, randomization; RAD, irradiation; SCB 1-4, SR consolidation 1-4 arm B; SCT, stem-cell transplantation; SIB, SR induction arm B; SMB, SR maintenance arm B; SR, standard risk group

2.6 Sample Size

After opening of all participating national groups, a recruitment of about 200 patients per year can be expected. It is expected to take up to 2 years to open all centres, so in the first and second year recruitment will be assumed to be half that given above. Follow-up will continue until 3 years after the last patient is recruited to give a median follow-up of at least 4 years.

2.6.1 Sample size for SR induction, randomisation SR-1

SR induction follows an open label 2-arm randomized design (SR1). The primary endpoint is EFS1 of arm A (ALL-REZ BFM 2002) versus arm B (ALL-R3). The expected 4 year EFS1 of arm A (ALL-REZ BFM 2002 arm) is 65%. The hypothesis of the study is that EFS1 can be improved by at least 10% with arm B (ALL-R3). To detect this difference at a power of 82% and a p-value of 0.05, with 100 recruited in the first and second year and 200 per year thereafter, allowing for 3 years follow-up from the last patient recruitment and 5% loss to follow-up at 7 years, 306 patients per arm is required (207 events in total). The expected recruitment time for the total of 612 patients is 4 years. The power is calculated at 82% rather than the more usual 80% in order to give enough patients from SR1 for a power of 80% to be achieved in SR2 (see section 2.6.2).

Note that it is possible that another country may take part in only the SR1 randomisation increasing the recruitment for SR1 in the first 2 years to 130 each year and to 260 per year thereafter. If this happens, then it is proposed to keep 4 years recruitment and 3 years follow-up and realise that there would be 80% power to show an 8.6% difference in 4 years EFS1 to be significant using a p-value of 0.05 and allowing for a 5% loss to follow-up at 7 years with a total of 786 patients (271 events in total).

2.6.2 Sample size for SR post-induction/ consolidation, epratuzumab, randomisation SR-2

Within open label randomization SR2 the prognostic effect of epratuzumab combined with the conventional intensive consolidation therapy of arm A or B is investigated. It is expected that the effect of epratuzumab will be independent from the respective treatment arm A or B but randomisation will be stratified by the induction arm to give balanced groups. Since patients with M1 or M2 bone marrow status after induction are included, thus nonresponse and induction death are possible events, event-free survival is applied as outcome parameter. The expected EFS2 after both SR arms A and B is 74% at 4 years (induction failures, M0 and M3 are off study and not randomized). The hypothesis of the study is that EFS2 can be improved by at least 10% with epratuzumab. To detect this difference at a power of 80% and a p-value of 0.05, with 75 recruited in each of the first two years and 150 per year thereafter, allowing for 3 years follow-up

from the last patient recruitment and 5% loss to follow-up at 7 years a number of 228 patients per arm is required (109 events in total). The expected recruitment time for the total of 456 patients is 4 years.

2.7 Randomization

Informed consent for participation in the randomized parts of the study is given separately from the consent to the enrolment into the study. This enables the patients and parents (or persons entitled to custody) to have the choice to participate in the randomization, or not to be randomized. Non-randomised patients will be treated as specified for the control arm but their data will not be used for the main analysis. After consent has been given and the patient has been randomized, consent can be withdrawn at any time. Randomization occurs when the diagnostic findings required for inclusion/exclusion criteria and randomisation stratification are complete and the informed consent for randomization has been obtained.

Randomization is performed centrally by the internet-based data entry system with stratification according to the clinical parameters relevant for each randomization. Randomization will use variable blocksize.

The first randomization of the SR group (SR1), arm A versus B, is performed immediately after first relapse diagnosis during the cytoreductive prophase and is stratified by country-group and site of relapse (IBM, CBM, IEM).

4 country-groups:

1. Germany, Austria, Switzerland, Italy, Czech Republic, Poland, Israel
2. UK, Netherlands, Ireland, Australia, New Zealand
3. Denmark, Finland, Norway, Sweden, France, Belgium, Portugal
4. Japan

The second randomization of the SR group (SR2), addition of epratuzumab with conventional consolidation is performed directly after the bone marrow assessment after induction (week 5/6). Patient with M3 marrow or induction failures (SR1) have reached their study endpoint as nonresponders and are not included into the 2nd randomization. Patients with M0 marrow will be reassessed after 1 week and are randomized only after proven M1 or M2 marrow, otherwise excluded. Drug labeling and shipment of epratuzumab to the local treating centre will be provided by a central institution in Europe for the first 2 doses independently of the randomization result, such that the drug will be immediately available after randomization. The randomization will be stratified by the induction regimen (arm A or B), country-group, and site of relapse (IBM, CBM, IEM).

3 country-groups:

1. Germany, Austria, Switzerland, Italy, Czech Republic, Poland, Israel
2. UK, Netherlands, Ireland, Australia, New Zealand
3. Denmark, Finland, Norway, Sweden, France, Belgium, Portugal

2.8 Definition of Primary and Secondary Outcomes

The null hypothesis for both primary outcomes is that there are no differences between the two treatment groups, ALLR3 and ALL-REZ BFM 2002 for EFS1, and with and without epratuzumab for EFS2.

Primary outcome:

- Event free survival (EFS1) for SR1 is defined as the time from randomisation to the first of induction failure, relapse, death from any cause or second malignancy, or is censored at the date of last follow-up. Note that SCT is not considered as event and is also not censored in the EFS1 analyses for SR1
- Event free survival (EFS2) for SR2 is defined as the time from randomisation to the first of persisting non-response, progression, death from any cause or second malignancy, or is censored at the date of last follow-up. Note that for SR2 analysis, M0 or M3 patients at the start of consolidation and induction failures (SR1) will be off study and not included in the analysis. Note that SCT is not considered as event and is also not censored in the EFS2 analyses for SR2

Secondary outcome:

- Overall Survival (OS) for SR1 and SR2 is defined as the time from randomisation to death from any cause. Note that SCT is not considered as event and is also not censored in the OS analyses
- Second complete remission rate (CR2) for SR1 patients is defined as the proportion of patients with remission bone marrow and no further evidence of persistent leukemic lymphoblast based on cytologic, histopathologic, radiologic or clinical findings on the total randomised patients for SR1 at the end induction
- Minimal residual disease (MRD) level is defined as good response for arm A (ALL-REZ BFM 2002) whether MRD at day 1 of week 5 $< 10^{-3}$ and for arm B (ALL-R3) whether MRD at day 1 of week 6 $< 10^{-4}$; poor response for arm A whether MRD at day 1 of week 5 $\geq 10^{-3}$ and for arm B whether MRD at day 1 of week 6 $\geq 10^{-4}$
- MRD reduction for SR2 is defined as the MRD change between MRD at randomisation and minimum MRD value post-induction phase
- Toxicity is defined using CTCAE grade version 4.0 in Liver, GI, Infection, renal and CNS providing the grade is ≥3. Excluding haematological, other toxicities with CTCAE grade ≥ 3 will be used combined as a further category (and split by category if any subcategory of others is large enough for consideration on its own). Toxicity is also described using SAE, using treatment related deaths and second malignancy. Treatment related death is defined as a death with temporal and/or causal relationship to treatment that occurs during continuous CR. Second malignancy is defined as a histologically distinct second cancer or neoplasm

2.9 Outcomes Assessment Schedule

At diagnosis, a complete work up of infection parameters and organ function is performed as basic values for further monitoring. This includes biochemistry, coagulation parameters, virology, HLA-typing, ultrasound of abdomen, testes, lymph nodes and venous vessels for central line implantation, echocardiogram, chest X-ray, electroencephalography, and MRI of the head. A flow chart for all diagnostic measures is given in the following table by protocol week and by risk group and treatment arm.

Table 4 Disease and drug specific diagnostic procedures by protocol week and by risk group and treatment arm

Parameters \ Week	Arm	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
BM Cytology	SR-A	+		(+)		+				+				+			+											(+)			
	SR-B	+		(+)			+				+			+		+															(+)
BM Immunophenotype		+																													
BM Genetics		+																													
BM MRD	A	+		(+)		+				+				+			+											(+)			
	B	+		(+)			+				+			+		+															(+)
CSF cytology		+	(+)	(+)																											
Asparaginase PK	SR-A		+	+	+	+	+	+	+						+	+		+	+		+	+		+	+		+	+			
	SR-B		+	+	+	+		+	+		+	+		+																	
Epratuzumab PK (serum levels)	SR-A					+	+	+	+	+	+	+	+	+	+	+	+	+	+													
	SR-B						+	+	+	+	+	+	+	+	+	+	+	+	+	+												
Epratuzumab PD (CD22 quantification & Ig levels)	SR-A					+		+		+		+		+		+		+		+							+				
	SR-B						+		+		+		+		+		+		+		+							+				

Abbreviations: BM, bone marrow; CSF, cerebro-spinal fluid; MRD, minimal residual disease analysis (PCR and flow); PD, pharmacodynamic analysis; PK, pharmacokinetic analysis; SR, standard risk;

2.9.1 Response to therapy

All manifestations of an ALL relapse have to be monitored until their complete resolution using the most appropriate diagnostic methods. These include bone marrow aspirates and lumbar punctures at the beginning of each treatment block, and imaging techniques for other sites. Complete blood and differential counts are performed daily until the disappearance of blasts, then 2 to 3 times per week.

Mandatory bone marrow aspirates are performed at the time of relapse, and further on as described in the respective chapter of the treatment arms. Additional bone marrow aspirates are performed prior to SCT and prior to start of maintenance therapy, to document a continuous cytological remission. Following SCT bone marrow aspirates are scheduled on day 30, 60, 100, 180 as well as during month 9, 12 and 18 after SCT.

Bone marrow aspirates are performed even after the achievement of a cytological remission to determine the level of MRD. This also applies to patients with an isolated extramedullary relapse.

3. DOCUMENTATION AND DATA MANAGEMENT

3.1 Documentation

3.1.1 Case Report Form (CRF)

Data collected on each subject will be recorded on an eCRF.

3.1.2 Investigator Site File (ISF)

All essential documents will be kept in the Investigator Site File which will be stored at the study site in accordance with ICH GCP chapter 8.

3.2 Data Entry and Data Management

All patient related data will be recorded in the electronic database under a pseudonym. Every patient will receive a pseudonym which will be unique for this individual patient. The database provides a confidential list, which relates these patient numbers to the patient's full name. This list will only be accessible to the treating physician and the monitor. Original patient files may be viewed by monitors, auditors and inspectors.

3.2.1 Data collection/Case Report Form (CRF)

All data will be recorded in an electronic Case Report Form. The study software used is based on the MARVIN system, Xclinical inc. The CRFs will be filled in by an authorized person (defined in the study team log) as soon as possible. They will be checked, dated and signed electronically by the investigator. All data will be recorded online. Data will be transferred between the workstation computer at the study site and the study server via a secure connection (secure socket layer /SSL) so that the data cannot be manipulated.

All patient data will be saved under a pseudonym. The study software will automatically generate a pseudonym for every new patient. A document for clear patient allocation will be printed and will be kept in the ISF.

3.2.2 Source Data and Patient Files

The information in original documents and records (e. g. patient files, laboratory notes) are defined as Source Data and will be reviewed by the Monitor for Source Data Verification. All data

that will be recorded directly in the eCRF without prior written or electronic record will be described in the protocol and considered to be Source Data.

3.2.3 Data processing

The biometrical centre will file all data electronically. To verify accuracy of the data, range, validity and consistency checks will be performed automatically by the database and additionally by the biometrical centre. Implausible or missing data can be corrected or added after consulting the Investigator. Documentation for these corrections will be stored with the eCRFs.

All validated data will be stored in the database MARVIN. After termination of the study and after completion of all entries, the database will be closed for further entries. This process will be documented.

4. QUALITY CONTROL AND DATA VALIDATION

4.1 Direct Access to Source Data

According to ICH-GCP and to the European laws (if applicable), the principal investigator must permit all authorized third parties access to the trial site and insight into the medical records of the trial subjects (source data). This permission includes the clinical trial monitors, auditors and other authorised employees of the sponsor, as well as members of the competent authorities. All these persons are sworn to secrecy.

4.2 Monitoring

The national cosponsors are responsible for the organisation of an adequate monitoring process in the respective country. An on-site monitoring in all participating centres and including all patients is planned. In case of frequent protocol violations, incomplete documentation, unanswered queries or other problems, for cause monitoring visits may be performed. The main emphasis of on-site monitoring should lie on the check of the informed consent forms and of the inclusion and exclusion criteria, as well as on the main efficacy and safety endpoints. In addition on-site monitoring visits make sure that the study is performed according to ICHGCP, and that the protocol is adhered to. Thus, on-site monitoring plays an important role in the support and training of participating trial sites.

A detailed monitor plan will be provided within a monitor manual by the sponsor.

The investigators allow the monitor to have access to all of the study materials needed for source data verification and proper review of the study process. All times, the sponsor/investigators/monitors will maintain the confidentiality of the study documents. Furthermore, problems with inconsistent and incomplete data will be discussed. By signing the declaration of informed consent the participants allow access to their documents. With the signature in the protocol, the investigators confirm that auditors and health authority inspectors may have access to the study documentation and accordant medical records. Auditors and inspectors are bound by professional confidentiality and may not pass on any personal information that comes to their knowledge. In the course of audits or inspections, data in the case report forms will be compared with the data for medical record. All the documentation held by the investigators within the scope of the clinical trial, as well as the drug logs of the study medications will be verified.

4.3 Audits/Inspections

Authorized representatives of the Sponsor, a regulatory authority, or an Independent Ethics Committee (IEC) may visit the centre to perform audits or inspections, including source data verification. The purpose of a Sponsor audit or inspections is to systematically and independently examine all the study activities and documents to determine whether these activities were conducted, and data were recorded, analysed, and accurately reported according to the protocol, Good Clinical Practise (GCP), guidelines of the International Conference on Harmonization (ICH), and any applicable regulatory requirements.

According to the pertinent European legislation, inspections of the trial sites may be performed by the competent authorities at any time during or after completion of the trial.

5. DATA AND SAFETY MONITORING COMMITTEE AND INTERIM ANALYSES

The study will be monitored by an independent Data and Safety Monitoring Committee (DSMC) that is specifically chosen to include an expert in paediatric oncology, in biostatistics and in clinical trials to ensure utmost competence and vigilance. The DSMC may meet before the start of the trial, then at 1 year recruitment or 100 patients recruited whichever occurs first, and then every 6 months during the first 4 years recruitment and annually thereafter in order to review the trial's progress, safety data (SAEs) and adherence to protocol. DSMC will make recommendations to the independent Trial Steering Committee (TSC) who will make decisions about whether the trial should continue.

Possible recommendations could include:

- No action needed, trial continues as planned
- Early stopping due, for example, to clear benefit or harm of a treatment, futility, or external evidence
- Stopping recruitment within a subgroup
- Extending recruitment (based on actual control arm response rates being different to predicted rather than on emerging differences) or extending follow-up
- Sanctioning and/or proposing protocol changes

TSC should also provide advice through its independent Chairman to the Inter-Group Trial Management Group (TMG), Charité – Universitätsmedizin Berlin (Sponsor), and trials office on all aspects of the trial.

During the course of the study one interim analysis after 4 years of recruitment for SR patients is planned to exclude significant results of the randomized questions that would require a preterm termination of the trial. The results of this analysis will be discussed with the DSMC and are not planned for general release. The DSMC will recommend discontinuing of the trial in all patients or in selected subgroups only if the result is likely to convince a broad range of clinicians (including participants in the trial) and the general clinical community. The ultimate decision for discontinuation of the trial lies with the independent TSC.

According to principal investigator and co-chair investigator the interim analysis will be done by the trial statistician and will be reviewed by an independent statistician who could be the one within the DSMC.

The open report for the DSMC will have overall results (not split by treatment arm) which will be seen by the PI and with permission from the DSMC can be distributed as a confidential document to the Inter-Group TMG, the funder and the sponsor.

5.1 Early stopping guidelines for treatment related mortality

Based on I-BFM and the R3 trial, the best estimate of expected mortality is 6% treatment related mortality (TRM) in SR patients. The stopping guideline for each randomisation is: Stop if the difference in number of deaths exceeds 3 standard deviations between the two randomised arms. This will be shown by a p-value of 0.001 or less on a two-sided Fisher's exact test. For this analysis, TRM will be estimated by all deaths in randomised patients regardless of cause and timing. This is done in order to remove bias introduced by multiple centres classifying deaths as treatment related (or not) and in light of the fact that this guideline will be used early in the trial when the majority of the deaths will be treatment related. This rule will be agreed by the DSMC and the DSMC will take this guideline into account when assessing the trial.

The following table illustrates for each expected DSMC, the border of tolerable differences in proportions of TRM between ALL-REZ BFM and ALL-R3 for the first randomization (SR1). The boundary is three times the standard deviation for the overall proportion of TRM based on the expected simple size for each DSMC. The boundary of tolerable difference in TRM between both arms will be calculated and a graph drawn for each DSMC according to the number of patients randomised and the overall proportions of TRM.

Table 7 Border of tolerable difference in proportion of TRM between randomised arms for first randomisation (SR1)

Overall proportion of TRM	SR1 - DSMC (expected sample size)[1,2]						
	1 year (N=100)	1.5 year (N=150)	2 years (N=200)	2.5 years (N=300)	3 years (N=400)	3.5 years (N=500)	≥4 years (N=612)
5%	6.5%	5.3%	4.6%	3.8%	3.3%	2.9%	2.6%
10%	9.0%	7.3%	6.4%	5.2%	4.5%	4.0%	3.6%
15%	10.7%	8.7%	7.6%	6.2%	5.4%	4.8%	4.3%
20%	12.0%	9.8%	8.5%	6.9%	6.0%	5.4%	4.9%
25%	13.0%	10.6%	9.2%	7.5%	6.5%	5.8%	5.3%
30%	13.7%	11.2%	9.7%	7.9%	6.9%	6.1%	5.6%
35%	14.3%	11.7%	10.1%	8.3%	7.2%	6.4%	5.8%
40%	14.7%	12.0%	10.4%	8.5%	7.3%	6.6%	5.9%
45%	14.9%	12.2%	10.6%	8.6%	7.5%	6.7%	6.0%
50%	15.0%	12.2%	10.6%	8.7%	7.5%	6.7%	6.1%

[1] SR1 - early stopping if the difference in proportion of TRM between ALL-REZ BFM 2002 and ALL-R3 arms exceeds three times the standard deviation (SD) for the overall proportion of TRM based on the expected sample size
[2] Note that despite defining the expected boundary for each DSMC, sample size and proportion of TRM could be different and boundaries should be recalculated

The following table illustrates for each expected DSMC, the border of tolerable differences in proportions of TRM between consolidation therapy with and without epratuzumab for the second randomization (SR2).

Table 8 Border of tolerable difference in proportion of TRM between randomised arms for second randomisation (SR2)

Overall proportion of TRM	SR2 - DSMC (expected sample size)[1,2]						
	1 year (N=75)	1.5 year (N=113)	2 years (N=150)	2.5 years (N=225)	3 years (N=300)	3.5 years (N=375)	≥4 years (N=456)
5%	7.5%	6.2%	5.3%	4.4%	3.8%	3.4%	3.1%
10%	10.4%	8.5%	7.3%	6.0%	5.2%	4.6%	4.2%
15%	12.4%	10.1%	8.7%	7.1%	6.2%	5.5%	5.0%
20%	13.9%	11.3%	9.8%	8.0%	6.9%	6.2%	5.6%
25%	15.0%	12.2%	10.6%	8.7%	7.5%	6.7%	6.1%
30%	15.9%	13.0%	11.2%	9.2%	7.9%	7.1%	6.4%
35%	16.5%	13.5%	11.7%	9.5%	8.3%	7.4%	6.7%
40%	17.0%	13.9%	12.0%	9.8%	8.5%	7.6%	6.9%
45%	17.2%	14.1%	12.2%	9.9%	8.6%	7.7%	7.0%
50%	17.3%	14.1%	12.2%	10.0%	8.7%	7.7%	7.0%

[1] SR2 - early stopping if the difference in proportion of TRM between consolidation therapy with and without epratuzumab exceeds three times the SD for the overall proportion of TRM based on the expected sample size
[2] Note that despite defining the expected boundary for each DSMC, sample size and proportion of TRM could be different and boundaries should be recalculated

For each expected DSMC, the proportion of TRM in both arms and the border of tolerable difference in proportion of TRM between arms for SR1 and SR2 will be also tabulated by country.

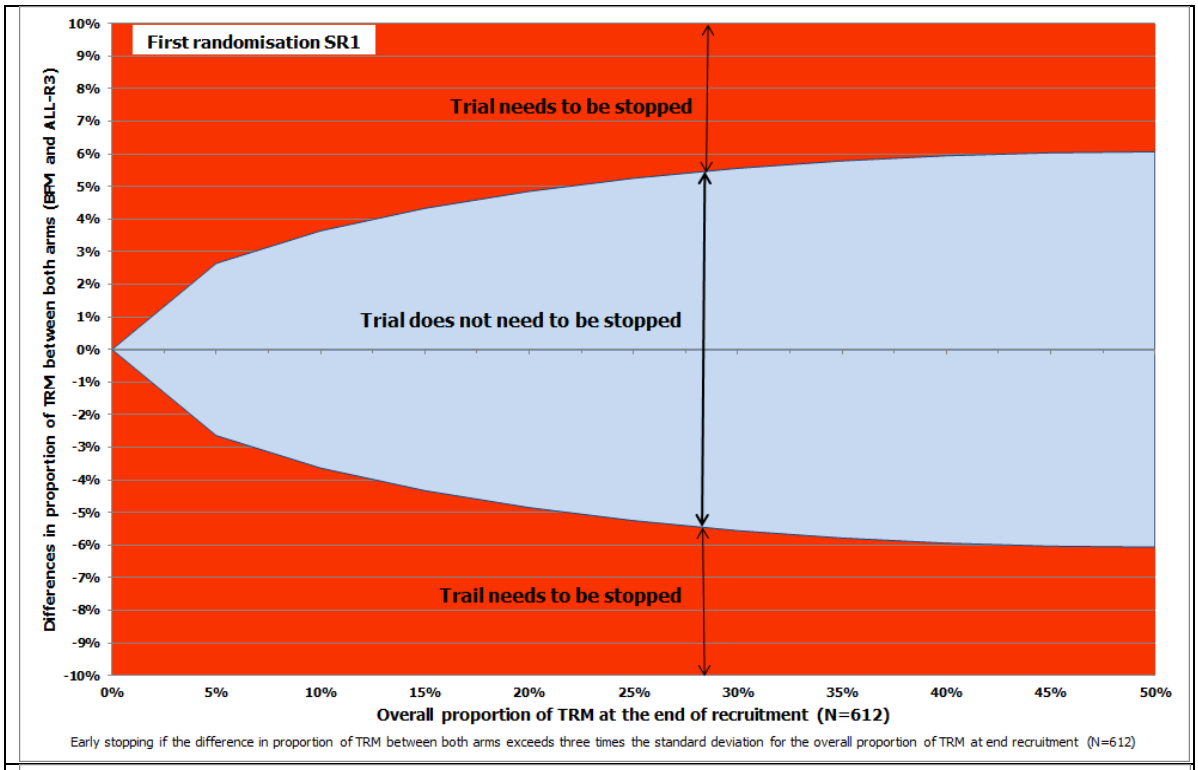
5.2 Interim safety analysis

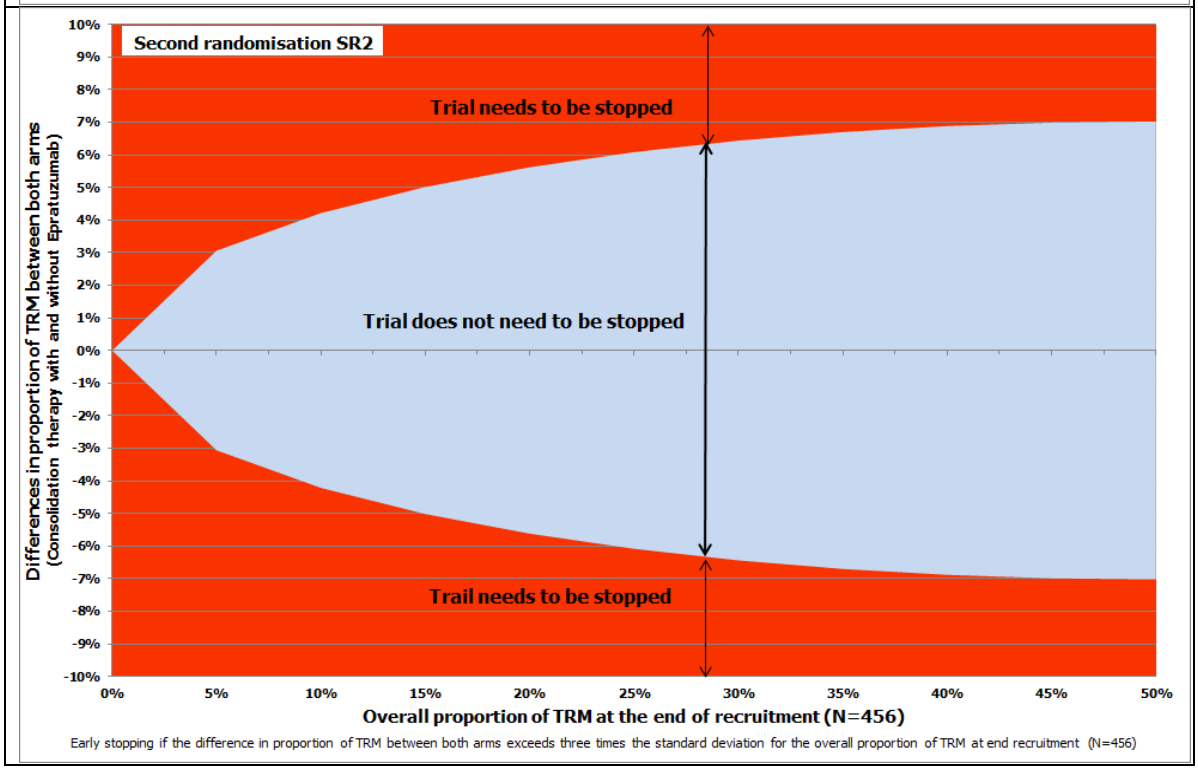
This analysis will check the safety and tolerability for treatment of SR childhood relapsed ALL. Detailed data on toxicity and proportion of patients experiencing any toxicity will be given overall and by stratification groups. SAEs, deaths and main characteristics of them will be assessed overall and by stratification groups. MRD results and proportion of negative/positive MRD will also be made available if applicable.

The proportion of patients experiencing any toxicity (both for any CTCAE grade and CTCAE grade ≥ 3) and any SAEs, the proportion of deaths and MRD negativity will be compared between treatment groups, ALLR3 versus ALL-REZ BFM 2002 and consolidation with versus without epratuzumab, using Chi-square or Fisher's exact test as appropriate.

The following graphs show for each randomisation, the border of tolerable differences in proportions of TRM between treatment arms. The boundary is three times the standard deviation for the overall proportion of TRM based on the expected simple size at the end of recruitment. Note that despite defining expected boundary in the following graphs, sample size at the end recruitment could be different and this should be recalculated.

Figure 3 Boundary of tolerable differences in proportions of TRM between arms





For interim safety analysis, the proportion of TRM in both arms and the border of tolerable difference in proportion of TRM between arms for SR1 and SR2 will be also tabulated by country.

5.3 Interim efficacy analysis

The primary objectives of the study will be investigated. EFS will be shown using Kaplan-Meier curves overall and by stratification groups.

Efficacy will be investigated after 4 years of recruitment. For SR1 analysis, 81 EFS1 events are expected at 4 years (end recruitment) and 207 EFS1 events at the end follow-up (calculated using STATA software). Consequently, interim efficacy analysis will be conducted after 39.1% of EFS1 events (81/207 events) have occurred. EFS1 will be investigated at this analysis using a Cox analysis of treatment effect on EFS1 adjusting for the factors used in the randomization stratification. Consideration will be given to stopping the trial early if the significance level for treatment effect in the Cox analysis is < 0.00034 following the Lan de Mets alpha spending approach using the O-Brien spending function and one interim analysis at 39.1% information fraction. This does not require adjustment of the p-value at the end of the trial (calculated using the WinLD software provided by Lan and DeMets[1,2]).

For SR2 analysis, 42 EFS2 events are expected at the end recruitment and 109 at the end follow-up. Then interim efficacy analysis will be conducted after 38.5% of EFS2 events have occurred. EFS2 will be investigated at this analysis using a Cox analysis of treatment effect on EFS2 adjusting for the factors used in the randomization stratification. Consideration will be given to stopping the trial early if the significance level for treatment effect in the Cox analysis is < 0.0003 following the Lan de Mets alpha spending approach using the O-Brien spending function and one interim analysis at 38.5% information fraction. This does not require adjustment of the p-value at the end of the trial (calculated using the WinLD software provided by Lan and DeMets).

EFS1 for SR1 and EFS2 for SR2 will be considered at this analysis. Note that EFS1 for SR1 is likely to have more events. If efficacy is shown only for EFS1 for SR1 and not for EFS2 for SR2, it may be prudent to continue the trial to fully evaluate the efficacy.

6. DESCRIPTIVE ANALYSES

6.1 Representativeness of Study Sample and Patient Throughput

Patient throughput will be illustrated using CONSORT flow diagram (below). In particular numbers will be presented for all patients registered, showing numbers in each randomised arm, compliant, and analysed with respect to the primary objectives. The number of patients who are ineligible or excluded (failure to satisfy inclusion and exclusion criteria, refusal to participate) or withdrawn from treatment (with reasons) will also be recorded.

[1] Lan KKG & DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika (1982) **70**, 659-663 (1982)
[2] DeMets DL & Lan KKG. Interim Analysis: The alpha spending approach. Stats in Med (1994) **13**:1341-1352

Figure 4 Consort diagram for SR-group

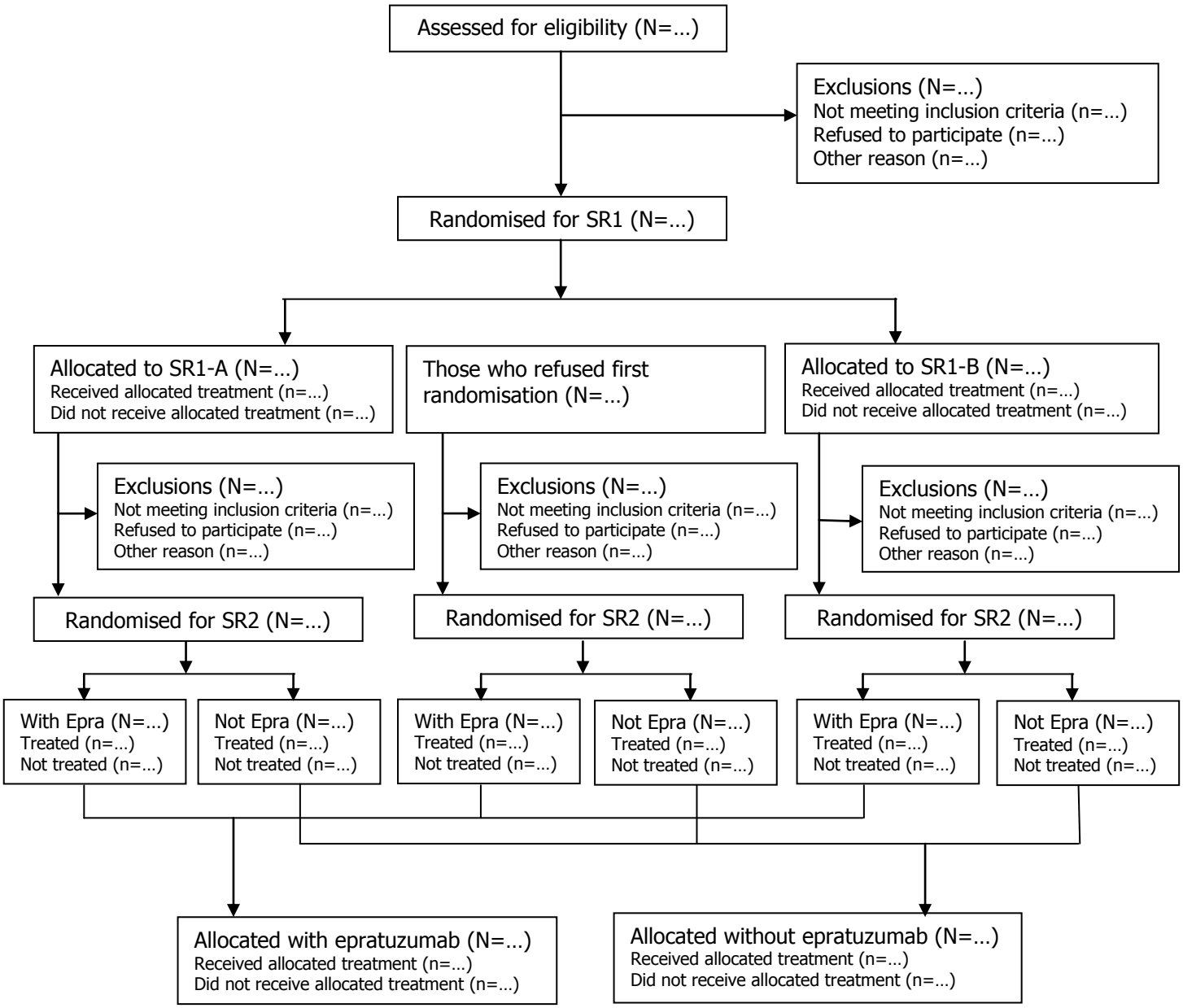


Table 9 Expansion of reasons for exclusion

Reason for exclusion	N	Subtotal
		Not meeting inclusion criteria n=
		Refused to participate n=
		Other reason n=
Total patients screened but not randomised		

6.2 Baseline Comparability of Randomised Groups

Baseline characteristics will be compared between randomised groups to check that they are similar. Characteristics to be described in tabular form are as follows:

Table 10 Baseline characteristics

Characteristics	RISK/TREATMENT GROUPS			
	SR-1		SR-2	
	SR-A	SR-B	With epratuzumab	Without epratuzumab
Variables used to define risk group				
Time-point of relapse from date of primary diagnosis				
Site of relapse				
Immunophenotype defined according to EGIL criteria				
Demographics				
Age at first diagnosis				
Age at first relapse				
Sex				
Treating country				
Disease specification				
White cell count at first diagnosis				
Age at first relapse				

Each of the characteristics shown above will be summarised by showing number and percentage for SR1 (SR-A vs SR-B) and SR2 (epratuzumab vs without). Numbers (with percentages) for binary and categorical variables and means (and standard deviations), or medians (with lower and upper quartiles) for continuous variables will be presented. The trial statistician, DSMC and Inter-Group TMG will consider if an imbalance is likely to cause concern and will agree an action accordingly. There will be no tests of statistical significance nor confidence intervals for

differences between randomised groups on any baseline variable because the treatment arms will be allocated at random, therefore any difference are due to chance variation.

The follow-up will be summarised within each treatment group for each time to event endpoint using the median from the reverse Kaplan-Meier method and the median follow-up for those without an event.

6.3 Comparison of Loses to Follow-up

The numbers and distribution of losses to follow-up (defaulters and withdrawals) over the 7 year period of the study will be reported overall and for each randomization. For each randomization, the median follow up and the number withdrawn from treatment will be reported overall and for each randomisation group.

For time to event analyses, the baseline characteristics of those lost to follow-up and those to be analysed will be summarised for each randomization. No test of significance will be used. Those randomised may be balanced, but those with data to analyse on the secondary objectives may not be. The consequences of this examination of the data will be described separately for each objective.

6.4 Description of Available Data

The patterns of availability of primary and secondary outcome data, from baseline to end of follow-up, will be summarised for the two treatment groups (SR-1 and SR-2), and for all patients.

6.5 Description of Compliance with Therapy

A summary of patient compliant, partially compliant and non-compliant will be made by randomised arm. Details will be listed of any major protocol deviations. The number of patients accepted to the trial but refusing randomisation to the two groups will be reported.

6.6 Reliability

MRD will be measured using standard European wide MRD procedures. MRD measured sequentially during therapy allows assessing the dynamics of response to treatment and can be considered as early efficacy endpoint in randomized trials. Taken together, quantification of MRD at diagnosis of ALL relapse and regularly during therapy is an essential tool to characterize the responsiveness of the disease and to allocate the patients to a risk adapted treatment. MRD quantification at validated time-point within controlled treatment strategies has in the meantime achieved the status of standard of care and is in Germany financed by health insurance system.

Survival times (EFS) will be calculated using a computer and will be checked by hand calculations for a minimum of 5% or 20 patients, where appropriate.

BMI is calculated via different formula in different countries so doses of drug received will differ by as much as 10%. This cannot be changed but must be mentioned.

6.7 Description of Country Differences

The age at first relapse, time to first relapse and site of first relapse by country will be tabulated.

6.8 Compare R3 in IntReALL with R3 in ALLR3

3 year PFS will be described overall and by country for R3 in IntReALL and R3 in ALLR3. PFS is defined as time from randomisation to the first occurrence of progression, relapse, death from any cause or second malignancy, or is censored at the date of last follow-up.

6.9 Compare BFM in IntReALL with historical BFM

3 year PFS will be described overall and by country for BFM in IntReALL and BFM as used prior to IntReALL

7. PATIENT GROUPS FOR ANALYSIS

All patients randomised will be analysed on an intention-to-treat basis.

Intention to treat population:

- ITT- SR1: all patients randomised to treatment in SR1 only.
- ITT- SR2: all patients randomised to treatment in SR2 only.

For the primary, secondary and exploratory situations, analyses will be conducted on an intention to treat (ITT) sample, defined in this case to mean all patients randomised irrespective of how much or which treatment they actually received. In secondary and exploratory analysis the ITT sample may be modified (MITT) due to missing data in the variables required for the analysis, i.e. every randomised patient who has the data for the particular analysis will be included using the treatment code to which they were randomised.

As far as possible, data from all patients will be included in analysis, with censoring to deal with loss to follow-up. Patients withdrawn from treatment will be followed up where possible and included in analysis unless consent for inclusion in analysis has been withdrawn.

As a sensitivity analysis, all analyses will be repeated on the per protocol (PP) sample which is defined here as the ITT sample without patients who did not receive sufficient trial treatment, or have extensive missing data or a protocol deviation. Inclusion of patients in the PP sample will be made on a patient by patient basis and agreed by the DSMC, but are likely to be guided as below:

- Insufficient trial treatment is defined as patients who received less than 60% of the recommended doses for induction, consolidation or maintenance therapies according to a IntReALL SR 2010 protocol
- Extensive missing data is defined as patients with missing data required for the analysis. The primary outcomes for SR1 and SR2 are time to event analysis so data should be available on all patients – not for secondary outcomes.
- Protocol deviation is defined as the following situations: withdrawal of consent, pregnancy, retroactive failure to fulfil inclusion/exclusion criteria, significant noncompliance, new medical conditions not allowing for continuation of the protocol conform treatment

The safety sample is those patients who received at least one dose of at least one drug. At each DSMC safety and toxicity data will be summarised between the patient groups and also by stratification factors within each randomisation. The SAE will be investigated by the DSMC and treatment related deaths corroborated. Any further safety or toxicity data will be provided to the DSMC upon request.

For pre-specified subgroups see section 8.4.

8. ANALYSES TO ADDRESS PRIMARY AIMS

All validated data will be stored in the database MARVIN. After termination of the study and after completion of all entries, the database will be closed for further entries. This process will be documented. The analysis will be performed using the following commercial software: SAS or SPSS or STATA with the (Program / version specified).

8.1 Statistical Methods Used for Analysis of Primary Outcome

This is expected to take place 3 years after final patient recruitment for both SR randomisations.

a) SR1 – time to an event

A Cox analysis of treatment effect on EFS1 adjusting for the factors used in the randomisation stratification to remove possible bias between arms due to chance unbalance and to give more statistical power and unbiased results.

b) SR2 – time to an event

A Cox analysis of treatment effect on EFS2 adjusting for the factors used in the randomisation stratification.

8.2 Adjustment of P values for Multiple Testing

No formal adjustment for multiple significance testing is intended.

8.3 Missing Data

The primary outcomes for SR-1 and SR-2 are time to event analysis so data should be available on all patients.

As a part of the review of the study data, recruitment, loss follow-up, missing MRD, and other relevant missing data will be checked. This information will be reviewed by the DSMC.

8.4 Pre-specified Subgroup Analysis

Within SR analyses, results will be specified for

T cell
B cell
Week 5 MRD high
Week 5 MRD low
Extramedullary
Not extramedullary
Relapse at 12 months
Relapse at 24 months
Relapse at 36 months
Age 1-<6
Age 6-<10
Age 10+
Cytology groupings
BFM countries
R3 countries
Non BFM and non R3 countries

Using a Forest plot to show the results.

8.5 Consideration of Treatment Interaction

A Kaplan-Meier plot will be drawn with a line for each of the four SR1 SR2 treatment combination groups to look for a visual sign of interaction between the treatments. An interaction term will also be added to the Cox models containing both of the treatment main effect to test for significance (although this will have low power). If statistically significant interaction is found, the results will also be presented for SR2 within each SR1 treatment group.

8.6 Treatment by Centre Interaction

Consistency of effect will be assessed across the *N* centres by informal examination of the within centre effects. There will be limited capacity to investigate these formally and it is noted that such centre effects are expected by chance.

8.7 Sensitivity Analysis

As a sensitivity analysis, all analyses will be repeated on the per protocol (PP) sample.

The results will be compared between previous BFM countries, previous R3 countries and countries previously using other regimens.

The results will be checked for a difference between ages and for differences between the genders.

9. ANALYSIS TO ADDRESS SECONDARY AIMS

These investigate other aspects of differences between treatment arms.

9.1 Evaluation/Definition of Secondary Outcomes (where applicable)

Secondary outcomes

- Improvement of OS with arm B compared to arm A
- Improvement of OS after consolidation with versus without epratuzumab
- Rate of CR2 of arm B versus arm A
- Rate of SCT of arm B versus arm A
- MRD reduction during consolidation with versus without epratuzumab
- Rate of MRD negativity prior to SCT of arm B versus arm A
- Rate of MRD negativity prior to SCT after consolidation with versus without epratuzumab
- Toxicity of induction arm B versus arm A
- Toxicity of consolidation with versus without epratuzumab
  Pharmacokinetic of epratuzumab in context with Arm A and B

9.2 Statistical Methods Used for Analysis of Secondary Outcomes

This expected to take place 3 years after final patient recruitment for both SR randomisations.

a) SR1 – time to death
A Cox analysis of treatment effect on OS adjusting for the factors used in the randomisation stratification

b) SR2 – time to death
A Cox analysis of treatment effect on OS adjusting for the factors used in the randomisation stratification

c) SR1 - Rate of CR2 at the end induction with arm B versus arm A
Comparison of rate of patients with CR2 prior to consolidation phase between treatment groups using chi-square or Fisher's exact test as appropriate

d) SR1 - Rate of SCT at the end induction with arm B versus arm A
Comparison of rate of patients with SCT prior to consolidation phase between treatment groups using chi-square or Fisher's exact test as appropriate

e) SR2 - Improvement of MRD reduction during consolidation with epratuzumab versus without
MRD levels between treatment groups will be assessed using Mann-Whitney or ANOVA as appropriate

f) SR1 - Rate of MRD negativity prior to SCT with arm B versus arm A
Comparison of rate of patients with no detectable MRD (MRD negative) prior to SCT between treatment groups using chi-square or Fisher's exact test as appropriate

g) SR2 - Rate of MRD negativity prior to SCT after consolidation with epratuzumab versus without
Comparison of rate of patients with no detectable MRD (MRD negative) prior to SCT between treatment groups using chi-square or Fisher's exact test as appropriate

h) SR1 - Toxicity of arm B versus arm A
For each of liver, GI, infection, renal, CNS and other compare count with CTCAE grade ≥3 between treatment groups using chi square and chi square for trend. Comparison of incidence of SAE, treatment related deaths and second malignancy between groups using chi-square or Fisher's exact test as appropriate. A p-value adjustment for multiple testing will be made using Bonferroni corrections to protect against making false positive conclusions

i) SR2 - Toxicity during consolidation with epratuzumab versus without
For each of liver, GI, infection, renal, CNS and other compare count with CTCAE grade ≥3 between treatment groups using chi square and chi square for trend. Comparison of incidence of SAE, treatment related deaths and second malignancy between groups using chi-square or Fisher's exact test as appropriate. A p-value adjustment for multiple testing will be made using Bonferroni corrections to protect against making false positive conclusions.

j) SR2 - Pharmacokinetic of epratuzumab in context with Arm A and B
For each pharmacokinetic parameter of epratuzumab, comparison between treatment arms using chi-square or Mann-Whitney test as appropriate

9.3 Additional Exploratory Analysis Not Specified Prior to Receiving Data

In the final publication it will be reported that additional exploratory analyses not specified in section 6 and 7 was not pre - specified. The rationale for carrying out further analyses will be justified.

10. SERIOUS ADVERSE EVENTS

For definition of a SAE refer to protocol section 14.1.2.

All Serious Adverse Events (SAEs) and all Adverse Events (AEs) need to be documented by the investigator, no matter if the Investigator suspects a causal connection to the investigational product or not. The documentation needs to include the type of event, start, duration, severity and causality. Related signs symptoms and laboratory changes should be summarized to a specific disease. The event will be recorded in the electronic CRF. SAEs need to be documented on a separate SAE form. Out of normal range laboratory data need to be analysed concerning their clinical relevance by the Investigator – and if relevant documented as an AE itself. All adverse events need to be followed until they subside or stabilize. The Sponsor will carefully document all Adverse Events (AEs) reported by the Investigator. His documentation will be sent to the relevant regulatory authorities and to relevant authorities of other European member states and other contracting states of the EWR agreement, if the study is run in their territory and if they so request. The Data Safety Monitoring Committee (DSMC) will look at SAEs periodically.

11. ADVERSE EVENT

An adverse event is any untoward medical occurrence in a patient or clinical trial subject administered a medicinal product and which does not necessarily have a causal relationship with this treatment. An AE can therefore be any unfavorable and unintended sign (including abnormal laboratory findings for example), symptom, or disease temporally associated with the use of a medicinal (investigational) or medical product, whether or not it is considered to be related to the medicinal (investigational) product. (ICH-Guideline E2A) Adverse events encompass illness, signs of illness (including pathological laboratory findings) and symptoms that initiate during the trial or previous conditions that become worse. An AE could be diseases, signs or symptoms which occur or worsen after enrolment of the patient in the clinical trial

12. DOCUMENT HISTORY

Ver	Date	Who	Comments
0.1	30Sep2011	CC	First version
0.2	24Jan2012	CC	Update based on protocol 5Dec11
0.3	3 Feb2012	CC	Update based on SBL comments
1.0	8 Feb2012	CC	Adding section 5.1 Early stopping guidelines for TRM & change TSC by Inter-Group TSC
1.1	28Feb2012	CC	Updated sections 5.1 and 5.2. Inclusion of tables and graphs showing the border of tolerable difference in proportion of TRM Clarified sentence in section 5.1 defining TRM
1.2	12Feb2012	CC	Updated based on Vaskar and Arend comments about SAP 3Feb12 and teleconference – There is a document that includes all their comments and changes done according with their comments on I drive analysis folder of IntReALL
1.3	27Mar2012	CC	Updated based on SBL comments about SAP 12Feb12
2.0	21May2012	CC	Updated based on protocol version 1.6 10May12

2.1	20Jun2012	CC	Updated based on protocol version 1.6 30May12
2.2	28Jun2012	CC	Updated based on SBL comments
3.1	4Jul2012	CC	New exclusion criteria added based on protocol V1.7 1Jul12
3.2	12Jul2012	CC	Changes related to M0, M1, M2 and M3 bone marrow status

13. APPENDIX 1 – STATA OUTPUT FOR SAMPLE SIZES

13.1 SR1

STATA version 11

*ALLREZ 65% ALLR3 75% n= 612 to get enough in SR2 power ??% 4r3fu considered Feb2011
. artsurv, alpha(0.05) detail(1) edf0(0.65,4) hratio(1,0.667) nperiod(7) n(612) recrt(4 0,0.166 0.166 0.333 0.333, 0) tunit(1) ldf(0.05, 7; 0.05, 7)

ART - ANALYSIS OF RESOURCES FOR TRIALS (version 1.0.4, 13 January 2005)

A sample size program by Abdel Babiker, Patrick Royston & Friederike Barthel,
MRC Clinical Trials Unit, London NW1 2DA, UK.

Type of trial	Superiority - time-to-event outcome
Statistical test assumed	Unweighted logrank test (local)
Number of groups	2
Allocation ratio	Equal group sizes
Total number of periods	7
Length of each period	One year
Survival probs per period (group 1)	0.898 0.806 0.724 0.650 0.584 0.524 0.471
Survival probs per period (group 2)	0.931 0.866 0.806 0.750 0.698 0.650 0.605
Number of recruitment periods	4
Number of follow-up periods	3
Method of accrual	Uniform
Recruitment period-weights	0.166 0.166 0.333 0.333 0 0 0
Hazard ratios as entered (groups 1,2)	1, 0.667
Hazard ratios per period (group 1)	1.000 1.000 1.000 1.000 1.000 1.000 1.000
Hazard ratios per period (group 2)	0.667 0.667 0.667 0.667 0.667 0.667 0.667
Alpha	0.050 (two-sided)
Power (calculated)	0.823
Total sample size (designed)	612
Expected total number of events	207

Adding Japan

. artsurv, alpha(0.05) detail(1) edf0(0.65,4) hratio(1,0.71) nperiod(7) power(0.8) recrt(4 0,0.166 0.166 0.333 0.333, 0) tunit(1) ldf(0.05, 7; 0.05, 7)

ART - ANALYSIS OF RESOURCES FOR TRIALS (version 1.0.4, 13 January 2005)

A sample size program by Abdel Babiker, Patrick Royston & Friederike Barthel,
MRC Clinical Trials Unit, London NW1 2DA, UK.

Type of trial	Superiority - time-to-event outcome
Statistical test assumed	Unweighted logrank test (local)
Number of groups	2
Allocation ratio	Equal group sizes
Total number of periods	7
Length of each period	One year
Survival probs per period (group 1)	0.898 0.806 0.724 0.650 0.584 0.524 0.471
Survival probs per period (group 2)	0.926 0.858 0.795 0.736 0.682 0.632 0.586
Number of recruitment periods	4
Number of follow-up periods	3
Method of accrual	Uniform
Recruitment period-weights	0.166 0.166 0.333 0.333 0 0 0
Hazard ratios as entered (groups 1,2)	1, 0.71
Hazard ratios per period (group 1)	1.000 1.000 1.000 1.000 1.000 1.000 1.000
Hazard ratios per period (group 2)	0.710 0.710 0.710 0.710 0.710 0.710 0.710
Alpha	0.050 (two-sided)
Power (designed)	0.800
Total sample size (calculated)	785
Expected total number of events	271

13.2 SR2

. artsurv, alpha(0.05) detail(1) edf0(0.74,4) hratio(1,0.5790) nperiod(7) power(0.8) recrt(4 0,0.166 0.166 0.333 0.333, 0) tunit(1) ldf(0.05, 7; 0.05, 7)

ART - ANALYSIS OF RESOURCES FOR TRIALS (version 1.0.7, 19 October 2009)

A sample size program by Abdel Babiker, Patrick Royston & Friederike Barthel,
MRC Clinical Trials Unit, London NW1 2DA, UK.

Type of trial	Superiority - time-to-event outcome
Statistical test assumed	Unweighted logrank test (local)
Number of groups	2
Allocation ratio	Equal group sizes
Total number of periods	7
Length of each period	One year
Survival probs per period (group 1)	0.927 0.860 0.798 0.740 0.686 0.637 0.590
Survival probs per period (group 2)	0.957 0.917 0.877 0.840 0.804 0.770 0.737
Number of recruitment periods	4
Number of follow-up periods	3
Method of accrual	Uniform
Recruitment period-weights	0.166 0.166 0.333 0.333 0 0 0
Hazard ratios as entered (groups 1,2)	1, 0.5790
Hazard ratios per period (group 1)	1.000 1.000 1.000 1.000 1.000 1.000 1.000
Hazard ratios per period (group 2)	0.579 0.579 0.579 0.579 0.579 0.579 0.579
Alpha	0.050 (two-sided)
Power (designed)	0.800
Total sample size (calculated)	455
Expected total number of events	109