

An Integrative Scoring System for Survival Prediction Following Umbilical Cord Blood Transplantation in Acute Leukemia



Roni Shouval^{1,2,3,4}, Annalisa Ruggeri^{5,6}, Myriam Labopin^{5,7}, Mohamad Mohty^{5,8}, Guillermo Sanz⁹, Gerard Michel¹⁰, Jürgen Kuball¹¹, Patrice Chevallier¹², Amal Al-Seraihy¹³, Noel-Jean Milpied¹⁴, Cristina Diaz de Heredia¹⁵, William Arcese¹⁶, Didier Blaise¹⁷, Vanderson Rocha^{6,18}, Joshua Fein^{1,3}, Ron Unger⁴, Frederic Baron¹⁹, Peter Bader^{20,21}, Eliane Gluckman^{6,22}, and Arnon Nagler^{1,3,7,8}

Abstract

Purpose: Survival of acute leukemia (AL) patients following umbilical cord blood transplantation (UCBT) is dependent on an array of individual features. Integrative models for risk assessment are lacking. We sought to develop a scoring system for prediction of overall survival (OS) and leukemia-free survival (LFS) at 2 years following UCBT in AL patients.

Experimental Design: The study cohort included 3,140 pediatric and adult AL UCBT patients from the European Society of Blood and Marrow Transplantation and Eurocord registries. Patients received single or double cord blood units. The dataset was geographically split into a derivation ($n = 2,362$, 65%) and validation set ($n = 778$, 35%). Top predictors of OS were identified using the Random Survival Forest algorithm and introduced into a Cox regression model, which served for the construction of the UCBT risk score.

Results: The score includes nine variables: disease status, diagnosis, cell dose, age, center experience, cytomegalovirus serostatus, degree of HLA mismatch, previous autograft, and anti-thymocyte globulin administration. Over the validation set an increasing score was associated with decreasing probabilities for 2 years OS and LFS, ranging from 70.21% [68.89–70.71, 95% confidence interval (CI)] and 64.76% (64.33–65.86, 95% CI) to 14.78% (10.91–17.41) and 18.11% (14.40–22.30), respectively. It stratified patients into six distinct risk groups. The score's discrimination (AUC) over multiple imputations of the validation set was 68.76 (68.19–69.04, range) and 65.78 (65.20–66.28) for 2 years OS and LFS, respectively.

Conclusions: The UCBT score is a simple tool for risk stratification of AL patients undergoing UCBT. Widespread application of the score will require further independent validation. *Clin Cancer Res*; 23(21): 6478–86. ©2017 AACR.

¹Hematology and Bone Marrow Transplantation, Chaim Sheba Medical Center, Tel-HaShomer, Ramat-Gan, Israel. ²Dr. Pinchas Bornstein Talpiot Medical Leadership Program, Chaim Sheba Medical Center, Tel-HaShomer, Ramat-Gan, Israel. ³Sackler Faculty of Medicine, Tel-Aviv University, Israel. ⁴Bar-Ilan University, Ramat Gan, Israel. ⁵Service d'Hématologie Clinique et de Thérapie Cellulaire, Hôpital Saint-Antoine, Paris, France. ⁶Eurocord, Hospital Saint Louis, Paris, France. ⁷Acute Leukemia Working Party EBMT Paris Office, Hôpital Saint-Antoine, Paris, France. ⁸Université Pierre et Marie Curie, Paris, France. ⁹Hospital Universitario La Fe, Valencia, Spain. ¹⁰Timone Enfants Hospital and Aix-Marseille University, Marseille, France. ¹¹University Medical Centre Utrecht, Utrecht, the Netherlands. ¹²CHU Nantes, Nantes, France. ¹³King Faisal Specialist Hospital & Research Center, Riyadh, Saudi Arabia. ¹⁴Hôpital du Haut Leveque, Bordeaux, France. ¹⁵Hospital Vall d'Hebron, Barcelona, Spain. ¹⁶University of Rome 'Tor Vergata', Rome, Italy. ¹⁷Institut Paoli Calmettes, Marseille, France. ¹⁸Churchill Hospital, Oxford, United Kingdom. ¹⁹University of Liege, Liege, Belgium. ²⁰Pediatric Diseases Working Party of the EBMT, Barcelona, Spain. ²¹University Hospital Frankfurt, Frankfurt, Germany. ²²Monacord, Centre Scientifique de Monaco, Monaco, France.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Corresponding Author: Roni Shouval, Sheba Medical Center, Tel-HaShomer, Ramat-Gan 5265601, Israel. Phone: 972-52-6668162; Fax: 972-3-530211; E-mail: shouval@gmail.com.

doi: 10.1158/1078-0432.CCR-17-0489

©2017 American Association for Cancer Research.

Introduction

Umbilical cord blood (UCB) is an established alternative source of hematopoietic stem cells for allogeneic transplantation when suitable HLA-matched sibling, or well-matched unrelated donors are unavailable (1). It may offer a cure for patients with acute leukemia (AL). Mounting experience with unrelated UCB transplantation (UCBT), modifications of conditioning regimens, and better choice of the UCB unit according to cell dose and HLA typing have led to improved outcomes (2). Various studies have linked individual parameters such as recipient's disease status, cell dose, degree of HLA match, graft type (single vs. double), and anti-thymocyte globulin (ATG) administration with mortality following UCBT (3–11). However, it remains unclear how these parameters should be best combined to optimize transplantation outcomes and obtain prognostic information. Furthermore, integrative models for prediction of UCBT outcomes are lacking.

We investigated a series of UCBT parameters, evaluating their individual and cumulative predictive weight in the prediction of overall survival (OS) at 2 years following transplantation in a large cohort of AL patients. Based on the top predictors, a risk score for 2 years OS and leukemia-free survival (LFS) was constructed.

Translational Relevance

Umbilical cord blood transplantation (UCBT) is a curative treatment for acute leukemia. Several individual features have been associated with outcomes following transplantation, but it remains unclear how these parameters should be best combined to predict outcomes. We have developed the first integrative scoring system for prediction of overall survival and leukemia-free survival following UCBT in acute leukemia patients. The score stratifies patients into six distinct risk groups by weighing patient, disease, donor, and transplantation-related features. Potential applications include pretransplant risk assessment and stratification, interpretation and analysis of retrospective data, patient counseling during informed consent sessions, and tailoring transplant regimens or referring to alternative treatments according to transplantation risk.

Materials and Methods

Source of data and participants

This was a retrospective prognostic modeling study. The data source was based on information reported to Eurocord and the Acute Leukemia Working Party, Paediatric Disease Working Party, and Cord Blood Committee of the European Society for Blood and Marrow Transplantation (EBMT). The EBMT registry is a voluntary working group of more than 500 transplant centers, whose participants routinely report data on patients undergoing hematopoietic stem cell transplantation (HSCT). Eurocord collects data on UCBT performed in >50 countries, mainly in EBMT centers. The population selection criteria included children and adults undergoing an UCBT (single/double unit) between the years 2004 and 2014, as a treatment for *de novo* or secondary AL in all disease statuses. Patients receiving myeloablative (MAC) or reduced intensity conditioning regimen (RIC) were included. MAC was defined as a regimen containing either total body irradiation (TBI) with a dose greater than 6 Gy, a total dose of oral Busulfan greater than 8 mg/kg, or a total dose of intravenous Busulfan greater than 6.4 mg/kg (11). Patients who had a previous allogeneic HSCT were excluded. All patients or legal guardians provided informed consent for UCBT according to the Declaration of Helsinki. The review board of Eurocord/EBMT approved this study.

To maximize the power and generalizability of the results, all available patients on the registry were included, provided they meet inclusion criteria and had no missing data on the measured outcome. Geographical splitting (i.e., random selection of the center's country; Supplementary Table S1) was applied to generate the derivation ($n = 2,362$, 65%) and validation datasets ($n = 778$, 35%) from the original cohort (i.e., geographical validation; ref. 12).

Predictors and outcomes

Nineteen variables detailing patient disease and UCB characteristics were considered (Table 1). These include, recipients' age, sex, recipient cytomegalovirus (CMV) serostatus, and Karnofsky/Lansky performance status at UCBT (<80 , ≥ 80), diagnosis [acute myeloid leukemia (AML) or acute lymphoblastic leukemia (ALL)], disease status [first complete remission (CR), second CR,

>second CR, advanced (i.e., active disease), months from diagnosis to transplantation, cytogenetics (good, intermediate, poor, or secondary AL); refs. 13, 14], previous autologous HSCT, graft type [single or double cord blood unit (sCBU or dCBU), respectively], total nucleated cells/kg body weight (TNC) at cryopreservation, degree of HLA mismatch as defined by antigen or allelic level DNA typing (≤ 1 , >1), donor-recipient ABO blood group match (major incompatibility, other), female donor to male recipient, mycophenolate mofetil for GVHD prophylaxis, ATG administration, and center experience as measured by the annual number of UCBT done in the individual center and reported to Eurocord/EBMT.

The primary and secondary predictive objectives were prediction of OS and LFS at 2 years following UCBT, respectively. An OS event was defined as death from any cause. Patients were censored if alive at 2 years or last follow-up. LFS was defined as survival with no evidence of relapse or progression. Probabilities of LFS and OS were calculated using Kaplan-Meier estimates; differences between groups were evaluated using the log-rank test. Cumulative incidence functions were used to estimate relapse incidence and non-relapse-related mortality (NRM) in a competing risks setting because death and relapse compete; differences between groups were evaluated using the Fine and Gray method.

Statistical analysis pipeline

Briefly, the analysis pipeline comprised of sequential stages: (i) preprocessing- data quality assurance and multiple imputations of missing values; (ii) estimation of predictor importance using the random survival forest (RSF) for feature selection; (iii) interaction analysis using RSF and Cox modelization; (iv) construction of the risk score. Principles for prognostic model development were adapted from the Transparent Reporting of a prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement (12).

Analyses were performed with SPSS version 21 (SPSS Inc./IBM) and R version 3.2.1 (R Development Core Team) including the *mi*, *randomForestSRC*, *ggRandomForests*, and *survival* packages.

Missing data

We assumed missing data occurred at random depending on the clinical variables and UCBT outcomes, and performed multiple imputations using chained equations (15). Multiple imputation allows to introduce the variability of imputed data to find a range of possible responses from which to work from. It has been shown to be an effective way of handling missing data and minimizes bias and loss in precision that may often result from excluding such patients. In addition, multiple imputation remains valid even if the proportion of missing data is large (12). Missing values were predicted for the entire cohort on the basis of all predictors listed in Table 1, demographical information on the transplant center, and patients overall survival (16). A total of 20 imputations were performed on the entire cohort, producing 20 distinct datasets.

Feature selection

The RSF algorithm was used to identify predictive features of 2 years OS following UCBT. RSF is an extension of random forest machine learning algorithm to analyze right-censored time-to-event data (17). A forest of survival trees is grown using a log-rank splitting rule to select the optimal candidate variables. A survival

Table 1. Patients characteristics

Variable		Value	Missing (%)
UCBT year	Median (IQR)	2,009 (2007–2012)	
Age at UCBT (years)	Median (IQR)	21.9 (7.3–43.8)	0 (0)
	<18	1,395 (44.4)	
Gender	≥18	1,745 (55.6)	23 (1)
	Male	1,693 (54.3)	
Karnofsky/Lansky PS	Female	1,424 (45.7)	765 (24)
	<80%	110 (4.6%)	
CMV serostatus	≥80	2,265 (95.4%)	318 (10)
	Negative	1,083 (38.4)	
Diagnosis	Positive	1,739 (61.6)	0 (0)
	ALL	1,397 (44.5)	
Cytogenetics	AML	1,743 (55.5)	887 (28)
	Good	129 (5.7)	
Months from diagnosis to UCBT	Intermediate/poor	1,946 (86.4)	
	Secondary AL	178 (7.9)	
	Median (IQR)	10.5 (5.82–22.44)	83 (3)
Previous autograft	≤12	1,668 (54.6)	
	>12	1,389 (45.4)	0 (0)
Disease status	No	2,944 (93.8)	
	Yes	196 (6.2)	197 (6)
Graft	First CR	1,332 (45.3)	
	Second CR	1,071 (36.4)	
	Other CR	159 (5.4)	
	Advanced disease	381 (12.9)	0 (0)
HLA mismatch	Single CB unit	2,139 (68.1)	
	Double CB unit	1,001 (31.9)	640 (20)
ABO major vs. other	≤1	1,179 (47.2)	
	>1	1,321 (52.8)	0 (0)
Female donor to male recipient	Other	2,281 (72.6)	
	Major incompatibility	859 (27.4)	89 (3)
TNCs cryopreserved ($\times 10^7$ cells/kg)	No	2,182 (71.9)	
	Yes	959 (31.4)	798 (25)
Conditioning	Median (IQR)	5 (3.8–6.9)	
	<3	234 (10)	
ATG	≥3	2,108 (90)	106 (3)
	MAC	2,182 (71.9)	
Mycophenolate mofetil	RIC	852 (28.1)	307 (10)
	No	1,064 (37.6)	
Center experience (UCBT/year)	Yes	1,769 (64.2)	384 (12)
	No	1,226 (44.5)	
Center experience (UCBT/year)	Yes	1,530 (55.5)	
	Median (IQR)	30 (12–53)	0 (0)
	<20	1,150 (36.6)	
	≥20	1,990 (63.4)	

Abbreviation: PS, performance status.

estimate for each observation is constructed with a Kaplan–Meier estimator within each terminal node and at each event time. RSF models can be used for prediction and knowledge extraction, including variable ranking and recovery of nonlinear effects and interactions, as they are fully non-parametric (18, 19). The predictive accuracy of an RSF model is assessed by the Harrell concordance (C)-index. It is a measure of discrimination

obtained by repetitive bootstrapping (1,000 iterations) when constructing the forest and is conceptually similar to the area under the receiver-operating characteristic curve (AUC), ranging from 0.5 to 1.

Minimal depth, a property derived from the construction of each tree within the forest can be used for predictor ranking. Minimal depth assumes that variables with high impact on the

prediction are those that most frequently split nodes nearest to the trunks of the trees (i.e., at the root node) where they partition large samples of the populations. Predictiveness of a variable is inversely related to the value of minimal depth. Smaller minimal depth value of a variable, the greater is its predictive impact (20). RSF and minimal depth are further explained in Appendix 1.

For all 20 versions of the imputed derivation datasets, we constructed comprehensive RSF models, using all available variables, for prediction of 2 years OS. The importance of variables in each model was determined using the minimal depth method. A list of variables, ordered by average importance across models is presented in Supplementary Fig. S1.

A threshold for variable importance was determined through a nested RSF approach (21), estimating a cut-off for predictive contribution by serially introducing variables into RSF models according to their importance, while measuring improvement in the C-index. For instance, in the first iteration, the top ranking variable was introduced; in the second, the top two variables; and so on. Variables were considered informative as long as the predictive performance improved.

Interaction analysis

Interactions were analyzed using partial dependence conditional plots (coplots). Partial dependence plots are generated by isolating the effects of variables other than the covariate of interest. This provides a qualitative, risk-adjusted visualization of the nature (e.g., linear, nonlinear, etc.) of a variable's effect on predicted response (17, 18). Coplots are in fact partial dependence plots conditioned on group membership (18). All interactions discovered by coplots were also validated by Cox modelization ($P < 0.1$). Interacting variables were combined to form groups based on joint values of the interacting variables.

Risk score construction

The top predictors and grouped interactions were introduced into a Cox model. Hazard ratio estimates were pooled over 20 imputed derivation datasets using Rubin's rules (22). The weight for each predictor was determined according to the hazard ratio ($HR = 1-1.49 = 1$; $HR = 1.5-2.49 = 2$; $HR = 2.5-3.49 = 3$), providing $P < 0.1$. The score was categorized, and intervals representing outliers were grouped. The associated risk for 2 years OS, LFS, NRM, and relapse incidence was calculated and plotted on a randomly selected imputed derivation dataset and its corresponding validation set (i.e., the derivation and validation sets were produced in the same iteration of the multiple imputation process). Calibration and discrimination, assessed with the time-dependent AUC (12, 23), were used to evaluate the score's quality.

Results

Patient characteristics and outcomes

Characteristics of 3,140 analyzed patients are listed in Table 1. Median age was 21.9 years (interquartile range [IQR]: 7.3–43.8). The majority of patients had AML (55.5%) were in first CR (CR1; 45.2%) and received MAC (71.9%). Grafts were mainly derived from a sCBU (68.1%) and had at least two HLA mismatches (52.8%). Median follow-up was 30 months. The 2 years OS, NRM, relapse incidence, and LFS rates were 47.7% (95% CI, 45.8–49.6), 29.9% (95% CI, 28.2–31.6), 26.8% (95% CI, 25.1–28.4) and 43.4% (95% CI, 41.5–45.2), respectively.

Predictor identification and ranking

Minimal depth, a dimensionless order statistic measuring the predictiveness of a variable in a survival tree, was used to estimate variable importance across the derivation datasets (Appendix 1). Disease status was consistently the most influential factor, followed by age and cell dose (TNCs cryopreserved; Supplementary Fig. S1). To determine a cut off for the variables' cumulative predictive contribution (i.e., the minimum number of variables required maximal predictive performance), a nested RSF approach was applied (21). RSF prediction models were constructed on serially introduced variables, according to their minimal depth ranking. Predictive performance on the derivation datasets, as measured by the C-index, incrementally improved until reaching a plateau after introducing the top 10 ranked variables (disease status, TNCs cryopreserved, age, center experience, interval from diagnosis to UCBT, year of UCBT, CMV serostatus, degree of HLA mismatch, previous autograft, and ATG administration; Fig. 1).

The UCBT risk score

For the score construction, we introduced the top predictors of 2 years overall survival into a Cox multivariate model. Age, cell dose (TNCs cryopreserved) and center experience were categorized into two groups according to acceptable thresholds (7, 24). Although not considered among top predictors, we included diagnosis into the model, as it carries important clinical information. The score was developed on all 20 versions of the imputed derivation datasets. In a preliminary phase, a randomly selected version of one of the imputed derivation datasets was used to analyze interactions; coplots derived from a RSF model showed that increasing age was associated with inferior predicted survival in the groups receiving ATG and that the effect of disease status was diagnosis dependent (Supplementary Fig. S2). These interactions were validated in a Cox model (Tables S2 and S3).

The UCBT risk score was derived from pooled estimates of Cox models over multiple imputed datasets (Table 2; ref. 22). The models included independent predictors and variables grouping interactions (ATG administration conditioned on age group and disease status conditioned on diagnosis). The variable year of UCBT was kept in the models for covariate adjustment, but excluded from the score. Duration from diagnosis to UCBT was not included in the Cox model as it is highly dependent on disease and disease status, resulting in complex three-way interactions with 16 possible combinations. Furthermore, its addition to the Cox models had minimal impact on the model's performance (i.e., the difference in the Akaike's Information Criterion between the two models was 0.04%).

On a randomly selected derivation dataset and its corresponding validation set (patients' characteristics are listed in Table S4), the UCBT score ranged from 0 to 8 (Table 3). Calibration was excellent (Supplementary Fig. S3). An increasing score corresponded with increasing hazard and decreasing probability for 2 years OS and LFS (log-rank $P < 0.001$ for both, Table 4 and Supplementary Table S5). Over the validation cohort, probabilities of 2 years OS and LFS ranged from 70.21% (95% CI, 68.89–70.71) and 64.76% (95% CI, 64.33–65.86), respectively, for patients having a score of 0 to 1, to 14.78% (95% CI, 10.91–17.41) and 18.11% (95% CI, 14.40–22.30) for a score of 6 to 8 (Fig. 2). Increasing scores were also associated with a greater probability for 2 years NRM, ranging from 14.10% to 56.45%. However, considerable overlap was noted between patients

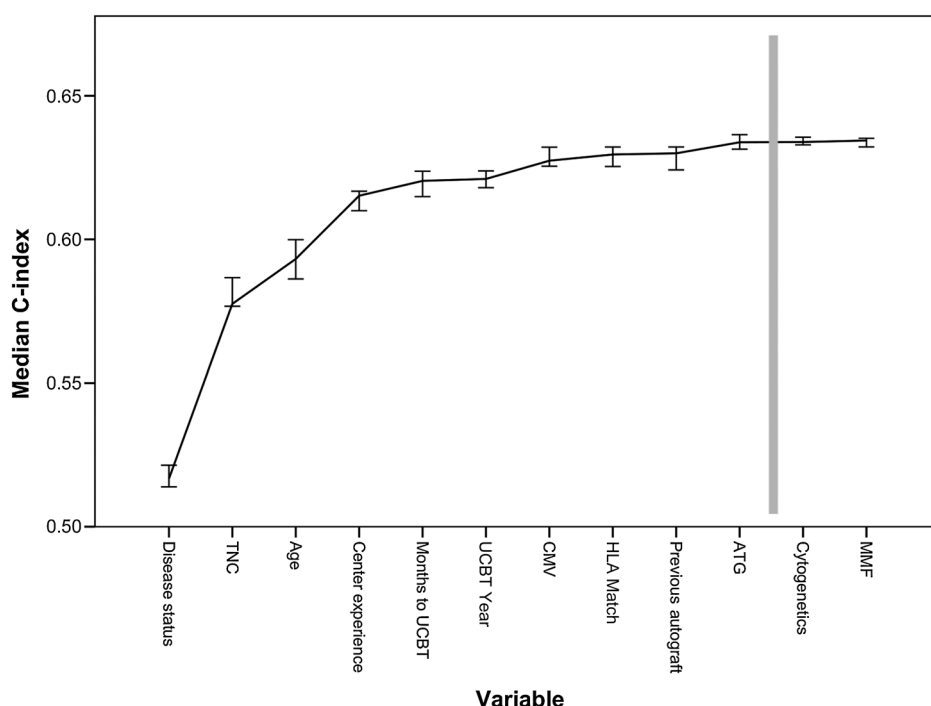


Figure 1. Cumulative predictive value of variables. Variables are ordered by predictive value as assessed with the minimal depth method. Performance (c-index) of nested Random Survival Forest models, trained with an increasing number of variables, is plotted on the y-axis. Error bars indicate 95% confidence intervals over the imputed derivation datasets. Improvement in performance was observed when the top 10 variables were added to the models, whereas other variables (to the right of the gray line) did not further improve the accuracy of prediction.

assigned scores of 3 to 5 (Table 4). Similar results were obtained when analyzing all versions of the imputed validation datasets (Supplementary Table S6). Patients assigned with a score of 0 to 1 were more likely to be in first CR (78.1%), receive 3×10^7 /kg or more TNCs (99.4%), have donors with 0 to 1 HLA mismatches (84.3%), and be transplanted in centers performing a high number of UCBT/year (89.6%). Very high-risk patients, with scores of 6 to 8 were transplanted mainly in late CRs or advanced disease (87.3%), be adults receiving ATG (72.0%), have a donor with over one HLA mismatch (79.9%), and transplanted in center performing 20 or less UCBT/year (60.8%; Table S7). The score's median discrimination (AUC) for 2 years OS overall imputation of the

Table 2. Association between predictors and 2 years OS over 20 imputed derivation datasets

Variable	Hazard ratio (95% CI)	P-value
Recipient CMV serostatus (positive vs. negative)	1.28 (1.12-1.47)	<0.001
Previous autograft (yes vs. no)	1.39 (1.12-1.69)	0.002
HLA mismatch (>1 vs. ≤1)	1.22 (1.07-1.37)	0.003
TNC (≥3 vs. <3 × 10 ⁷ /kg)	1.19 (0.98-1.43)	0.073
Center experience (UCBT/year) (≥20 vs. <20)	1.17 (1.04-1.33)	0.011
UCBT year	0.97 (0.95-0.97)	0.006
Age and ATG (reference: <18 years and no ATG)		
<18 years and ATG	1.04 (0.81-1.35)	0.715
≥18 years and no ATG	0.99 (0.76-1.28)	0.831
≥18 years and ATG	1.36 (1.05-1.77)	0.018
Diagnosis and disease status (reference: ALL and CR1)		
ALL and CR2	1.37 (1.11-1.70)	0.004
ALL and other CR	2.53 (1.83-3.50)	<0.001
ALL and advanced	3.07 (2.32-4.06)	<0.001
AML and CR1	1.14 (0.93-1.39)	0.214
AML and CR2	1.24 (1.00-1.55)	0.051
AML and other CR	1.61 (1.04-2.50)	0.034
AML and advanced	2.64 (2.1-3.31)	<0.001

Abbreviation: CI, confidence interval.

Table 3. The UCBT risk score

	Score points	Derivation set (%) ^a	Validation set (%) ^a
Recipient CMV serostatus			
Negative	0	887 (37.6)	362 (46.5)
Positive	1	1,475 (62.4)	416 (53.5)
Previous autograft			
No	0	2,191 (92.8)	753 (96.8)
Yes	1	171 (7.2)	25 (3.2)
HLA mismatch			
≤1	0	1,110 (47)	440 (56.6)
>1	1	1,252 (53)	338 (43.4)
TNC (×10 ⁷ /kg)			
TNC≥3	0	2,088 (88.4)	722 (92.8)
TNC<3	1	274 (11.6)	56 (7.2)
Center experience (UCBT/year)			
>20	0	1,632 (69.1)	358 (46)
≤20	1	730 (30.9)	420 (54)
Age and ATG			
Other	0	1,601 (67.8)	630 (81)
≥18 years and receiving ATG	1	761 (32.2)	148 (19)
Diagnosis and disease status			
ALL and CR1, AML and CR1	0	1,059 (44.8)	318 (40.9)
ALL and CR2, AML and CR2	1	828 (35.1)	304 (39.1)
AML and other CR	2	44 (1.9)	19 (2.4)
ALL and other CR, ALL and advanced, AML and advanced	3	431 (18.2)	137 (17.6)
AML and advanced			
0-1		521 (22.1)	178 (22.9)
2		559 (23.7)	208 (26.7)
3		496 (21.0)	154 (19.8)
4		344 (14.6)	110 (14.1)
5		229 (9.7)	73 (9.4)
6-8		213 (9.0)	55 (7.1)

^aThe derivation and validation data sets were generated in the same iteration of the multiple imputation process (i.e., they represent 1 out of 20 versions of imputed data sets).

Table 4. Probabilities for OS, LFS, NRM, and RI stratified by the UCBT score

Score	Training Set						Validation Set					
	1-year OS (95% confidence interval)	1-year LFS (95% confidence interval)	1-year NRM (95% confidence interval)	1-year RI (95% confidence interval)	2-years OS (95% confidence interval)	2-years LFS (95% confidence interval)	2-years NRM (95% confidence interval)	2-years RI (95% confidence interval)	2-years OS (95% confidence interval)	2-years LFS (95% confidence interval)	2-years NRM (95% confidence interval)	2-years RI (95% confidence interval)
0-1	73.3 (69.1-77.7)	66 (61.6-70.8)	13.6 (10.7-17.3)	20.4 (16.9-24.8)	67.3 (62.8-72.0)	60.8 (56.1-65.8)	14.7 (11.7-18.6)	24.5 (20.59-29.16)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)
2	64.5 (60.3-69.0)	58.9 (54.6-63.5)	22.5 (19.1-26.6)	18.6 (15.4-22.5)	56.0 (51.5-60.8)	50.6 (46.2-55.5)	25.5 (21.9-29.8)	23.86 (20.2-28.17)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)
3	58.3 (53.9-63.1)	53.1 (48.7-58.0)	28.4 (24.5-32.9)	18.5 (15.2-22.5)	49.3 (44.7-54.4)	45.6 (41.1-50.7)	31.9 (27.8-36.6)	22.48 (18.82-26.84)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)
4	46.5 (41.7-51.8)	43.2 (38.6-48.5)	33.2 (28.9-38.2)	23.6 (19.7-28.2)	38.8 (34.0-44.2)	36.9 (32.2-42.2)	36.2 (31.7-41.4)	26.92 (22.76-31.84)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)
5	36.0 (30.3-42.8)	32.5 (27.0-39.2)	34.4 (28.9-41.0)	33.1 (27.6-39.7)	27.6 (22.1-34.5)	26.4 (21.1-33.0)	36.6 (30.9-43.34)	36.98 (31.2-43.9)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)
6-8	25.9 (20.7-32.3)	22.1 (17.3-28.2)	43.1 (37.2-50.0)	34.8 (29.2-41.6)	17.0 (12.5-23.0)	15.9 (11.6-21.8)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)
Score	Training Set						Validation Set					
0-1	78.7 (72.1-85.9)	73.2 (66.1-81.0)	10.1 (6.16-16.7)	16.7 (11.5-24.3)	71.7 (64.3-79.9)	68.0 (60.4-76.5)	12.8 (8.14-20.0)	19.3 (13.6-27.3)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)
2	69.5 (63.1-76.5)	63.1 (56.5-70.6)	17.5 (12.8-23.9)	19.3 (14.3-26.1)	57.7 (50.7-65.8)	55.3 (48.3-63.4)	21.7 (16.4-28.8)	23.0 (17.4-30.3)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)
3	60.3 (53.0-68.7)	57.8 (50.4-66.3)	25.6 (19.5-33.6)	16.6 (11.6-23.8)	52.8 (45.2-61.8)	48.8 (41.2-57.8)	29.8 (23.2-38.4)	21.4 (15.6-29.3)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)
4	54.1 (45.8-63.9)	50.3 (42.0-60.1)	26.4 (19.6-35.5)	23.4 (16.8-32.5)	44.5 (35.9-54.8)	44.8 (36.4-55.0)	29.8 (22.5-39.5)	25.5 (18.6-34.9)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)
5	40.2 (30.7-52.8)	37.7 (28.4-49.9)	37.6 (28.4-49.9)	24.7 (16.8-36.3)	31.2 (21.9-44.4)	29.3 (20.4-41.9)	39.3 (29.8-51.8)	31.4 (22.3-44.2)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)
6-8	20.0 (11.7-34.3)	18.5 (9.5-32.7)	51.4 (39.7-66.6)	30.1 (20.0-45.4)	13.0 (6.2-27.0)	14.0 (7.1-27.9)	53.4 (41.6-68.6)	32.5 (21.9-48.2)	44.84 (38.79-51.84)	39.2 (33.3-46.3)	39.2 (33.3-46.3)	39.2 (33.3-46.3)

derivation and validation datasets was 68.76 (range = 68.19-69.04) and 68.12 (range = 67.43-68.69), respectively. Similarly, for 2 years LFS the median AUC was 66.68 (range = 66.05-66.98) and 65.78 (range = 65.20-66.28), respectively.

Discussion

Umbilical cord blood transplantation is a valid alternative source of stem cells in AL patients (25). Nonetheless, the concerns associated with UCBT, including an increased risk of graft failure, delayed immune reconstitution, and unavailability of the donor for additional donations warrants careful evaluation of transplantation candidates. Motivated by the need for a predictive prognostic model in UCBT, we have developed and internally validated, a risk score for 2 years overall survival and LFS following UCBT in patients with AL. The new UCBT score is based on nine variables; age, diagnosis, disease status, previous autologous HSCT, recipient CMV serostatus, HLA matching, cell dose, ATG administration, and center experience (annual UCBT/center). This is the first risk score developed specifically for UCBT. The score demonstrates distinctiveness and monotonicity; it categorizes patients' survival to unique groups, and an increasing score is associated with decreased OS and LFS.

Several scores, including the EBMT risk score, the hematopoietic cell transplantation comorbidity index (HCT-CI) and the Pre-transplantation Assessment of Mortality (PAM) score have been developed in the allogeneic HSCT setting (26-28). These have been validated and could be incorporated in the therapeutic algorithm of HSCT candidate assessment. Prior to the current study, integrative prognostic models dedicated to AL UCBT recipients were lacking. The new UCBT risk score may serve as a prognostic tool for patient stratification, interpretation of retrospective and prospective studies, and potentially for treatment allocation.

The methodological process of the UCBT risk score construction was unique, involving an initial screening stage for predictors using the RSF machine learning technique, followed by incorporation of the selected variables into a standard Cox regression model. The integration of the two approaches is appealing, as RSF is a nonparametric method, successfully applied in a variety of clinical scenarios (19-21). It accounts for censored data, refrains for assumptions on data distribution, and allows for feature selection, whereas Cox modeling is easy to interpret and frequently used in clinical scenarios. Overall, key principles from the TRIPOD guidelines for predictive modeling were followed, promoting standardization and transparency in prognostic research (12).

The determinants identified by RSF and incorporated into the UCBT score recapitulates established risk factors in UCBT. Disease status remains the most powerful predictor (5, 24, 26), and is diagnosis specific. The accepted threshold for graft selection, having at least 3×10^7 TNC/kg and the importance of a low degree of HLA mismatch were corroborated (7, 29). Interestingly, an age-dependent effect of ATG was noted. Similar to findings reported by Pascal et al., ATG had a detrimental impact among adults in our cohort (8). However, in children, no harm was discovered. It is likely that adults are more susceptible than children to the risks ATG-related immune suppression (e.g., immune reconstitution, infections, and posttransplant lymphoproliferative disorder). However, given ATG's potential role in reducing GVHD, there is a need for prospective evaluation of its

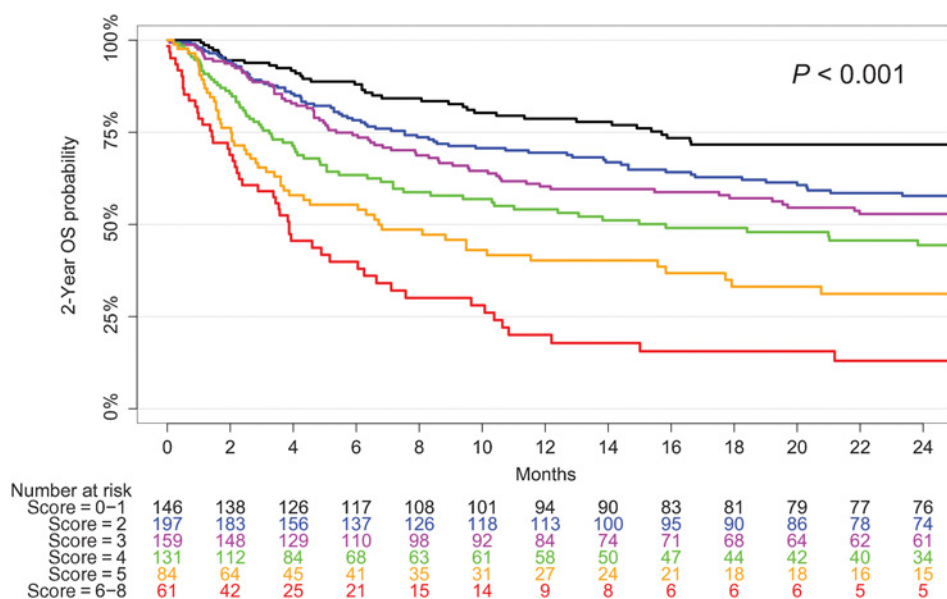
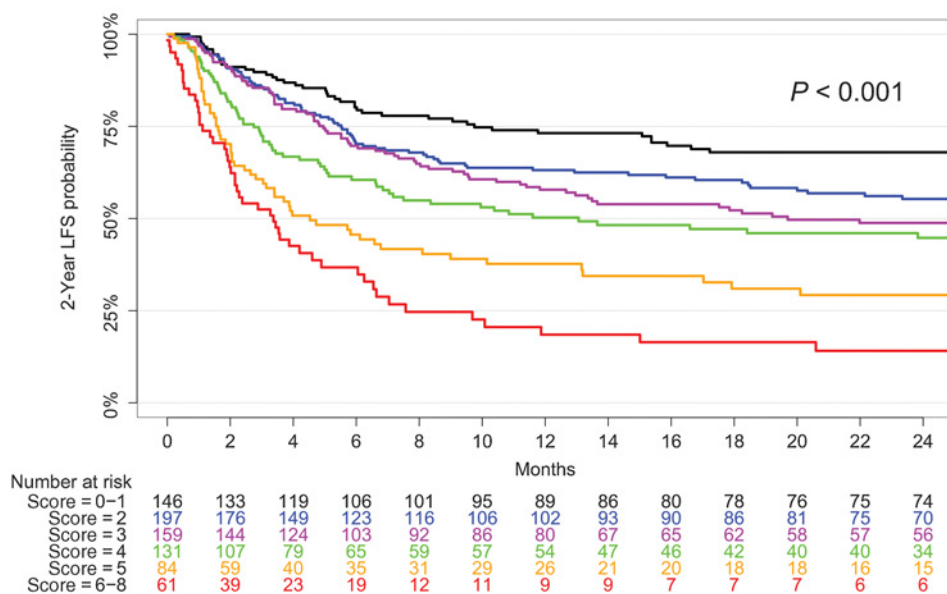


Figure 2. Survival and LFS by the UCBT risk score.



role (8, 30, 31). In line with previous publications, the conditioning regimen and graft type (single/double unit) were not identified as risk factors (32–34), most likely reflecting the proper selection of condition type and cell dose. Overall, the score includes the main clinical risk factors impacting on UCBT outcomes. Nonetheless, the score captures the expanding corpus of UCBT knowledge, allowing for an integrative evaluation of transplantation risk.

When dissecting the factors discriminating between patients in differing risk score categories (Supplementary Table S7), the importance of center experience is striking. Almost 90% of patients receiving a score of 0 to 1 were transplanted in centers performing 20 or more transplants per year, whereas in patients assigned with scores of 6 to 8, only a minority (39.2%) were

transplanted in well-experienced centers. These findings stress the importance of center experience and accreditation (35–37). Nevertheless, patients' profile could be linked to experience, thereby introducing a selection bias; centers performing a low volume of UCBT might resort to UCBT as last option (e.g., when haploidentical or 9/10 HLA unrelated donors are not available), whereas high-volume centers may choose to pursue UCBT as a valid option early on. Other discriminating factors were CMV seropositivity (88.4%) and two or more than one HLA mismatch (79.9%) in the group with the highest risk. Patients with scores over 5 should be carefully evaluated taking into account the associated increase in risk.

The study has several limitations. First, the UCBT score was not validated in an external cohort. Nonetheless, geographical

validation (i.e., the transplantation centers' countries in the derivation and validation dataset differed) was used to ensure generalizability of the results (12). Also, the UCBT score weights were assigned by pooling Cox estimations over multiple versions of imputed derivation datasets (22). Second, in the registry certain granular data are missing at least in part (e.g., disease genetic features.) which could further improve models. Third, the analysis was directed towards factors affecting 2 years OS and LFS. However, NRM incidence is of high interest, especially when comparing to alternative therapies. Fourth, missing data were imputed on the entire cohort; interdependencies between the derivation and validation cohort in the process of imputation could lead to overoptimistic estimations. However, in multiple imputation algorithms reliable prediction of missing data values is dependent on the quality and extent of available data. Therefore, we opt to maximize data exposure in the imputation process and took measures to reduce the risk for overfitting, as discussed above. Finally, categorization of continuous variables introduced into the Cox model, which was used for model construction, might have led to loss of predictive information. Nevertheless, categorization promotes simplicity. Furthermore, the score was rather discriminative, despite this transformation.

In conclusion, we have developed and internally validated the first risk score for stratification of overall survival and LFS in AL patients undergoing a UCBT. External validation is warranted for widespread application. The score is simple and stratifies patients into distinct risk groups. Its potential applications include pre-transplant risk assessment and stratification, interpretation and analysis of retrospective data, patient counseling during informed consent sessions, and tailoring transplant regimens or referring to alternative treatments according to transplantation risk. The score's discrimination (i.e., AUC) is comparable to similar prognostic models in HSCT (26, 28, 38); integration of detailed data on comorbidities, transplant regimens, and the genetic features

of the disease may further enhance predictive accuracy, allowing for individualized prediction rather than stratification.

Disclosure of Potential Conflicts of Interest

J. Kuball is a consultant/advisory board member for Gadeta and reports receiving commercial research grants from Gadeta, Miltenyi, and Novartis. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: R. Shouval, A. Ruggeri, M. Mohty, R. Unger, F. Baron, P. Bader, A. Nagler

Development of methodology: R. Shouval, R. Unger, A. Nagler

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A. Ruggeri, M. Mohty, G. Sanz, G. Michel, J. Kuball, P. Chevallier, N.-J. Milpied, C.D. de Heredia, W. Arcese, D. Blaise, V. Rocha, F. Baron, A. Nagler

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R. Shouval, A. Ruggeri, M. Labopin, G. Sanz, V. Rocha, J. Fein, P. Bader, A. Nagler

Writing, review, and/or revision of the manuscript: R. Shouval, A. Ruggeri, M. Labopin, M. Mohty, G. Sanz, G. Michel, J. Kuball, P. Chevallier, A. Al-Seraihi, N.-J. Milpied, C.D. de Heredia, W. Arcese, D. Blaise, R. Unger, F. Baron, P. Bader, E. Gluckman, A. Nagler

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M. Mohty, V. Rocha, A. Nagler

Study supervision: R. Unger, A. Nagler

Acknowledgments

This study was supported by The Varda and Boaz Dotan Research Center in Hemato-Oncology affiliated with the CBRC of Tel Aviv University and The Shalvi Foundation for the Support of Medical Research.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 18, 2017; revised May 25, 2017; accepted July 20, 2017; published OnlineFirst July 28, 2017.

References

- Milano F, Gooley T, Wood B, Woolfrey A, Flowers ME, Doney K, et al. Cord-blood transplantation in patients with minimal residual disease. *N Engl J Med* 2016;375:944-53.
- Ballen KK, Gluckman E, Broxmeyer HE. Umbilical cord blood transplantation: the first 25 years and beyond. *Blood* 2013;122:491-8.
- Eapen M, Rubinstein P, Zhang MJ, Stevens C, Kurtzberg J, Scaradavou A, et al. Outcomes of transplantation of unrelated donor umbilical cord blood and bone marrow in children with acute leukaemia: a comparison study. *Lancet* 2007;369:1947-54.
- Herr AL, Kabbara N, Bonfim CM, Teira P, Locatelli F, Tiedemann K, et al. Long-term follow-up and factors influencing outcomes after related HLA-identical cord blood transplantation for patients with malignancies: an analysis on behalf of Eurocord-EBMT. *Blood* 2010;116:1849-56.
- Cohen YC, Scaradavou A, Stevens CE, Rubinstein P, Gluckman E, Rocha V, et al. Factors affecting mortality following myeloablative cord blood transplantation in adults: a pooled analysis of three international registries. *Bone Marrow Transplant* 2011;46:70-6.
- Barker JN, Scaradavou A, Stevens CE. Combined effect of total nucleated cell dose and HLA match on transplantation outcome in 1061 cord blood recipients with hematologic malignancies. *Blood* 2010;115:1843-9.
- Eapen M, Klein JP, Ruggeri A, Spellman S, Lee SJ, Anasetti C, et al. Impact of allele-level HLA matching on outcomes after myeloablative single unit umbilical cord blood transplantation for hematologic malignancy. *Blood* 2014;123:133-40.
- Pascal L, Tucunduva L, Ruggeri A, Blaise D, Ceballos P, Chevallier P, et al. Impact of ATG-containing reduced-intensity conditioning after single- or double-unit allogeneic cord blood transplantation. *Blood* 2015;126:1027-32.
- Hough R, Danby R, Russell N, Marks D, Veys P, Shaw B, et al. Recommendations for a standard UK approach to incorporating umbilical cord blood into clinical transplantation practice: an update on cord blood unit selection, donor selection algorithms and conditioning protocols. *Br J Haematol* 2016;172:360-70.
- Scaradavou A, Brunstein CG, Eapen M, Le-Rademacher J, Barker JN, Chao N, et al. Double unit grafts successfully extend the application of umbilical cord blood transplantation in adults with acute leukemia. *Blood* 2013;121:752-8.
- Ruggeri A, Sanz G, Bittencourt H, Sanz J, Rambaldi A, Volt F, et al. Comparison of outcomes after single or double cord blood transplantation in adults with acute leukemia using different types of myeloablative conditioning regimen, a retrospective study on behalf of Eurocord and the Acute Leukemia Working Party of EBMT. *Leukemia* 2014;28:779-86.
- Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1-W73.
- Pui CH, Robison LL, Look AT. Acute lymphoblastic leukaemia. *Lancet* 2008;371:1030-43.
- Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 2010;115:453-74.

15. Su YS, Yajima M, Gelman AE, Hill J. Multiple imputation with diagnostics (mi) in R: opening windows into the black box. *J Stat Software* 2011;45: 1–31.
16. Moons KG, Donders RA, Stijnen T, Harrell FE Jr. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol* 2006;59:1092–101.
17. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat* 2008;2:841–60.
18. Ehrlinger J, Rajeswaran J, Blackstone EH. ggRandomForests: Exploring random forest survival. 2015. arXiv:1612.08974.
19. Hsieh E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS. Identifying important risk factors for survival in patient with systolic heart failure using random survival forests. *Circulation* 2011;4:39–45.
20. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-dimensional variable selection for survival data. *J Am Stat Assoc* 2010;105: 205–17.
21. Inuzuka R, Diller G-P, Borgia F, Benson L, Tay EL, Alonso-Gonzalez R, et al. Comprehensive use of cardiopulmonary exercise testing identifies adults with congenital heart disease at increased mortality risk in the medium term. *Circulation* 2012;125:250–9.
22. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol* 2009;9:57.
23. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 2013;32:5381–97.
24. Shouval R, Labopin M, Bondi O, Mishan-Shamay H, Shimoni A, Ciceri F, et al. Prediction of allogeneic hematopoietic stem-cell transplantation mortality 100 days after transplantation using a machine learning algorithm: a European group for blood and marrow transplantation acute leukemia working party retrospective data mining study. *J Clin Oncol* 2015;33:3144–51.
25. Ruggeri A, Labopin M, Sanz G, Piemontese S, Arcese W, Bacigalupo A, et al. Comparison of outcomes after unrelated cord blood and unmanipulated haploidentical stem cell transplantation in adults with acute leukemia. *Leukemia* 2015;29:1891–900.
26. Gratwohl A, Stern M, Brand R, Apperley J, Baldomero H, de Witte T, et al. Risk score for outcome after allogeneic hematopoietic stem cell transplantation a retrospective analysis. *Cancer* 2009;115:4715–26.
27. Parimon T, Au DH, Martin PJ, Chien JW. A risk score for mortality after allogeneic hematopoietic cell transplantation. *Ann Intern Med* 2006;144: 407–14.
28. Sorror ML, Maris MB, Storb R, Baron F, Sandmaier BM, Maloney DG, et al. Hematopoietic cell transplantation (HCT)-specific comorbidity index: a new tool for risk assessment before allogeneic HCT. *Blood* 2005; 106:2912–9.
29. Eapen M, Klein JP, Sanz GF, Spellman S, Ruggeri A, Anasetti C, et al. Effect of donor–recipient HLA matching at HLA A, B, C, and DRB1 on outcomes after umbilical-cord blood transplantation for leukaemia and myelodysplastic syndrome: a retrospective analysis. *Lancet Oncol* 2011;12: 1214–21.
30. Admiraal R, van Kesteren C, Jol-van der Zijde CM, Lankester AC, Bierings MB, Egberts TC, et al. Association between anti-thymocyte globulin exposure and CD4+ immune reconstitution in paediatric haemopoietic cell transplantation: a multicentre, retrospective pharmacodynamic cohort analysis. *Lancet Haematol* 2015;2:e194–e203.
31. Lindemans CA, Chiesa R, Amrolia PJ, Rao K, Nikolajeva O, de Wildt A, et al. Impact of thymoglobulin prior to pediatric unrelated umbilical cord blood transplantation on immune reconstitution and clinical outcome. *Blood* 2014;123:126–32.
32. Wagner JE Jr, Eapen M, Carter S, Wang Y, Schultz KR, Wall DA, et al. One-unit versus two-unit cord-blood transplantation for hematologic cancers. *N Engl J Med* 2014;371:1685–94.
33. Baron F, Ruggeri A, Beohou E, Labopin M, Sanz G, Milpied N, et al. RIC versus MAC UCBT in adults with AML: a report from Eurocord, the ALWP and the CTIWP of the EBMT. *Oncotarget* 2016;7:43027–38.
34. Michel G, Galambrun C, Sirvent A, Pochon C, Bruno B, Jubert C, et al. Single versus double-unit cord blood transplantation for children and young adults with acute leukemia or myelodysplastic syndrome. *Blood* 2016;127:3450–7.
35. Frassoni F, Labopin M, Powles R, Mary JY, Arcese W, Bacigalupo A, et al. Effect of centre on outcome of bone-marrow transplantation for acute myeloid leukaemia. Acute leukaemia working party of the European group for blood and marrow transplantation. *Lancet* 2000;355: 1393–8.
36. Gratwohl A, Brand R, Niederwieser D, Baldomero H, Chabannon C, Cornelissen J, et al. Introduction of a quality management system and outcome after hematopoietic stem-cell transplantation. *J Clin Oncol* 2011;29:1980–6.
37. Giebel S, Labopin M, Mohty M, Mufti GJ, Niederwieser D, Cornelissen JJ, et al. The impact of center experience on results of reduced intensity: allogeneic hematopoietic SCT for AML. An analysis from the acute leukemia working party of the EBMT. *Bone Marrow Transplant* 2012;48: 238–42.
38. Shouval R, Bonifazi F, Fein J, Boschini C, Oldani E, Labopin M, et al. Validation of the acute leukemia EBMT score for prediction of mortality following allogeneic stem cell transplantation in a multicenter GITMO cohort. *Am J Hematol* 2017;92:429–34.