A validated novel continuous prognostic index to deliver stratified medicine in pediatric acute lymphoblastic leukemia

Amir Enshaei, PhD[1], David O'Connor, MD[2,13], Jack Bartram, MD[2], Jeremy Hancock, PhD[3], Christine J. Harrison, PhD[1], Rachael Hough, MD[4], Sujith Samarasinghe, MD[2], Monique L. den Boer, PhD[5,6], Judith M. Boer, PhD[5], Hester A. de Groot-Kruseman, PhD[6], Hanne Vibeke Marquart, MD[7], Ulrika Noren-Nystrom, MD[8], Kjeld Schmiegelow, MD[9], Claire Schwab, PhD[1], Martin A. Horstmann, MD[10], Gabriele Escherich, MD[10], Mats Heyman, MD[11], Rob Pieters, MD[5], Ajay Vora, MD[2], John Moppett, MD[12], Anthony V Moorman, PhD[1]

1. Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne, UK
2. Department of Haematology, Great Ormond Street Hospital, London, UK
3. Bristol Genetics Laboratory, North Bristol NHS Trust, Bristol, UK
4. Department of Haematology, University College Hospital, London, UK
5. Princess Máxima Center for Paediatric Oncology, Utrecht, Netherlands
6. Dutch Childhood Oncology Group, Utrecht, Netherlands
7. Dept. Clinical Immunology, Copenhagen University Hospital, Copenhagen, Denmark
8. Dept. of Clinical Sciences, Paediatrics, Umeå University, UMEÅ, Sweden
9. Department of Paediatrics and Adolescent Medicine, University Hospital Rigshospitalet, Copenhagen, Denmark
10. Dept. of Paediatric Hematology and Oncology, University Medical Center, Hamburg, Germany
11. Department of Paediatric Oncology, Karolinska University Hospital and Karolinska Institutet, Stockholm, Sweden,
12. Department of Haematology, Royal Hospital for Sick Children, Bristol, UK
13. Department of Haematology, UCL Cancer Institute, London, UK

Correspondence: Professor Anthony V Moorman, Leukaemia Research Cytogenetics Group, Wolfson Childhood Cancer Research Centre, Northern Institute for Cancer Research, Newcastle University, Level 6, Herschel Building, Brewery Lane, Newcastle upon Tyne, NE1 7RU
anthony.moorman@ncl.ac.uk

Short title: A validated novel prognostic index in ALL

Key words: minimal residual disease, genetics, risk stratification, risk groups, prognosis, outcome

Words: 3934 words; Abstract: 250 words; Tables/Figures: 5; References: 23

Key points

- We have developed and validated a prognostic index that assigns patient-specific risk scores and defines clinically relevant risk groups.
- The prognostic index, PI_{UKALL}, integrates existing risk factors and leverages continuous data to out-perform existing risk algorithms.

Abstract

Risk stratification is essential for the delivery of optimal treatment in childhood acute lymphoblastic leukemia. However, current risk stratification algorithms dichotomise variables and apply risk factors independently which may wrongly assume identical associations across biologically heterogeneous subsets and reduce statistical power. Accordingly, we developed and validated a prognostic index (PI_{UKALL}) which integrates multiple risk factors and uses continuous data. We created discovery (n=2,405) and validation (n=2,313) cohorts using data from four recent trials (UKALL2003, COALL-03, DCOG-ALL10, NOPHO-ALL2008). Using the discovery cohort, multivariate Cox regression modelling defined a minimal model that included white cell count at diagnosis, pre-treatment cytogenetics and end of induction minimal residual disease. Using this model we defined PI_{UKALL} - a continuous variable that assigns personalised risk scores. The PI_{UKALL} correlated with risk of relapse and validated in an independent cohort. Using PI_{UKALL} to risk stratify patients improved the C-index for all endpoints compared to the traditional algorithms. We used PI_{UKALL} to define four clinically relevant risk groups which had differential but similar relapse rates at 5 years in the discovery and validation cohorts respectively: low 3% (95% CI 2-4)/4%(3-6); standard 8%(6-10)/9%(6-12); intermediate 17%(14-21)/17%(14-21) and high 48%(36-60)/35%(24-48). An analysis of the area under the curve confirmed the risk groups were significantly better at predicting outcome than the algorithms employed in each trial. The PI_{UKALL} developed in this study provides an accurate method for predicting outcome and a more flexible method for defining risk groups in future studies. Personalised risk scores can facilitate the design of future risk algorithms.

Introduction

Accurate risk stratification is essential for the delivery of optimal treatment in pediatric acute lymphoblastic leukemia (ALL). Experimental therapeutic approaches are needed to improve cure rates for high-risk (HR) patients. Conversely, treatment de-intensification to reduce long-term toxicity, can only be justified for patient subgroups with a very low relapse risk. Minimal residual disease (MRD) during the first month of therapy is the most powerful prognostic factor in both pediatric and adult ALL and can be used to guide both therapy intensification and reduction.[1-3] However, MRD alone is not sufficient to fully predict outcome. We have recently shown that the prognostic effect of MRD differs significantly according to the genetic make-up of the leukemic clone.[4] Other patient- and disease-specific characteristics, including age and white cell count (WCC), have also been shown to independently influence outcome.[5]

The multitude of risk factors in pediatric ALL poses significant challenges to the development of risk algorithms. Risk factors have been used in different ways that has hindered the direct comparison of cure rates. Crucially, the requirement for simple clinical stratification has driven the use of categorical thresholds of continuous variables. However, dichotomisation of continuous variables leads to significant loss of statistical power.[6] Moreover, categorising continuous variables that are unevenly distributed produces risk groups of unequal and fixed size. This approach reduces flexibility when defining treatment groups by both size and relapse risk when designing clinical trials.

We recently analysed MRD data as a continuous variable for the first time in pediatric ALL and demonstrated that at the end of induction (EOI) disease levels were log normally distributed and that each log reduction in disease burden achieved by EOI decreased the risk of relapse by 20%.[4] In addition, a meta-analysis of 39 MRD studies concluded that achieving MRD negativity (<0.01%) by the EOI reduced a patients risk of relapse four-fold.[2] These results are consistent with one another and are both clinically important.

In this study, we use continuous data from more than 4,700 patients across four large international contemporaneous trials to build and validate an integrated prognostic index which enhances predictive power in pediatric ALL.

Methods

Study Participants, Treatment and Oversight

Individual patient data used in this post-hoc analysis was derived from patients who consented to treatment on UKALL2003 (ISCTRN 07355119), Nordic Society of Paediatric Haematology and Oncology (NOPHO) ALL2008 (Eudract 2008-003235-02)[7], Dutch Children's Oncology Group (DCOG)-ALL10 or German Co-operative Study Group (CoALL)-07-03. Full details of the recruitment, treatment and outcome have been published: UKALL2003[1,3], NOPH0-ALL2008, DOCG-ALL10[8] and CoALL-07-03[9]. All four protocols excluded infants (<1 year old) but had variable upper age limits: 18 years (DCOG-ALL10, CoALL-07-03), 24 years (UKALL2003), and 45 years (NOPHO-ALL2008). Each protocol risk stratified patients into two or three risk groups based on a combination of risk factors that included age, WCC, genetics and MRD (Table S1). Each trial was approved by the relevant ethics committee and patients or parents gave written informed consent in accordance with the declaration of Helsinki.

Minimal residual disease (MRD) and genetic studies

MRD was evaluated by PCR analysis of Ig/TCR rearrangements (UKALL2003, DCOG-ALL10 and CoALL-07-03) or flow cytometry using six-colour MRD panels to detect leukemia-associated immunophenotypes (NOPHO-ALL2008). To examine MRD as a continuous variable, we log transformed the raw MRD value calculated at EOI, τ(MRD).[4] Patients with undetectable MRD were assigned a value of $1x10^{-6}$ (one log below the minimum detection level of $1x10^{-5}$). MRD values <$1x10^{-5}$ were rounded up to $1x10^{-5}$ while values ≥1 were rounded down to 0.99999.

For the discovery cohort, pre-treatment cytogenetic and immunophenotyping analysis was used to classify patients into four mutually exclusive subtypes: (1) cytogenetic good risk (CYTO-GR), *ETV6-RUNX1*, high hyperdiploidy 51-67 chromosomes (HeH); (2) cytogenetic high risk (CYTO-HR), *KMT2A/MLL* fusions, near-haploidy (<30 chromosomes), low hypodiploidy (30-39 chromosomes), intrachromosomal amplification of chromosome 21q (iAMP21) and t(17;19)(q23;p13)/*TCF3-HLF*; (3) cytogenetic intermediate risk (CYTO-IR): t(1;19)(q23;p13)/*TCF3-PBX1* and B-other; and (4) T-ALL.[10] For the validation cohort, we collected the data required to calculate the prognostic index, i.e. the presence or absence of good and HR cytogenetics. Copy number data derived from MLPA analysis using the P335 SALS kit (MRC Holland) was available for UKALL2003 and DCOG-ALL10 and was analysed and coded as previously described.[11,12]

Eligibility criteria, endpoints and statistical analysis

Figure 1 provides details of the cases included in this analysis. To enable meaningful cross cohort comparison we applied multiple exclusion criteria. The excluded cohort was enriched, by definition,

for HR patients but overall the analysed cohort was representative of the vast majority of pediatric and adolescent ALL (Table S2).

Survival analysis considered three endpoints. Event-free survival (EFS) was defined as time to relapse, second tumour or death, censoring at date of last contact. Relapse rate (RR) was defined as time to relapse for those achieving a complete remission, censoring at date of death in remission or last contact. Overall survival (OS) was defined as time to death, censoring at date of last contact. Patients who relapsed were classified as having a standard or high-risk relapse. Standard risk relapses comprised (a) late (>6 months after stopping frontline therapy) isolated extra-medullary (EM) relapses; (b) BCP-ALL late relapses involving the bone marrow (BM) or early (<6 months from stopping frontline therapy) isolated EM and combined relapses and (c) T-ALL patients with early isolated EM relapses. HR relapses comprised (a) patients with a very early relapse (<18 months from initial diagnosis); (b) all patients with HR cytogenetics; (c) T-ALL relapses involving the marrow and (d) BCP-ALL patients with an early isolated BM relapses.[13]

Univariate Cox regression analysis was used to estimate the risk of relapse associated with individual risk factors. Multivariate Cox regression analysis was used to build a model for predicting relapse. We used two modelling strategies: (a) forward selection - adding each variable to the model (according to the univariate hazard ratio and p value) and only retaining variables if they improved the fit of the model; (b) backward selection - all variables started off in the model with non-significant variables removed according to their p value and checking that their removal did not reduce the fit of the model. Models were compared using the likelihood ratio test and a threshold of p=0.05 was applied to retain or exclude individual variables. The proportionality assumption of the models were assessed by visualising the log-log plot of survival, the Kaplan–Meier and predicted survival plot and tested using Schoenfeld residuals. The final model was internally validated using cross-validation techniques (100 repeats of a random 70% selection) and bootstrapping (1000-fold).[14] The fit of the final model was assessed using Harrell's c-index. The discrimination, calibration and fit of the model was validated using the principles and methods described by Royston and Altman.[15] The model was calibrated by comparing the predicted and observed even probability. Forest plots and the test of heterogeneity were used to examine hazard ratios across different patient subgroups or cohorts. The area under the ROC (receiver operator characteristic) curve was used to compare the predictive power of the prognostic index and the original trial risk groups. To identify the thresholds for the exemplar risk groups, we sorted the prognostic index, divided the cohort into bins comprising 25 cases (~1% cohort) and sequentially tested each threshold until the exemplar clinical criteria were met. Due to the

investigative nature of this analysis, all tests were conducted at the 1% significance level. All analyses were performed using Intercooled Stata 13.0 (Stata Corporation, USA).

Results

Development of the prognostic index using the discovery cohort

Univariate Cox regression analysis of 2,405 patients treated on UKALL2003 revealed all major risk factors were associated with significant increases or decreases in the risk of relapse (Table 1). Next, we performed multivariate Cox regression modelling to identify the minimum number of independent variables required to predict relapse. The final model comprised τ(MRD), WCC and genetics (Table 1). None of the other variables considered improved the ability of this model to predict relapse. Using the coefficients from this model (Table 1), we derived a linear model (Figure 2A) from which we calculated patient-specific risk scores. This prognostic index (PI_{UKALL}) was directly associated with risk of relapse (Figure 2B). Univariate models of the PI as a linear variable gave hazard ratios of 2.5-3.2 for EFS, RR and OS (Figure 2D). Sensitivity analyses revealed that these hazard ratios were consistent across all major patient and treatment subgroups, including T-ALL, illustrating the robustness of PI_{UKALL} to predict outcome independently of other risk factors and at different treatment intensities (Figure S1).

Validation of the prognostic index

PI_{UKALL} was validated using 2,313 patients derived from three contemporaneous clinical trials with equivalent baseline characteristics and outcomes (Figure 1, Table S2, Figure S2). The distribution of EOI MRD was significantly different across the trials (Figure S3) reflecting the different induction regimens (Table S1). We calculated PI_{UKALL} scores for each patient in the validation cohort using the same linear model (Figure 2A) and observed equivalent distributions in the combined validation cohort and individual datasets despite differences in MRD methodology and EOI distributions (Figure 2B, 2C, S4). As in the discovery cohort, a rising PI_{UKALL} was associated with relapse and each unit increase produced comparable hazard ratios for all three endpoints considered (Figure 2D) which were stable across patient and treatment subgroups (Figure S5). Further validation tests confirmed the ability of the PI_{UKALL} to predict outcome in both low and HR patients (Figure S6) and that each component of the prognostic index contributed equivalently in the individual validation datasets (Figure S7).

Using the PI_{UKALL} as a linear variable resulted in significantly improved C-indexes compared to the standard risk groups (Table 2). Furthermore, we used PI_{UKALL} to define comparable risk groups, in terms

of number and size, for NOPHO-ALL2008 and DOCG-ALL10 patients (n=2,053) (Table S3). Using the PI_{UKALL} defined risk groups would have resulted in 762 (37%) patients being assigned to a different risk group, with 384 (19%) assigned more treatment and 378 (18%) less therapy. Importantly, the outcome of the patients who would have moved risk groups fitted more closely with the PI_{UKALL} defined risk group than the original risk groups (Table S3).

Clinical benefit of using the prognostic index in protocol design

To explore the usefulness of PI_{UKALL} to define novel clinically meaningful risk groups, we used a scenario whereby a hypothetical new trial required patients to be assigned to 4 risk groups. The criteria for the groups were: (1) a low risk (LR) group comprising ~50% cases, with a RR <5% and OS ~98% which could be considered for treatment de-intensification; (2) a HR group comprising ~5% cases, with a RR >40% which could be considered for experimental therapy; (3) equal-sized standard (SR) and intermediate (IR) risk groups with RR </>10% respectively which could be randomised to novel agents or schedules. As PI_{UKALL} is a continuous variable, thresholds that define subgroups of the required size and outcome were readily identifiable (Figure 3). Importantly, applying the same thresholds to the validation cohort produced subgroups of near identical size and outcome (Figure 3).

To demonstrate how this novel PI_{UKALL} driven system could have improved the risk classification of patients in UKALL2003 we compared the distribution and outcome of patients using the two systems, (Figure S8). There was a strong correlation between the original and PI_{UKALL} driven classifications; which was expected because they use the same underlying risk factors. However, the PI_{UKALL} classification offered greater granularity. In particular, there were 229 (12%) patients treated on lower intensity regimens (A/B) which the PI_{UKALL} identified as IR/HR. These patients had a higher RR compared with those patients classified as LR/SR (4% v 21%, p<0.0001). In contrast, the RR of the 250 (45%) patients treated on regimen C, but identified by PI_{UKALL} as LR/SR, was significantly lower than the remaining regimen C patients (6% v 21%, p<0.0001). The RR in the four PI_{UKALL} defined risk groups was clearly distinct, rising from 3% to 48% in the discovery cohort (Figures 3). Examining the distribution of relapses also showed significant benefit for the PI_{UKALL}, with the LR group accounting for 55% cases but only 25% relapses, significantly better than regimen A which accounted for 51% cases and 36% relapses (p=0.014). Clearly the PI_{UKALL} HR group was highly significantly enriched for relapses (Figure S8) but it was striking that the IR group, although slightly smaller than regimen C (19% v 23%), captured the same proportion of relapses (38% v 38%). Patients with SR relapses (supplementary methods) have a better outcome than patients with HR relapse.[13] Hence it is

noteworthy that proportion of relapses that were HR relapses differed across the four PI_{UKALL} risk groups: LR 4/54 (7%), SR 19/46 (41%), 41/82 (50%), 26/31 (84%), p<0.0001 (Figure S8).

The risk stratification algorithms used by each trial in the validation cohort were different (Table S1) and the distribution of cases across the SR, IR and HR groups was 45%, 46%, 9% which is different to UKALL2003. Accordingly, there was a very strong correlation between the original and PI_{UKALL} defined HR groups (Figure S9). In this scenario, the benefit of the PI_{UKALL} defined risk groups was shown most clearly within the IR group that comprised 46% patients and had a 8% RR. PI_{UKALL} identified 398 (42%) patients with a significantly lower RR (4%, p=0.04), 305 (32%) patients with a significantly higher RR (13%, p<0.001) and 18 (2%) patients with a much higher RR (47%, p<0.001). As in the discovery cohort, there was a strong relationship between PI_{UKALL} risk group and the percentage of relapses classified as HR: LR 11/52 (21%), SR 17/37 (46%), 41/82 (62%), 21/24 (88%), p<0.0001 (Figure S9).

The current UK trial, UKALL2011, uses EOI MRD and HR cytogenetics to assign patients to treatment on regimen C. Applying these risk criteria to the UKALL2003 cohort did result in a stronger correlation with the PI_{UKALL} driven risk groups (Figure S10). In this scenario, the advantage of PI_{UKALL} system was the identification of 198 (17%) and 428 (37%) who have low PI_{UKALL} scores and RR of 2% and 7% respectively. Thus while the UKALL2011 criteria captured 73% relapses in the HR group it was at a cost of assigning 48% patients to more intensive chemotherapy.

Impact of the prognostic index in special patient subgroups

Stem cell transplant (SCT) is an important treatment option for HR patients but carries a significant risk of treatment related mortality. The criteria used to select patients for SCT in first remission differed by trial; so we excluded these patients from the cohort used to develop the PI_{UKALL} (Figure 1, Table S1). To assess whether the PI_{UKALL} could reliably identify these HR patients despite their omission from the discovery cohort, we retrospectively calculated the PI_{UKALL} for these 235 patients. We found that 134 (57%) patients had PI_{UKALL} values that assigned them to the HR group, 83 (35%) to the IR group and just 8% to the LR and SR groups combined. This was different to the overall distribution of cases across these four subgroups: 3%, 20%, 22% 55% respectively (p<0.0001). Interestingly when we examined each trial separately, we observed that SCT patients assigned by PI_{UKALL} to the IR group had significantly or borderline better OS than SCT patients assigned to the HR group: UKALL2003 87% (95% CI 83-89) v 81% (77-83), p=0.02; DCOG-ALL10 86% (77-92) v 80% (72-85), p=0.09; NOPHO-ALL2008 86% (82-89) v 67% (59-74), p<0.001, respectively.

During the development of PI_{UKALL} we considered the seven canonical chromosomal abnormalities in pediatric ALL. In order to examine the impact of PI_{UKALL} in the context of newly defined genomic abnormalities, we calculated the PI_{UKALL} for patients treated on UKALL2003/DCOG-ALL10 harbouring an ABL-class fusion, *IKZF1* deletion, *CRLF2* rearrangement and according to the UKALL-CNA profile.[11,12] A total of 29 patients with an ABL-class fusion were identified and these patients were unevenly distributed across the four risk groups: LR:SR:IR:HR 1:1:5:22. In keeping with previous observations[16], >50% (15/27) ABL-class patients classified in the IR/HR groups suffered an adverse event within 5 years. In contrast, when we calculated PI_{UKALL} values for the patients with an *IKZF1* deletion or *CRLF2* gene rearrangement, they were more evenly distributed across the four risk groups: LR/SR:IR/HR 63%:37% and 57%:43% respectively. Patients with an *IKZF1* deletion who were assigned by PI_{UKALL} to the IR/HR groups had a significantly inferior outcome (Table S3). As expected UKALL-CNA good risk patients were more likely to be assigned to the lower risk groups compared with the UKALL-CNA poor risk patients (p=0.001) (Table S3). For both UKALL-CNA good and poor risk patients, there was a significant difference in outcome when stratified by PI_{UKALL} defined risk groups (Table S3).

Discussion

We have developed and validated a prognostic index, PI_{UKALL}, which uses four weighted variables representing disease burden, treatment response and genetics. The key feature of the index is the use of continuous data for WCC and MRD which outputs patient specific rather subgroup specific risks. One of the major strengths of the index is that it was developed and validated using large, well-annotated cohorts of patients treated on modern protocols. While all four trials produced equivalent outcomes, they did so using different risk stratification algorithms, MRD methodologies and treatment regimens. This variation demonstrates the robustness of PI_{UKALL} and widespread clinical applicability.

The key question for any novel prognostic marker or system relates to its clinical impact and deliverability. We have demonstrated that using PI_{UKALL} is better than the current algorithms despite using fewer variables. Using PI_{UKALL} does not require any new variables or data; it simply uses existing information more efficiently. PI_{UKALL} is a continuous variable, so can define the number and size of risk groups that match the treatment options or randomisations being considered; rather than the other way round. This is a significant advantage over traditional systems as well as newly described integrated risk scores.[17] The validation of the exemplar risk groups in an independent cohort (figure 3) illustrate that PI_{UKALL} can be implemented without further development. PI_{UKALL} has been designed

to assist with the allocation of patients to risk groups at the EOI and does not preclude the reallocation of patients at other time-points in light of additional information, e.g. Downs syndrome, refractory disease or persistent MRD. PI_{UKALL} is flexible and can be used to define all risk groups or to split as pre-existing IR group; as illustrated in the validation cohort (Figure S9) where PI_{UKALL} can identify subsets of this group that have very different outcomes. So like other risk factors PI_{UKALL} is best employed in conjunction with other decision-making tools. In addition, a strategy for dealing with missing data would be required. Here PI_{UKALL} has the advantage that only a small number of variables are required for its implementation and, importantly, all the variables are already assessed in most modern protocols; so no new tests are required. Novel strategies for improving MRD detection and the advent of genomic technologies will minimise the number of patients with missing MRD and genetic data.[18,19] Hence PI_{UKALL} can be used now to improve the allocation of patients to risk groups as well as providing a flexible method for designing a trial with more than the traditional number of risk groups.

Improvement in the outcome for low risk patients must focus primarily on reducing treatment-related mortality, which accounts for almost half of the deaths in this group.[20] Therefore it is essential that such patients are identified early and treated on low intensity protocols to reduce mortality and morbidity.[21] Using PI_{UKALL}, we have demonstrated that it is feasible to define a LR group with a relapse rate of <5%. The advent of highly effective novel therapies, such as CAR-T cell therapy, provides the exciting possibility of cure in very HR patients.[22] However, the widespread use of such therapies will be limited by cost and complexity, thus it is essential that they are used to treat the most appropriate patients. Current classifications can struggle to define clinically useful HRHR groups. For example, UKALL2011 regimen C captures a very high percentage of relapses but it comprises nearly 50% of patients and has an overall relapse risk of 13%.

PI_{UKALL} can be used to define two clinically useful higher risk groups: (1) the IR group which comprises ~20% cases, captures ~40% relapses and has a RR of ~15-20% and could be suitable for novel drugs; and (2) a small HR with extremely poor outcome that could be used to assign patients to more experimental therapies. Crucially, given the recent increase in novel therapies, it allows the selection of specific patient risk groups for the precise allocation of treatment. All retrospective studies proposing new risk factors or prognostic indices are limited by the fact that the patients were treated according to different criteria. Identifying risk factors associated with HR of relapse among patients treated on lower intensity protocols is relatively straightforward. However, the reverse is more complicated. We have presented data suggesting that some patients treated according to UKALL2003 regimen C (a high intensity protocol) have a low risk of relapse and therefore should be prospectively

assigned to a LR or SR group. Whilst these patients could be genuine low risk patients, it is also possible that they only had a low risk of relapse because they received more intensive therapy. Retrospective studies cannot distinguish between the two scenarios. However, there is indirect evidence to support our assertion that they are truly low risk patients. Firstly, 72/82 (88%) patients treated on UKALL2003 regimen C and classified into the LR group had a good risk chromosomal abnormality - *ETV6-RUNX1* or high hyperdiploidy. Patients with good risk chromosomal abnormalities have excellent outcomes despite moderate levels of MRD after induction.[4] Secondly, the difference in relapse rate between UKALL2003 regimen C treated patients in the LR and HR groups is substantial: 4% to 43%. Whilst treatment intensification has been shown to reduce relapse risk, no one has ever reported such a large drop in relapse rate.

Even though PI_{UKALL} was based purely on MRD, WCC and a small selection of genetic abnormalities, sensitivity analyses demonstrated that it is effective at predicting outcome in all major patient subsets including T-ALL (Figure S1, S5). Developing and validating prognostic indices requires large uniformly annotated cohorts with extensive follow-up. We were only able to consider the seven canonical chromosomal abnormalities in pediatric ALL. Thus, one limitation of the PI_{UKALL} is that newly defined high and low risk abnormalities will not receive any weighting within the model. However, many HR genetic abnormalities correlate with WCC and MRD[23], so are likely to have high PI_{UKALL} values based on these risk factors alone. When we examined the distribution and outcome of patients with ABL-class fusions and key copy number alterations, we observed a strong correlation with PI_{UKALL} defined risk groups but also evidence of the additional predictive power associated with applying a multivariate rather than a univariate risk model. Nevertheless, it is likely that in the future when comprehensive screening of large cohorts becomes feasible, re-calibration of the index incorporating additional genomic and genetic data will improve its accuracy. The fact that the PI_{UKALL} does not rely on expensive genomic analyses means that it can be employed in a wide range of countries including those with more limited resources.

In conclusion, we have integrated multiple variables, including continuous data, into a single numeric PI that validated in independent datasets. PI_{UKALL} allocates individual risk scores that allow the accurate selection of patients with an explicit risk of relapse for the precise allocation of treatment. This novel approach to risk stratification offers clear benefits over current algorithms and because it uses the same information used for existing algorithms it can be adopted immediately. This study demonstrates that the future of risk stratification in ALL lies in integrating all known risk factors and utilizing all the available data with continuous variables.

Acknowledgments

We would like thank all the patients who took part in this trial as well as their families. We acknowledge Bloodwise (formerly Leukaemia & Lymphoma Research) and Children with Cancer for financial support and the member laboratories of the UK Cancer Cytogenetic Group (UKCCG) for providing cytogenetic data and material. We acknowledge the input of all the scientists and technicians working in the MRD laboratories: Bristol Genetics Laboratory, Southmead Hospital, Bristol; Molecular Biology Laboratory, Royal Hospital for Sick Children, Glasgow; Molecular Haematology Laboratory, Royal London Hospital, London; and the Molecular Genetics Service, Sheffield Children's Hospital, Sheffield; University Medical Center Rotterdam, Rotterdam; Sanquin Research, Amsterdam. Primary childhood leukemia samples used in this study were provided by the Bloodwise Childhood Leukaemia Cell Bank working with the aforementioned MRD laboratories.

Author contributions

Conception and Design: Anthony V Moorman, Amir Enshaei, David O'Connor, John Moppett

Collection and assembly of data: Anthony V Moorman, Jack Bartram, Jeremy Hancock, Christine J Harrison, Claire Schwab, Monique L. den Boer, Judith M. Boer, Hester A. de Groot-Kruseman, Hanne Vibeke Marquart, Ulrika Noren-Nyström, John Moppett

Data analysis and interpretation: Amir Enshaei, Anthony V Moorman, David O'Connor, John Moppett,

Financial support: Christine J Harrison, Anthony V Moorman

Administrative support: Christine J Harrison, Anthony V Moorman

Provision of study materials or patients: Rachael Hough, Sujith Samarasinghe, Hanne Vibeke Marquart, Ulrika Noren-Nystrom, Kjeld Schmiegelow, Martin Horstmann, Gabriele Escherich, Mats Heyman, Rob Pieters, Ajay Vora, John Moppett

Manuscript writing: Anthony V Moorman, Amir Enshaei

Final approval of manuscript: All authors

Conflicts of interests

None of the authors have any conflicts of interest to disclose.

References

1. 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 14:199-209, 2013
2. Berry DA, Zhou S, Higley H, et al: Association of Minimal Residual Disease With Clinical Outcome in Pediatric and Adult Acute Lymphoblastic Leukemia: A Meta-analysis. JAMA Oncol 3:e170580, 2017
3. Vora A, Goulden N, Mitchell C, et al: Augmented post-remission therapy for a minimal residual disease-defined high-risk subgroup of children and young people with clinical standard-risk and intermediate-risk acute lymphoblastic leukaemia (UKALL 2003): a randomised controlled trial. Lancet Oncol 15:809-18, 2014
4. O'Connor D, Enshaei A, Bartram J, et al: Genotype-Specific Minimal Residual Disease Interpretation Improves Stratification in Pediatric Acute Lymphoblastic Leukemia. J Clin Oncol 36:34-43, 2018
5. Hunger SP, Mullighan CG: Acute Lymphoblastic Leukemia in Children. N Engl J Med 373:1541-52, 2015
6. Royston P, Altman DG, Sauerbrei W: Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 25:127-41, 2006
7. Toft N, Birgens H, Abrahamsson J, et al: Results of NOPHO ALL2008 treatment for patients aged 1-45 years with acute lymphoblastic leukemia. Leukemia, 2017
8. 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 34:2591-601, 2016
9. Escherich G, Zimmermann M, Janka-Schaub G, et al: 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 60:254-7, 2013
10. Moorman AV, Ensor HM, Richards SM, et al: Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial. Lancet Oncol 11:429-38, 2010
11. Moorman AV, Enshaei A, Schwab C, et al: A novel integrated cytogenetic and genomic classification refines risk stratification in pediatric acute lymphoblastic leukemia. Blood 124:1434-44, 2014
12. Hamadeh L, Enshaei A, Schwab C, et al: Validation of the United Kingdom copy-number alteration classifier in 3239 children with B-cell precursor ALL. Blood Adv 3:148-157, 2019
13. Irving JAE, Enshaei A, Parker CA, et al: Integration of genetic and clinical risk factors improves prognostication in relapsed childhood B-cell precursor acute lymphoblastic leukemia. Blood 128:911-922, 2016

14. Sauerbrei W: The use of resampling methods to simplify regression models in medical statistics. Journal of the Royal Statistical Society Series C-Applied Statistics 48:313-329, 1999

15. Royston P, Altman DG: External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 13:33, 2013

16. Schwab C, Ryan SL, Chilton L, et al: EBF1-PDGFRB fusion in pediatric B-cell precursor acute lymphoblastic leukemia (BCP-ALL): genetic profile and clinical implications. Blood 127:2214-8, 2016

17. Sutton R, Venn NC, Law T, et al: A risk score including microdeletions improves relapse prediction for standard and medium risk precursor B-cell acute lymphoblastic leukaemia in children. Br J Haematol 180:550-562, 2018

18. Wood B, Wu D, Crossley B, et al: Measurable residual disease detection by high-throughput sequencing improves risk stratification for pediatric B-ALL. Blood 131:1350-1359, 2018

19. Lilljebjorn H, Fioretos T: New oncogenic subtypes in pediatric B-cell precursor acute lymphoblastic leukemia. Blood 130:1395-1401, 2017

20. 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 124:1056-61, 2014

21. Hunger SP: More Is Not Always Better: The Perils of Treatment Intensification in Pediatric Acute Lymphoblastic Leukemia. J Clin Oncol 37:1601-1603, 2019

22. Maude SL, Frey N, Shaw PA, et al: Chimeric antigen receptor T cells for sustained remissions in leukemia. N Engl J Med 371:1507-17, 2014

23. O'Connor D, Moorman AV, Wade R, et al: Use of Minimal Residual Disease Assessment to Redefine Induction Failure in Pediatric Acute Lymphoblastic Leukemia. J Clin Oncol 35:660-667, 2017

Table 1: Univariable and Multivariable Cox models for the risk of relapse for patients treated on UKALL2003

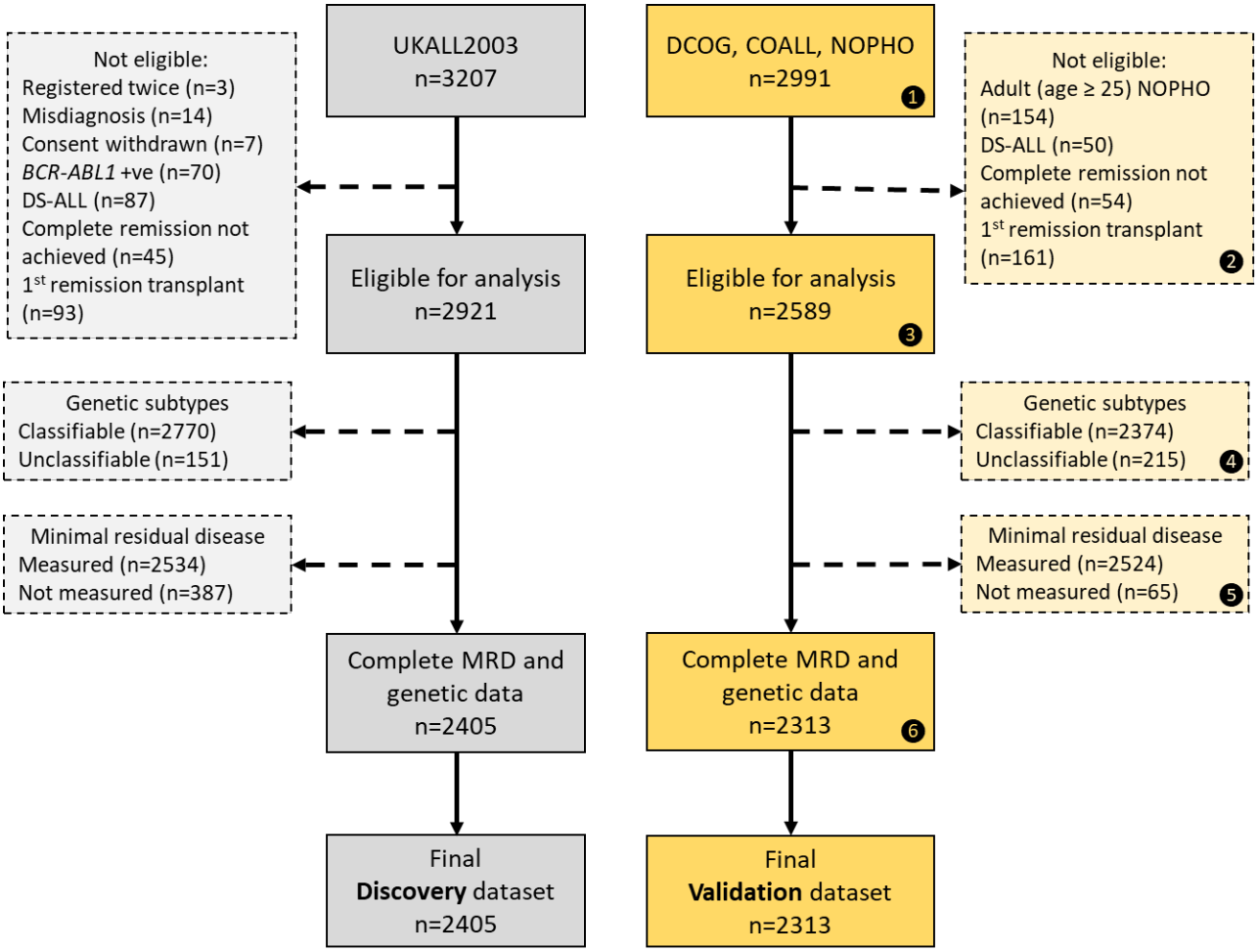
Univariate analysis	**Variable structure**	**Hazard ratio for risk of relapse (95% CI)**	**Coefficient (95% CI)**	**p-value**
Sex	Male v Female	1.39 (1.04-1.84)	0.33 (0.05-0.61)	0.022
Age (years)	Continuous	1.06 (1.03-1.08)	0.06 (0.03-0.08)	<0.001
White cell count (x10^9/L) [1]	Continuous (log)	1.27 (1.16-1.39)	0.24 (0.15-0.33)	<0.001
CNS disease [2]	Yes v No	3.09 (1.59-6.03)	1.12 (0.46-1.80)	0.001
T-cell status	Yes v No	1.85 (1.30-2.63)	0.61 (0.26-0.96)	0.001
τ(MRD) [3]	Continuous (log)	0.79 (0.75-0.82)	-0.24 (-0.28-(-0.20))	<0.001
Slow early responder	Yes v No	2.99 (2.18-4.11)	1.09 (0.78-1.41)	<0.001
Cytogenetic risk group				
Good risk [4]	Yes v No	0.39 (0.30-0.52)	-0.94 (-1.22-(-0.66))	<0.001
High risk [5]	Yes v No	3.92 (2.45-6.28)	1.37 (0.89-1.84)	<0.001
Multivariate model[6]	**Variable structure**	**Hazard ratio for risk of relapse (95% CI)**	**Coefficient (95% CI)**	**p-value**
τ(MRD) [3]	Continuous (log)	0.80 (0.77-0.84)	-0.22 (-0.26-(-0.18))	<0.001
Cytogenetic Good risk [4]	Yes (1) v No (0)	0.64-0.47-0.88)	-0.43 (-0.75-(-0.13))	0.005
Cytogenetic High risk [5]	Yes (1) v No (0)	2.90 (1.79-4.72)	1.07 (0.58-1.55)	<0.001
White cell count [1]	Continuous (log)	1.15 (1.05-1.26)	0.14 (0.05-0.23)	0.003

Notes: (1) White cell count was transformed as follows: ln(WCC+1); (2) Central nervous system (CNS) disease at diagnosis defined as the presence of >5/mm^3 unequivocal lymphoblasts in the CSF or cranial nerve palsy, parenchymal brain infiltrate or ocular infiltrate even in the absence of CSF blasts; (3) τ(MRD), log transformed minimal residual disease value (see methods); (4) Good risk cytogenetics: *ETV6-RUNX1*, high hyperdiploidy; (5) High risk cytogenetics: *KMT2A/MLL* fusions, near-haploidy, low hypodiploidy, iAMP21 and *TCF3-HLF;* (6) All variables significant in univariate analysis were included in the multivariate modelling.

Table 2: Cox Models for relapse rate, event-free and overall survival using the UKALL prognostic index and original risk definition in the discovery and validation cohorts.

Outcome measure **Prognostic factor**	**Discovery Cohort**	**Validation Cohorts**		
	C-index (95% CI)			
Event Free Survival	ALL2003	DCOG-ALL10	COALL-07-03	NOPHO-ALL2008
Model 1: PI_{UKALL} - linear variable	0.73 (0.69-0.76)**	0.68 (0.61-0.74)**	0.70 (0.61-0.78)**	0.70 (0.66-0.75)**
Model 2: PI_{UKALL} - 4 categories	0.70 (0.67-0.74)**	0.64 (0.57-0.70)**	0.68 (0.60-0.76)**	0.68 (0.63-0.72)**
Model 3: Original risk groups	0.60 (0.57-0.64)	0.59 (0.52-0.65)	0.51 (0.43-0.60)	0.66 (0.62-0.71)
Relapse Rate	ALL2003	DCOG	COALL	NOPHO
Model 1: PI_{UKALL} - linear variable	0.74 (0.70-0.77)**	0.68 (0.61-0.75)**	0.69 (0.60-0.79)**	0.76 (0.72-0.81)**
Model 2: PI_{UKALL} - 4 categories	0.72 (0.68-0.75)**	0.64 (0.57-0.71)**	0.69 (0.59-0.78)**	0.73 (0.69-0.78)**
Model 3: Original risk groups	0.61 (0.57-0.64)	0.55 (0.49-0.62)	0.50 (0.41-0.59)	0.68 (0.62-0.73)
Overall Survival	ALL2003	DCOG	COALL	NOPHO
Model 1: PI_{UKALL} - linear variable	0.79 (0.75-0.82)**	0.73 (0.65-0.81)*	0.83 (0.76-0.90)**	0.74 (0.68-0.80)
Model 2: PI_{UKALL} - 4 categories	0.76 (0.72-0.80)**	0.67 (0.58-0.77)	0.80 (0.71-0.89)**	0.73 (0.67-0.79)
Model 3: Original risk groups	0.65 (0.61-0.69)	0.67 (0.59-0.74)	0.59 (0.48-0.70)	0.70 (0.64-0.76)

Abbreviations: PI, prognostic index; C-index, Harrell's concordance index; CI, confidence interval



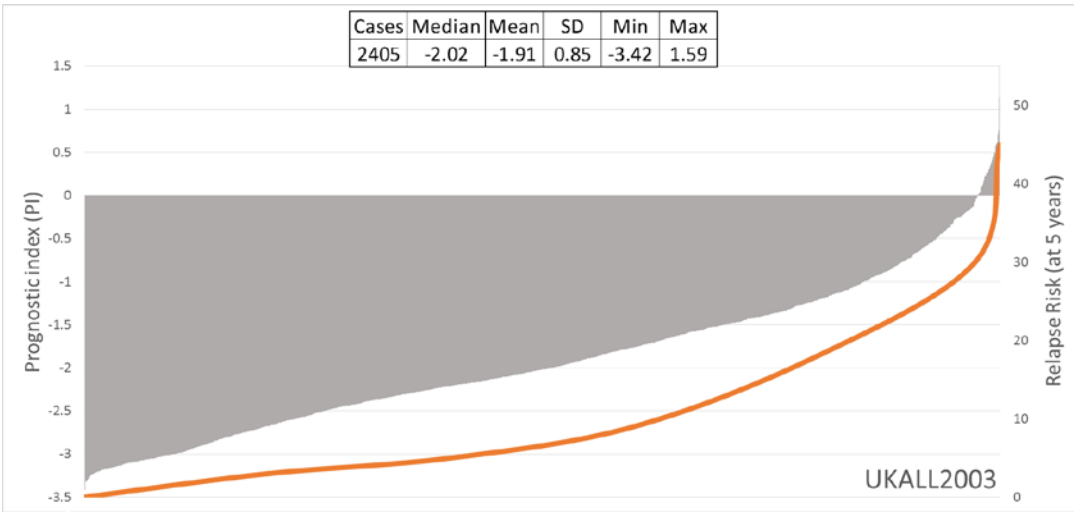
	COALL	DCOG	NOPHO
1	445	777	1769
2	36	95	117
3	409	682	1498
4	150	64	1
5	0	30	35
6	259	592	1462

Figure 1: CONSORT diagram for the discovery and validation datasets.
NB Excluded patients (dotted boxes) are counted in each applicable category.

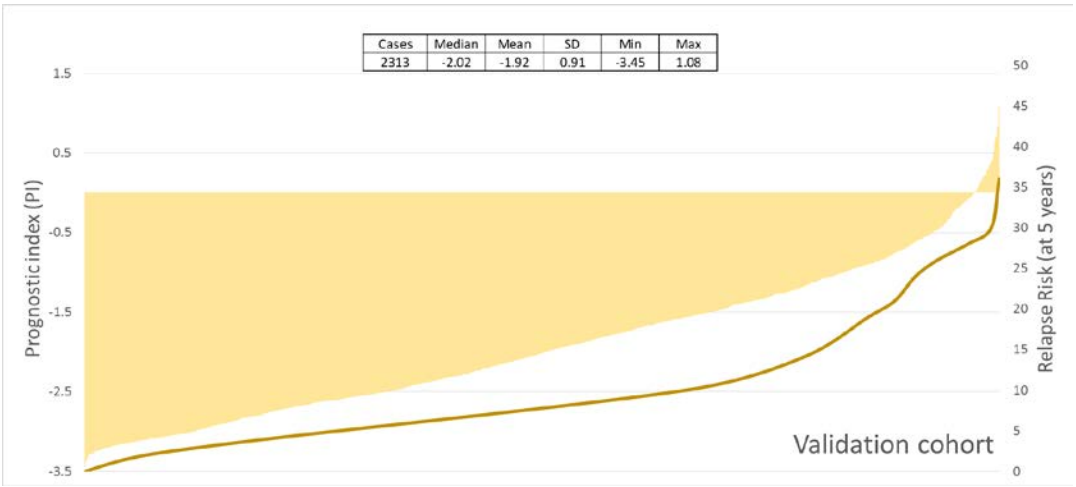
A

$$PI_{UKALL} = \tau(MRD) \times -0.218 + CYTO\text{-}GR \times -0.440 + CYTO\text{-}HR \times 1.066 + \tau(WCC) \times 0.138$$

B



C



D

Cohort / Trial	Hazard ratio (95% CI) for the risk of …		
	Event	Relapse	Death
Discovery/UKALL2003	2.53 (2.22-2.87)	2.72 (2.36-3.13)	3.18 (2.69-3.75)
Validation	2.17 (1.91-2.46)	2.33 (2.02-2.70)	2.66 (2.26-3.15)
NOPHO-ALL2008	2.34 (1.98-2.77)	2.80 (2.29-3.44)	2.74 (2.21-3.40)
DCOG-ALL10	2.07 (1.63-2.63)	2.05 (1.58-2.67)	2.35 (1.70-3.26)
CoALL-07-03	1.93 (1.37-2.73)	1.90 (1.31-2.74)	3.13 (1.91-5.13)

Figure 2: Definition (A) and distribution (B,C) of the UKALL prognostic index along with its association with risk of relapse (D). (A) The linear model derived from the coefficients of the multivariate model; (B &C) These bar charts show the distribution of the patient specific PI values derived from the model for the discovery (B) and validation (C) cohorts. The in-laid table gives the mean, median, standard deviation and minimum/maximum values of the distribution. The line shows the smoothed risk of relapse estimated for 10 equal-sized subgroups. (D) A table showing hazard ratios for the UKALL prognostic index as a continuous variable from univariate Cox models across the two cohorts and three trials within the validation cohort. Abbreviations: τ(MRD), log transformed minimal residual disease value; CYTO-GR, Cytogenetic Good Risk; CYTO-HR, Cytogenetic High Risk; τ(WCC), log transformed white cell count.

A

Risk Group	Discovery Cohort (UKALL2003)				Validation Cohort			
	n(%)	EFS	RR	OS	n(%)	EFS	RR	OS
LR	1319 (55)	96% (95-97)	3% (2-4)	99% (98-99)	1254 (54)	93% (91-94)	4% (3-6)	97% (96-98)
SR	553 (23)	90% (87-92)	8% (6-10)	95% (92-96)	490 (21)	90% (87-93)	9% (6-12)	96% (95-98)
IR	465 (19)	80% (76-83)	17% (14-21)	88% (85-91)	489 (21)	80% (76-83)	17% (14-21)	87%(83-90)
HR	68 (3)	51% (38-62)	48% (36-60)	63% (50-73)	80 (3)	55% (43-65)	35% (24-48)	69% (57-78)

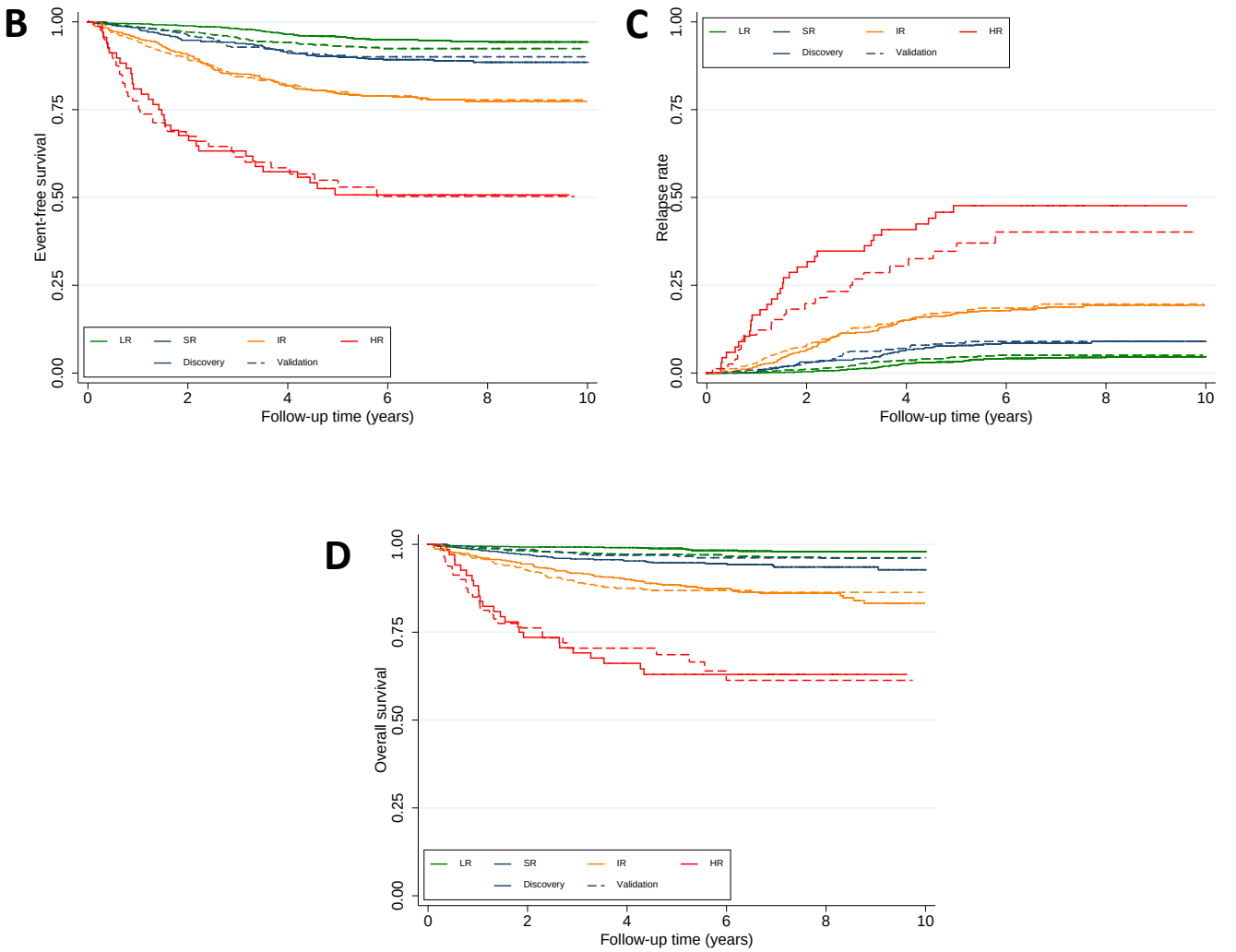
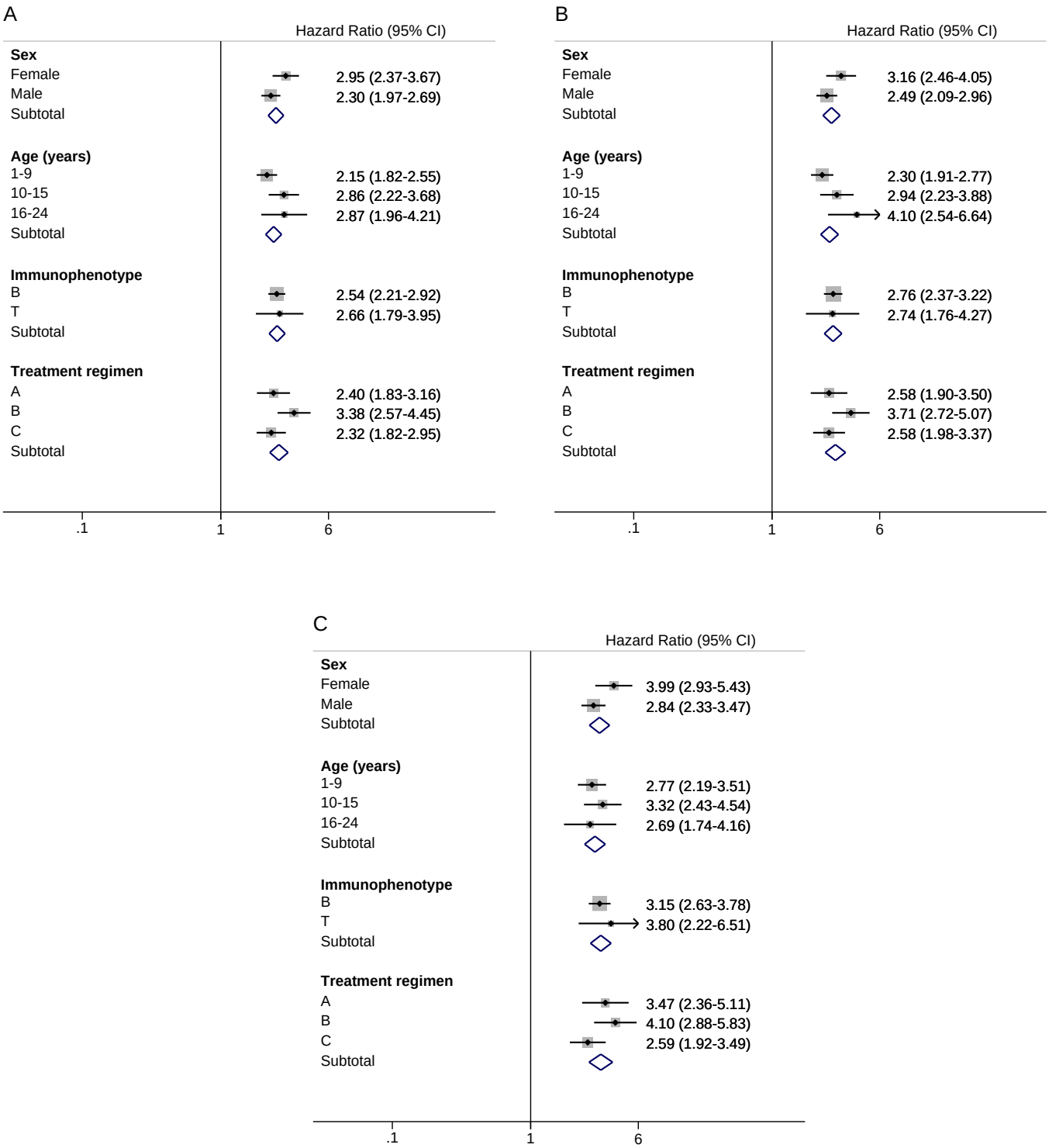
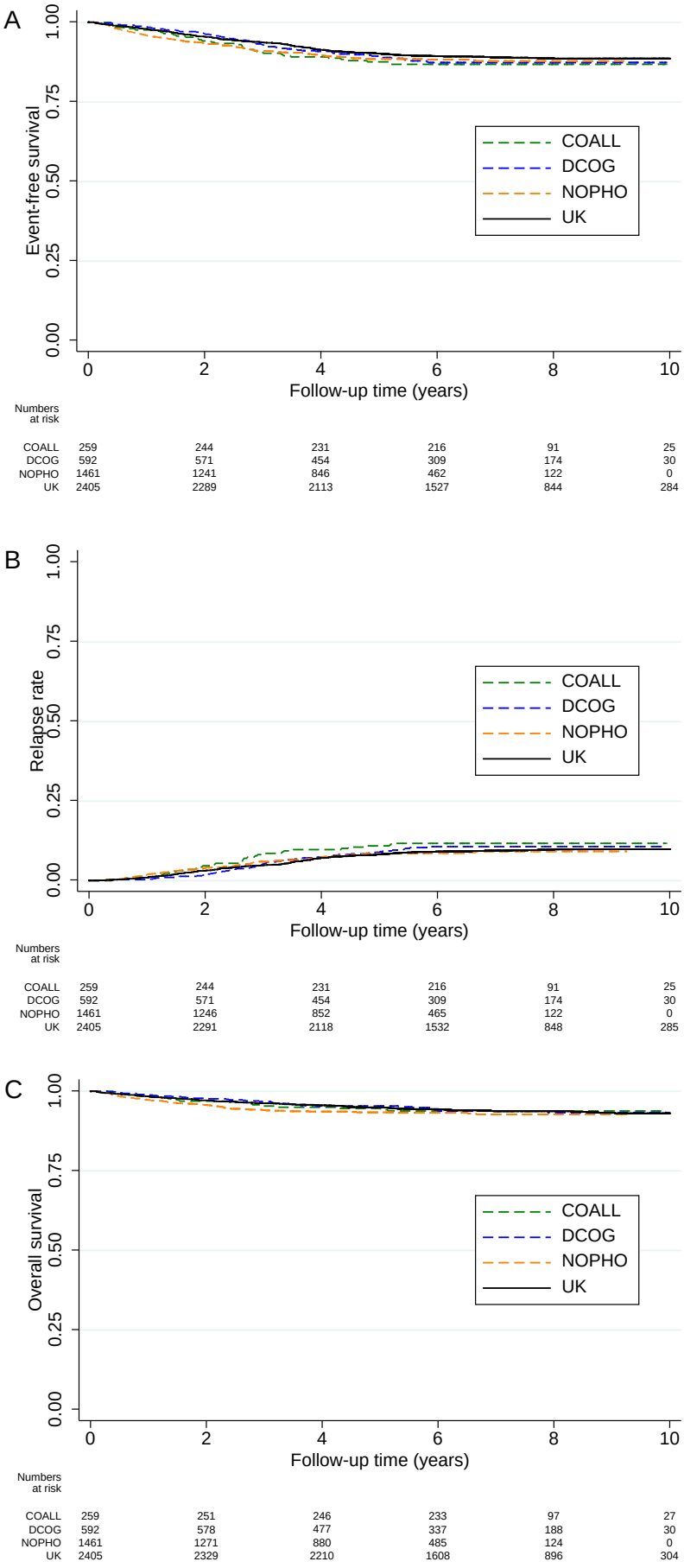


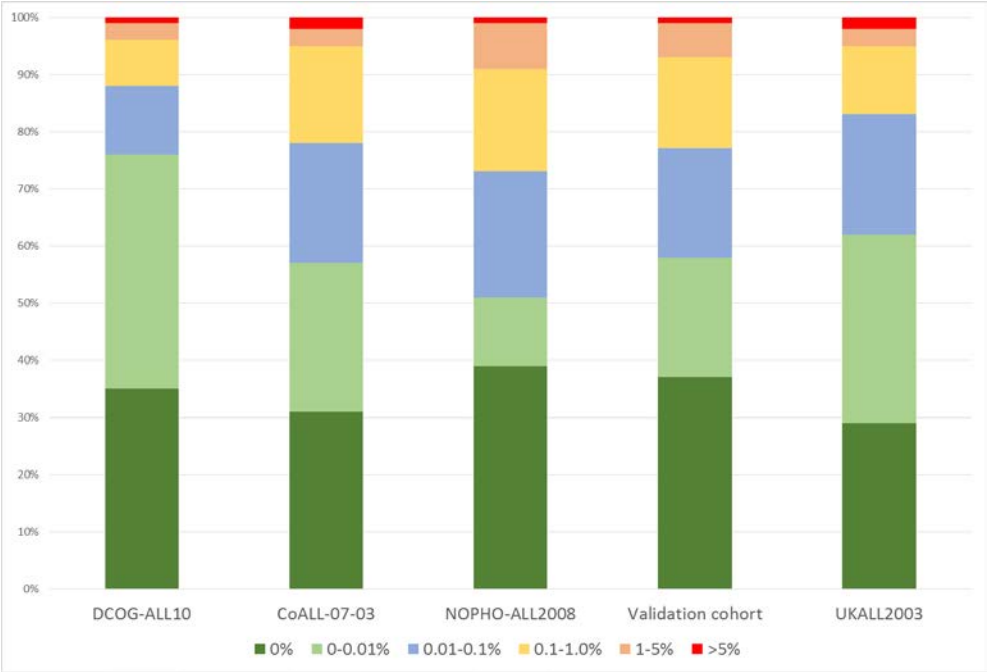
Figure 3: Outcome of patients in the discovery and validation cohorts sub-divided into four PI_{UKALL} defined risk groups. (A) Number of cases and event free survival (EFS), relapse rate (RR) and overall survival rates at 5 years. (B, C, D) Kaplan-Meier plots EFS, RR and OS. The PI_{UKALL} thresholds for defining each risk group were as follows: low risk (LR) ≤-1.894893; standard risk (SR) ≤-1.279577; intermediate risk (IR) ≤-0.0856656; high risk (HR) >-0.0856656.



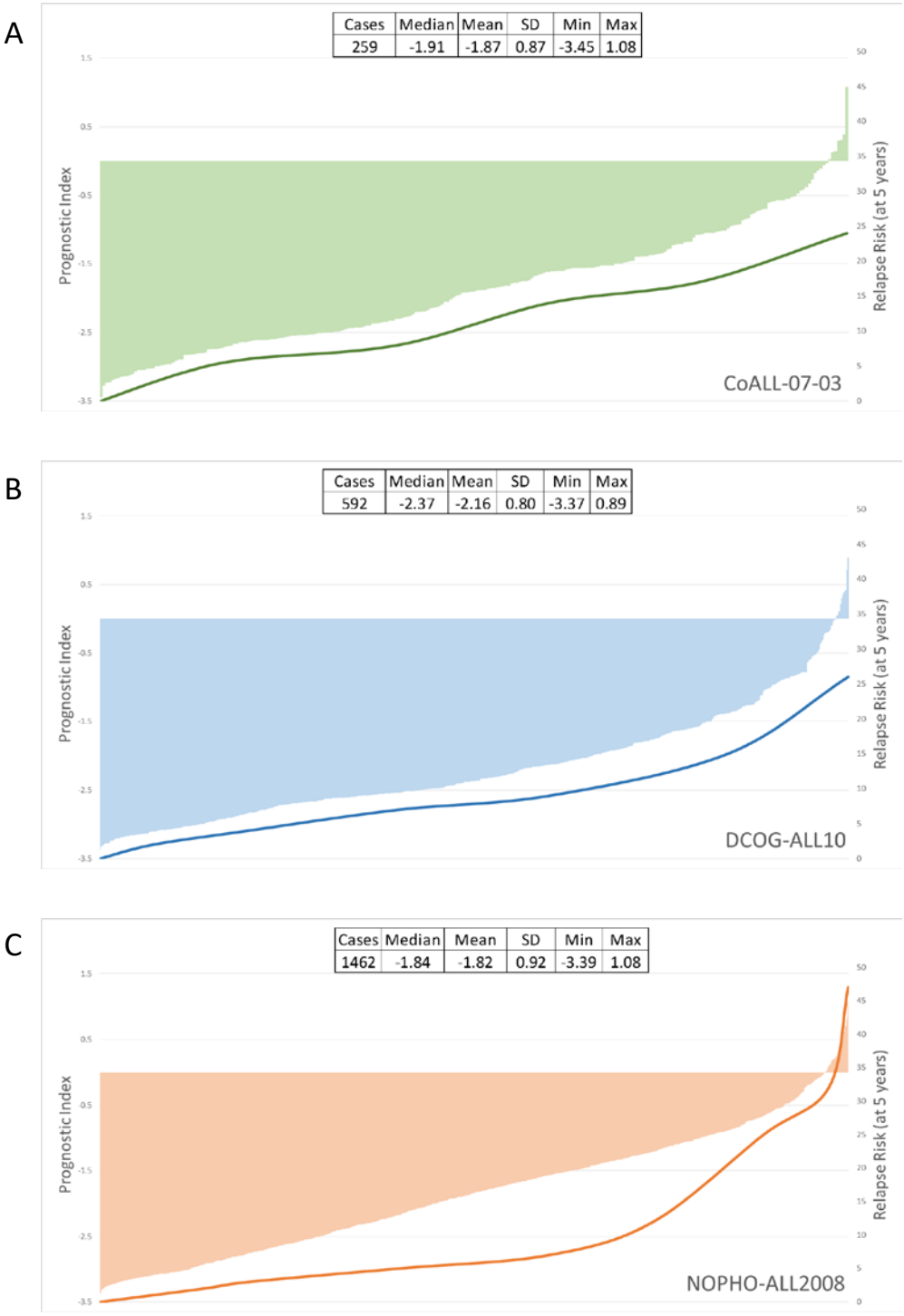
Supplementary Figure 1: Forest plots showing the hazard ratio for each unit increase in the UKALL prognostic index (PI_{UKALL}) across different patient and treatment subgroups in UKALL2003. The hazard ratio and 95% confidence interval are derived from univariate Cox models of PI_{UKALL} as a continuous variable and represent the increased risk for (a) event free survival (a); risk of relapse (b); overall survival (c) per unit increase and illustrates the robustness of the PI_{UKALL} to predict outcome independently of other risk factors and different intensities of chemotherapy. As WCC, MRD and genetics were used to derive the PI_{UKALL}, these subgroups have not been included in the Forest plot.



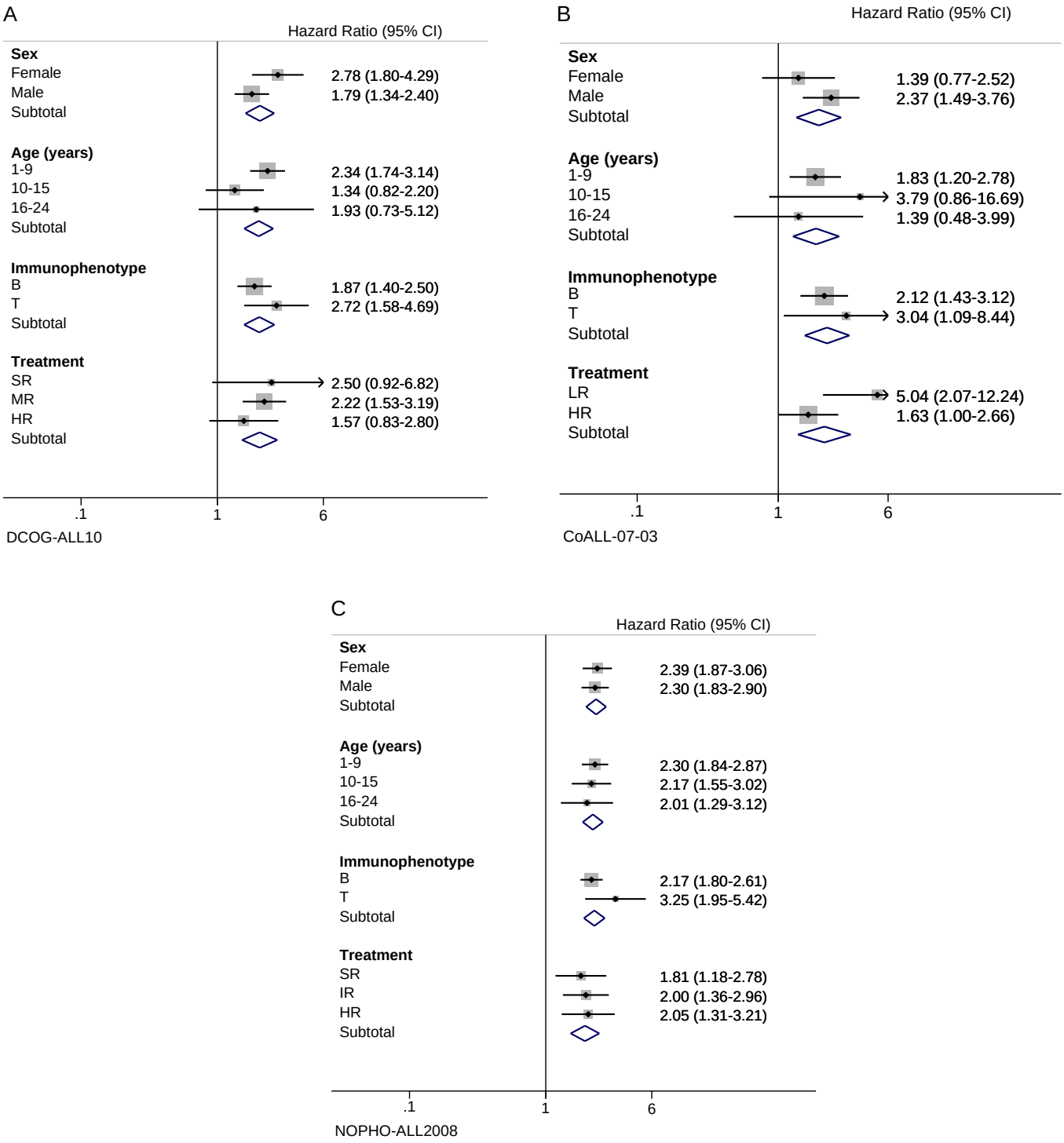
Supplementary Figure 2: Event free survival (A), relapse risk (B) and overall survival (C) for the three validation cohorts (DCOG-ALL10, CoALL-07-03 and NOPHO-ALL2008) and the discovery cohort (UKALL2003).



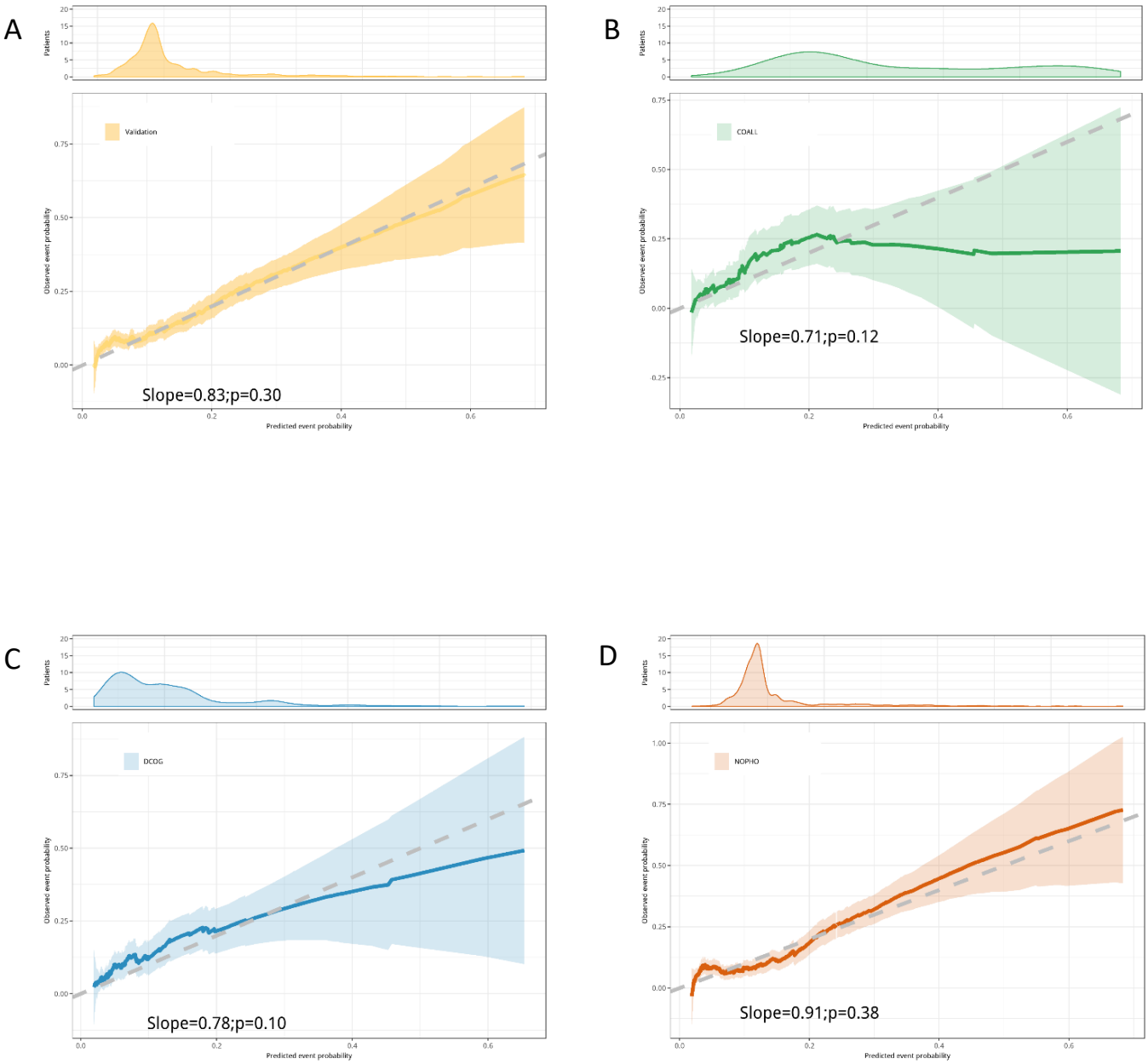
Supplementary Figure 3: Distribution of Minimal Residual Disease (MRD) across the four clinical trials used in this study. MRD was measured at the end of induction in all trials. UKALL2003, DCOG-ALL10 and CoALL-07-03 measured MRD by Ig/TCR PCR whereas NOPHO-ALL2008 used flow cytometry.



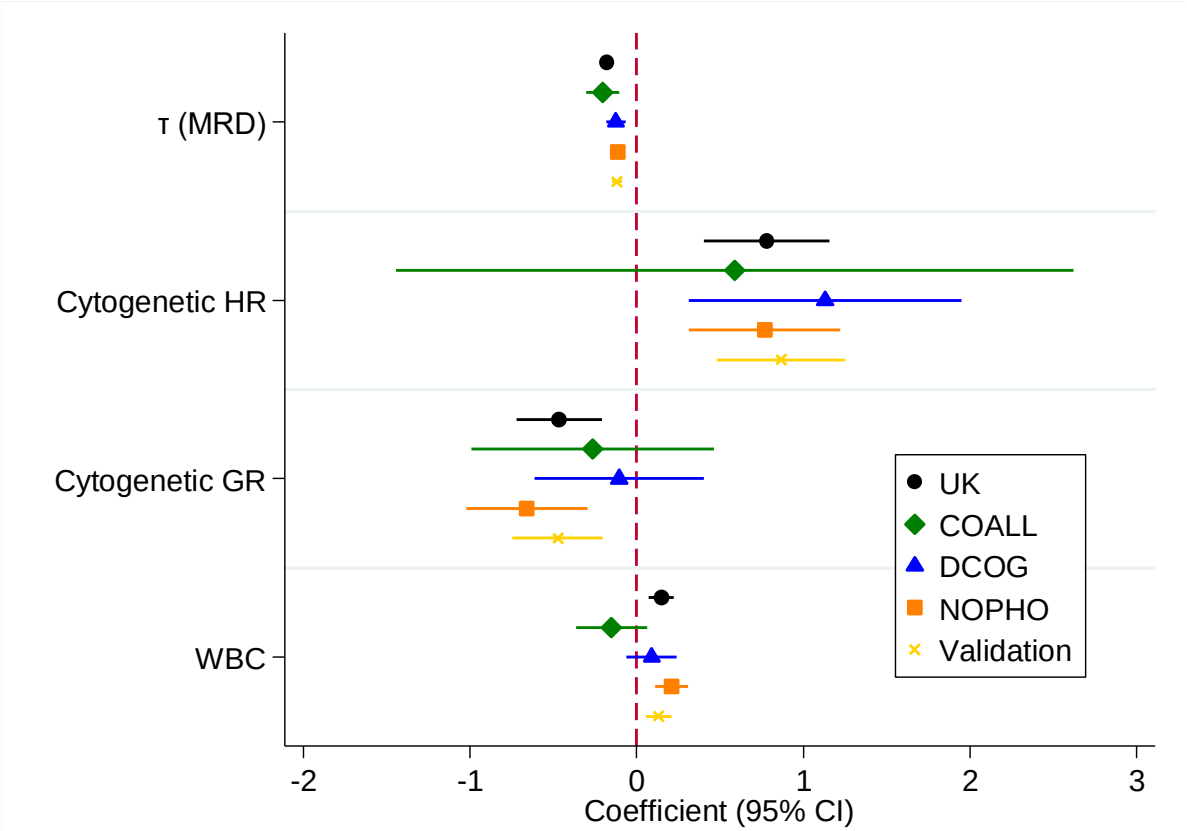
Supplementary Figure 4: Distribution of the UKALL prognostic index (PI_{UKALL}) in CoALL-07-03 (A), DCOG-ALL10 (B), NOPHO-ALL2008 (C) trials along with its association with risk of relapse. The bar chart component of each graph shows the distribution of the PI_{UKALL} values for each patient in the discovery (B) and validation (C) cohorts with the metrics for the distribution shown in each table. The risk of relapse was estimated for 10 equal-sized subgroups and plotted as a smoothed function (line).



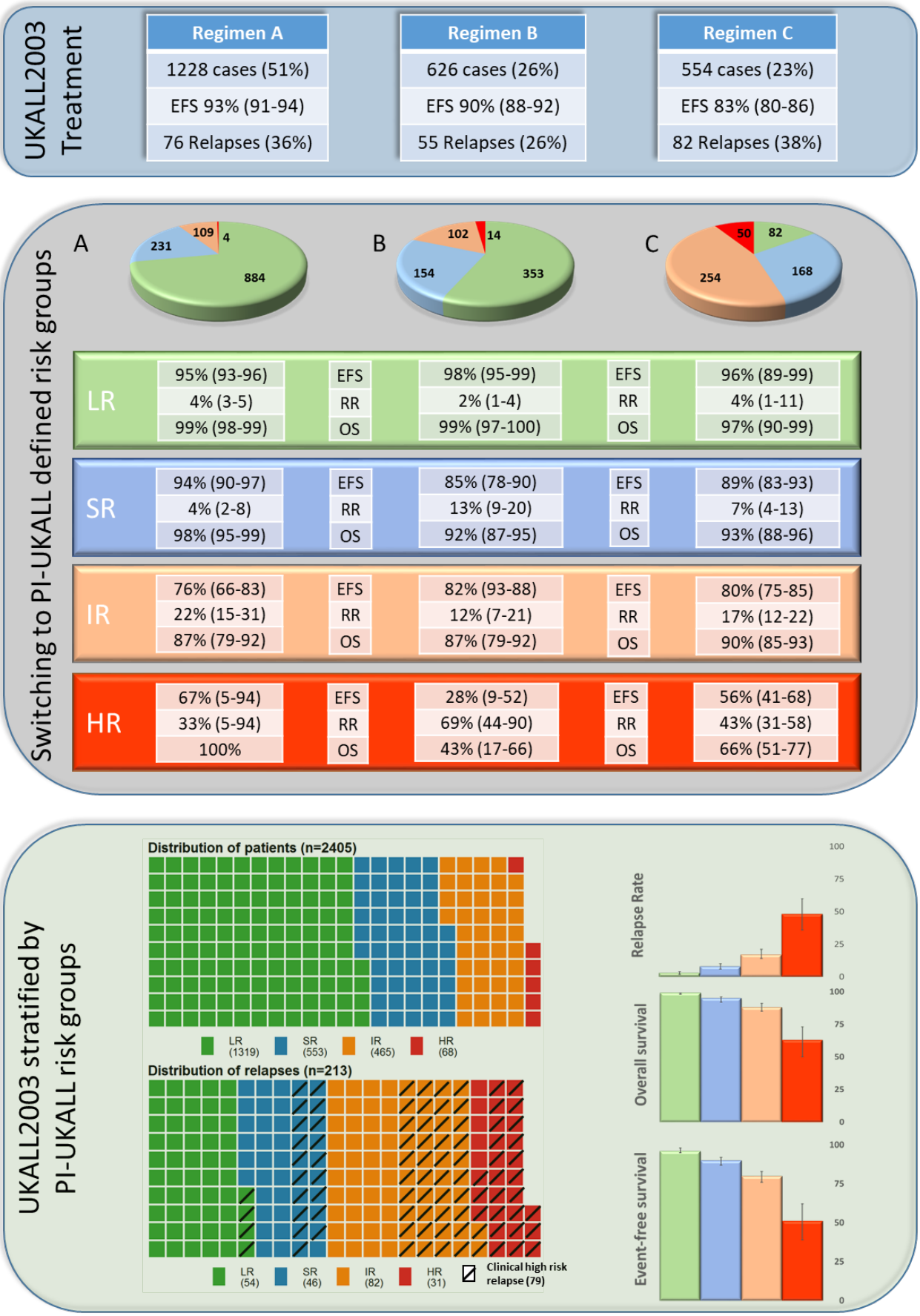
Supplementary Figure 5: Forest plots showing the hazard ratio for each unit increase in the UKALL prognostic index (PI_{UKALL}) across different patient and treatment subgroups in in CoALL-07-03 (A), DCOG-ALL10 (B) and NOPHO-ALL2008 (C). The hazard ratio and 95% confidence interval are derived from univariate Cox models of PI_{UKALL} as a continuous variable and represent the increased risk for event free survival per unit increase and illustrates the robustness of the PI_{UKALL} to predict outcome independently of other risk factors and different intensities of chemotherapy. As WCC, MRD and genetics were used to derive the PI_{UKALL}, these subgroups have not been included in the Forest plot.



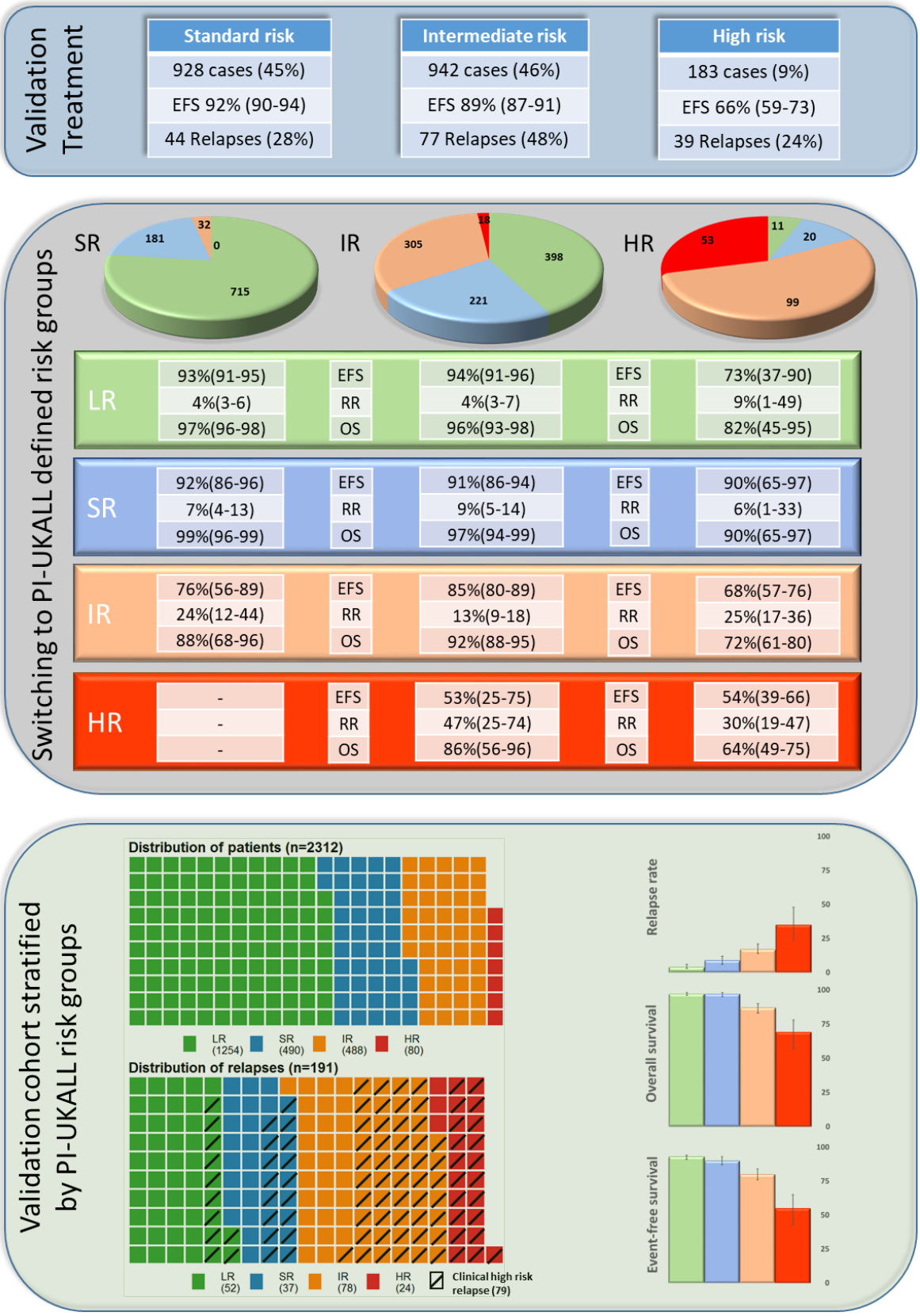
Supplementary Figure 6: Calibration the UKALL prognostic index (PI_{UKALL}) using the validation cohort. Each graph compares the predicted event probability (X axis) with the observed event probability (Y axis) for the whole validation cohort (A) and each constituent dataset (B, CoALL; C, DCOG; D, NOPHO). The dotted grey line represents perfect calibration (i.e. 1) whereas the solid coloured line represents the actual calibration. The shaded area represents the 95% confidence interval. Above each graph is a density plot showing the number of patients at the event probability. These graphs illustrates that PI_{UKALL} predicts outcome across the full spectrum of probabilities.



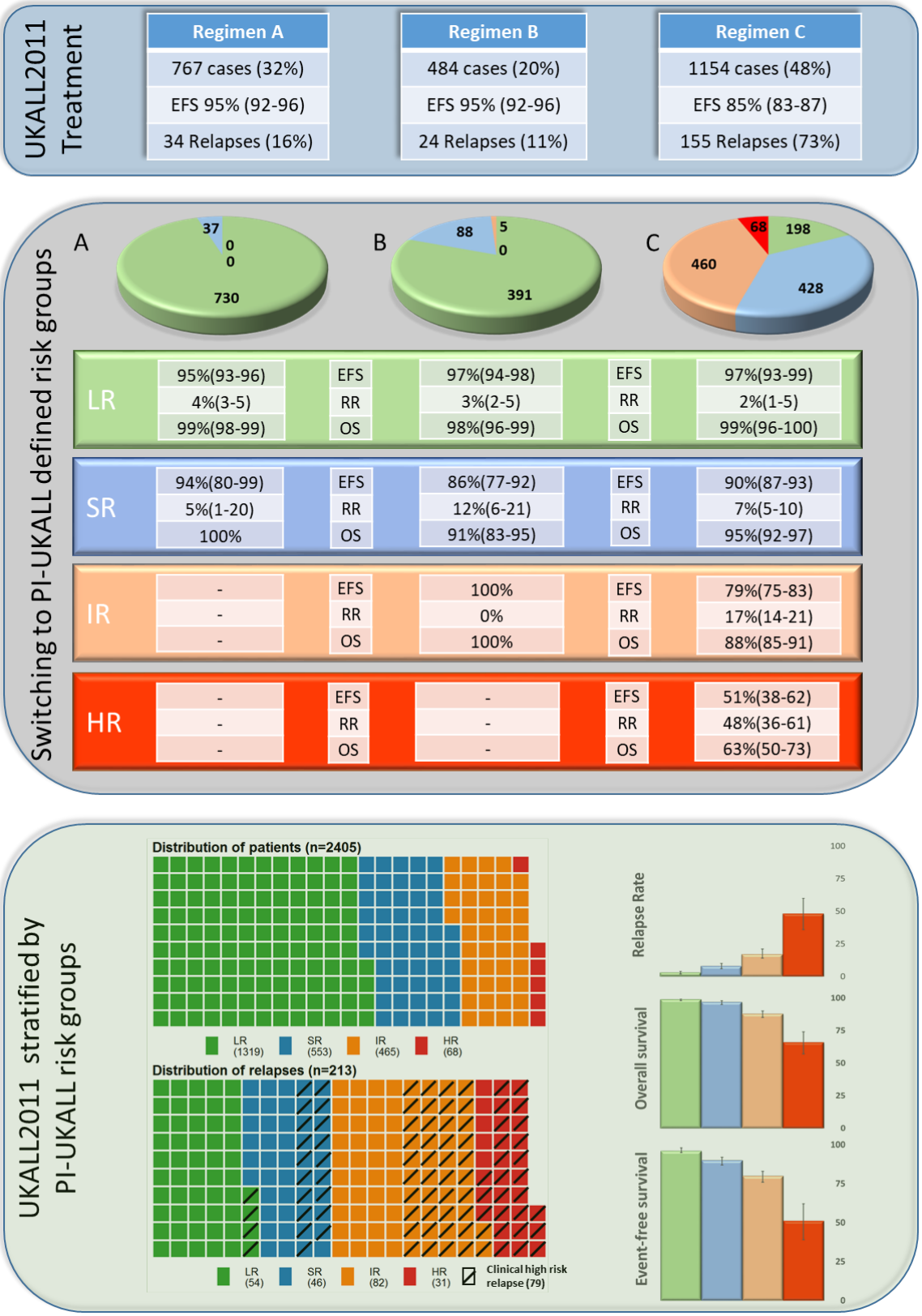
Supplementary Figure 7: Forest plot showing the coefficient and 95% confidence interval for each variable in the final model for the discovery and validation cohorts as well as each of the three datasets comprising the validation cohort. The similarity of each coefficient across the datasets confirms that each component of the PI_{UKALL} is contributing equivalently across the different datasets.



Supplementary Figure 8: Diagram illustrating the benefit of using PI_{UKALL} defined risk groups in the discovery cohort. The top panel shows the distribution and outcome of patients according to the risk groups used in the UKALL2003 trial. The middle panel shows how patients in each of the original risk groups distributes across the new PI_{UKALL} defined groups. The waffle plots in the bottom panel illustrates the distribution of patients and relapses according to the new PI_{UKALL} defined groups with the number of patients in each risk group shown at the bottom in parentheses. The definition of clinical high risk relapses is given in the supplementary methods. The three bar charts in the bottom panel show the vent-free survival (EFS), relapse rate (RR) and overall survival (OS) rates at 5 years across the four groups.



Supplementary Figure 9: Diagram illustrating the benefit of using PI_{UKALL} defined risk groups in the validation cohort. The top panel shows the distribution and outcome of patients according to the risk groups used in the validation cohort. The middle panel shows how patients in each of the original risk groups distributes across the new PI_{UKALL} defined groups. The waffle plots in the bottom panel illustrates the distribution of patients and relapses according to the new PI_{UKALL} defined groups with the number of patients in each risk group shown at the bottom in parentheses. The definition of clinical high risk relapses is given in the supplementary methods. The three bar charts in the bottom panel show the vent-free survival (EFS), relapse rate (RR) and overall survival (OS) rates at 5 years across the four groups. Only patients from DCOG-ALL10 and NOPHO-ALL2008 have been included in this figure because the CoALL-07-03 trial only used two risk groups.



Supplementary Figure 10: Diagram illustrating the benefit of using PI_{UKALL} defined risk groups in the discovery cohort using the UKALL2011 risk classification system. The top panel shows the distribution and outcome of UKALL2003 patients according to the UKALL2011 risk classification. Regimen A comprises all NCI standard risk BCP-ALL and Down Syndrome patients with an end of induction MRD level <0.005%. Regimen B comprises all remaining patients with an end of induction MRD level <0.005%. While regimen C comprises all patients an end of induction MRD level >=0.005% and patients with HR cytogenetics. The middle panel shows how patients in each of the original risk groups distributes across the new PI_{UKALL} defined groups. The waffle plots in the bottom panel illustrates the distribution of patients and relapses according to the new PI_{UKALL} defined groups with the number of patients in each risk group shown at the bottom in parentheses. The definition of clinical high risk relapses is given in the supplementary methods. The three bar charts in the bottom panel show the vent-free survival (EFS), relapse rate (RR) and overall survival (OS) rates at 5 years across the four groups.

Supplementary Table 1: Definition of risk group and details of induction therapy for the four clinical trial analysed in this study

Risk group	Risk Group Definition	Induction Therapy [a]
UKALL2003		
Regimen A	<10 years, WCC<50x10^9, <25% blasts @ day 15 and MRD<0.01% or MRD≥0.01% + Rx to ST	Dexamethasone 6mg/m^2 day 1-28 Vincristine 1.5mg/m^2 day 2,9,16,23,30 Pegylated L-asparaginase 1,000 IU/m^2 day 4,18 Intrathecal methotrexate 8-12mg by age day 1,8,28 Mercaptopurine 75mg/m^2 day 29-35
Regimen B	≥10 years, WCC≥50x10^9, <25% blasts @ day 8 and MRD<0.01% or MRD≥0.01% + Rx to ST	As above except Mercaptopurine 60mg/m^2 days 29-35 and plus Daunorubicin 25mg/m^2 days 2,9,16,23
Regimen C	HR cytogenetics [b] or ≥25% blasts at day 8/15 or MRD≥0.01% + Rx to AT	As above except Daunorubicin 45mg/m^2 days 2,9,16,23
DCOG-ALL10		
Standard risk	CR, PGR, MRD undetectable at time points 1 & 2 and No CNS/testicular disease	Prednisone 60 mg/m2 day 1-29 plus prophase plus tapering Vincristine 1.5 mg/m2 day 8, 15, 22, 29 Daunorubicin 30mg/m2 day 8, 15, 22, 29 Asparaginase (E. coli) 5,000 IU/m2 8 doses day 12-30
Intermediate risk	All other cases	Intrathecal methotrexate dose by age day 1;
High risk	No CR or PPR, MRD≥0.05% at time point 1 & 2 or *KMT2A-AF4*	Intrathecal methotrexate, cytarabine, prednisolone days 15 and 29 (plus day 8 and 22 in case of TLP+, CNS2 and CNS3).
NOPHO-ALL2008		
Standard risk	WCC<100x10^9/L, pre-B cell & MRD day 29 <0.1%. Not dic(9;20), iAMP21 or t(1;19). No CNS disease	Prednisolone 60 mg/m^2 day 1-29; Vincristine 2mg/m^2 day 1,8,15,22,29 Doxorubicin 40 mg/m^2 day 1,22 Intrathecal Methotrexate day 1,8,15,29
Intermediate risk	All other cases	Dexamethasone 10 mg/m^2 day 1–21 Vincristine 2.0 mg/m^2 day 1,8,15,22,29 Doxorubicin 40 mg/m^2 day 1,22 Intrathecal Methotrexate day 1,8,15,29
High risk	WBC≥100x10^9/L a/o T-cell & day 15 MRD≥25% or day 29 MRD≥0.1% or Any WBC/immunophenotype and day 29 MRD ≥5% or 79 MRD ≥0.1% Any WBC/response and KMT2A fusion or hypodiploidy (<45 chrs).	
CoALL-07-03		
Low risk	All other cases	Rx: Prephase Doxorubicin 30mg/m^2 or Daunorubicin 30/40mg/m^2 Prednisolone 60 mg/m2 days 1-28 Vincristine 1.5 mg/m2 day 1,8,15,22
High risk	≥10 years, ≥25x10^9/L, *KMT2A-AF4* , *BCR-ABL1* , No CR, T-ALL or pro-B ALL	Daunorubicin 36 mg/m2 day 1,8,15

Notes: (a) For the purposes of this paper we have induction therapy from the start of leukaemia to the first MRD time point; (b) KMT2A/MLL fusions, near-haploidy (<30 chromosomes), low hypodiploidy (30-39 chromosomes), intrachromosomal amplification of chromosome 21q (iAMP21) and t(17;19)(q23;p13)TCF3-HLF

Abbreviations: WCC, white cell count; CNS, Central Nervous System; MRD, Minimal Residual Disease at the end of induction therapy (unless otherwise stated); Rx, randomised; ST, Standard therapy; AT, augmented therapy; CR, complete remission; PGR, prednisone good response; PPR, prednisone good response.

Supplementary Table 2: Demographic, clinical, genetic and outcome features of the four datasets comprising the validation cohort

	UKALL2003				DCOG-ALL10				CoALL-07-03				NOPHO-ALL2008				Validation cohort		
	Total	Included	Excluded		Total	Included	Excluded		Total	Included	Excluded		Total	Included	Excluded		Total	Included	Excluded
	2921	2405	516		682	592	90		409	259	150		1498	1462	36		2589	2313	276
Sex																			
Female	1267(43)	1043(43)	224(43)		319(47)	273(46)	46(51)		182(44)	108(42)	74(49)		682(46)	669(46)	13(36)		1183(46)	1050(45)	133(48)
Male	1654(57)	1362(57)	292(57)		363(53)	319(54)	44(49)		227(56)	151(58)	76(51)		816(54)	793(54)	23(64)		1406(54)	1263(55)	143(52)
Age																			
1-9 years	2174(74)	1819(76)	355(69)		526(77)	462(78)	64(71)		335(82)	215(83)	120(80)		1131(76)	1113(76)	18(50)		1992(77)	1790(77)	202(73)
10-15 years	544(19)	429(18)	115(22)		126(18)	108(18)	18(20)		59(14)	35(14)	24(16)		203(14)	199(14)	4(11)		388(15)	342(15)	46(17)
16+	203(7)	157(7)	46(9)		30(4)	22(4)	8(9)		15(4)	9(3)	6(4)		164(11)	150(10)	14(39)		209(8)	181(8)	28(10)
WCC																			
<50	2294(79)	1890(79)	404(78)		572(84)	488(82)	84(94)		336(82)	199(77)	137(92)		1168(78)	1144(78)	24(67)		2076(80)	1831(79)	245(89)
<100	293(10)	242(10)	51(10)		58(9)	53(9)	5(6)		33(8)	28(11)	5(3)		147(10)	145(10)	2(6)		238(9)	226(10)	12(4)
100+	334(11)	273(11)	61(12)		51(7)	51(9)	0(0)		39(10)	32(12)	7(5)		183(12)	173(12)	10(28)		273(11)	256(11)	17(6)
Cytogenetic good risk																			
Yes	1223(44)	1013(42)	210(58)		268(43)	254(43)	14(54)		136(53)	136(53)	0(.)		692(46)	670(46)	22(63)		1096(46)	1060(46)	36(59)
No	1547(56)	1392(58)	155(42)		350(57)	338(57)	12(46)		123(47)	123(47)	0(.)		805(54)	792(54)	13(37)		1278(54)	1253(54)	25(41)
Cytogenetic high risk																			
Yes	2686(97)	2338(97)	348(95)		597(97)	571(96)	26(100)		256(98)	255(98)	1(100)		1422(95)	1392(95)	30(86)		2275(96)	2218(96)	57(92)
No	84(3)	67(3)	17(5)		21(3)	21(4)	0(0)		4(2)	4(2)	0(0)		75(5)	70(5)	5(14)		100(4)	95(4)	5(8)
MRD																			
0%	726(25)	694(29)	32(6)		227(33)	208(35)	19(21)		123(30)	80(31)	43(29)		570(38)	569(39)	1(3)		920(36)	857(37)	63(23)
0-0.005%	621(21)	596(25)	25(5)		271(40)	241(41)	30(33)		110(27)	63(24)	47(31)		117(8)	117(8)	0(0)		498(19)	421(18)	77(28)
0.005-0.01%	217(7)	200(8)	17(3)		1(0)	1(0)	0(0)		5(1)	5(2)	0(0)		57(4)	57(4)	0(0)		63(2)	63(3)	0(0)
0.01-0.1%	524(18)	500(21)	24(5)		73(11)	68(11)	5(6)		80(20)	53(20)	27(18)		327(22)	327(22)	0(0)		480(19)	448(19)	32(12)
0.1-1.0%	298(10)	277(12)	21(4)		53(8)	47(8)	6(7)		68(17)	43(17)	25(17)		268(18)	268(18)	0(0)		389(15)	358(15)	31(11)
1-5%	87(3)	81(3)	6(1)		20(3)	20(3)	0(0)		16(4)	9(3)	7(5)		109(7)	109(7)	0(0)		145(6)	138(6)	7(3)
>5%	61(2)	57(2)	4(1)		7(1)	7(1)	0(0)		7(2)	6(2)	1(1)		15(1)	15(1)	0(0)		29(1)	28(1)	1(0)
NA	387(13)	0(0)	387(75)		30(4)	0(0)	30(33)		0(0)	0(0)	0(0)		35(2)	0(0)	35(97)		65(3)	0(0)	65(24)
Dead																			
Yes	2718(93)	2256(94)	462(90)		639(94)	558(94)	81(90)		379(93)	243(94)	136(91)		1393(93)	1368(94)	25(76)		2411(93)	2169(94)	242(89)
No	203(7)	149(6)	54(10)		43(6)	34(6)	9(10)		30(7)	16(6)	14(9)		101(7)	93(6)	8(24)		174(7)	143(6)	31(11)
Relapse																			
Yes	2650(91)	2192(91)	458(89)		621(91)	536(91)	85(94)		360(88)	229(88)	131(87)		1383(93)	1356(93)	27(82)		2364(91)	2121(92)	243(89)
No	271(9)	213(9)	58(11)		61(9)	56(9)	5(6)		49(12)	30(12)	19(13)		111(7)	105(7)	6(18)		221(9)	191(8)	30(11)
Event																			
Yes	2581(88)	2143(89)	438(85)		604(89)	524(89)	80(89)		353(86)	225(87)	128(85)		1332(89)	1309(90)	23(70)		2289(89)	2058(89)	231(85)
No	340(12)	262(11)	78(15)		78(11)	68(11)	10(11)		56(14)	34(13)	22(15)		162(11)	152(10)	10(30)		296(11)	254(11)	42(15)
Event free survival *	90%(88-91)	90%(89-91)	87%(84-90)		90%(87-92)	89%(87-92)	90%(81-95)		87%(83-90)	88%(83-91)	86%(79-90)		88%(86-90)	88%(87-90)	67%(46-81)		88%(87-89)	89%(87-90)	85%(80-89)
Relapse rate *	8%(7-9)	8%(7-9)	9%(7-12)		8%(7-11)	9%(7-12)	6%(3-14)		11%(9-15)	11%(8-15)	12%(8-19)		9%(7-10)	8%(7-10)	24%(11-46)		9%(8-10)	9%(8-10)	11%(8-16)
Overall survival *	94%(93-95)	95%(94-96)	90%(87-93)		95%(93-96)	95%(93-97)	91%(82-95)		94%(91-96)	95%(91-97)	92%(86-95)		93%(91-94)	93%(92-95)	76%(57-87)		93%(92-94)	94%(93-95)	89%(85-93)
Median follow-up time (years)	7.29				6.66				8.01				4.85				6.09		

* at five years

Supplementary Table 3: Distribution and outcome of DCOG-ALL10 and NOPHO-ALL2008 patients classified according to their original risk groups and equivalently sized risk groups defined using the UKALL prognostic index (PI_{UKALL})

A		Original definition, Number of patients			
		SR	IR	HR	Total
PI-defined	SR	618	302	8	928
	IR	308	566	68	942
	HR	2	74	107	183
	Total	928	942	183	2053*

B		Original definition, Event-free survival, % (95% CI)			
		SR	IR	HR	Total
PI-defined	SR	93% (91-95)	95% (92-97)	75% (31-93)	94% (92-95)
	IR	91% (86-93)	89% (85-91)	83% (72-90)	89% (86-91)
	HR	50% (1-91) *	71% (57-81)	55% (45-64)	61% (53-68)
	Total	92% (90-94)	90% (87-91)	66% (59-73)	

C		Original definition, Relapse rate, % (95% CI)			
		SR	IR	HR	Total
PI-defined	SR	5% (3-7)	2% (1-5)	12% (2-61)	4% (3-6)
	IR	7% (5-11)	10% (8-13)	10% (5-21)	9% (7-11)
	HR	50% (9-99) *	28% (18-42)	33% (24-44)	31% (24-40)
	Total	6% (4-8)	9% (7-11)	23% (17-31)	

D		Original definition, Overall survival, % (95% CI)			
		SR	IR	HR	Total
PI-defined	SR	98% (96-99)	97% (94-98)	87% (39-98)	97% (96-98)
	IR	97% (94-99)	94% (92-96)	85% (73-91)	95% (93-96)
	HR	50% (1-91) *	88% (76-94)	63% (53-71)	72% (65-79)
	Total	97% (96-98)	95% (93-96)	72% (65-78)	

Notes: * Treatment risk group missing for one patient

Supplementary Table 4: Distribution and outcome of patients with an *IKZF1* deletion, *CRLF2* gene rearranagement or UKALL-CNA profile according to the risk group defined by the UKALL prognsotic index (PI_{UKALL})

PI_{UKALL} defined risk group	Number of cases with an CRLF2 Rearrangement [1]	Event Free Survival (95% CI)	Relapse Rate (95% CI)	Overall Survival (95% CI)
LR/SR	20 (57%)	89% (62-97)	11% (3-38)	94% (65-99)
IR/HR	15 (43%)	71% (40-88)	29% (12-60)	93% (59-99)
p value	-	0.3	0.3	0.4

PI_{UKALL} defined risk group	Number of cases with an IKZF1 deletion [1]	Event Free Survival (95% CI)	Relapse Rate (95% CI)	Overall Survival (95% CI)
LR/SR	62 (63%)	90% (79-95)	9% (5-21)	97% (87-99)
IR/HR	37 (37%)	64% (47-77)	30% (17-48)	78% (61-88)
p value	-	0.002	0.008	0.0005

PI_{UKALL} defined risk group	Number of cases with good risk UKALL-CNA profile [2]	Event Free Survival (95% CI)	Relapse Rate (95% CI)	Overall Survival (95% CI)
LR	259 (60%)	96% (92-98)	2% (1-5)	99% (96-100)
SR	104 (24%)	96% (90-98)	4% (1-10)	99% (93-100)
IR	61 (14%)	87% (75-93)	12% (6-23)	92% (81-96)
HR	7 (2%)	69% (21-91)	31% (9-79)	No deaths
p value		0.003	0.0003	0.002

PI_{UKALL} defined risk group	Number of cases with poor risk UKALL-CNA profile [2]	Event Free Survival (95% CI)	Relapse Rate (95% CI)	Overall Survival (95% CI)
LR	118 (47%)	96% (90-98)	3% (1-9)	97% (92-99)
SR	64 (25%)	94% (84-98)	5% (6-14)	95% (86-98)
IR	57 (23%)	70% (56-80)	25% (15-39)	80% (67-88)
HR	12 (5%)	55% (23-79)	44% (21-77)	67% (34-86)
p value		<0.0001	<0.0001	<0.0001

Notes: 1) DCOG-ALL10 and UKALL2003 patients only; 2) UKALL2003 patients only