

MYELOID NEOPLASIA

Use of machine learning techniques to predict poor survival after hematopoietic cell transplantation for myelofibrosis

Juan Carlos Hernández-Boluda,^{1,*} Adrián Mosquera-Orgueira,^{2,*} Luuk Gras,³ Linda Koster,³ Joe Tuffnell,³ Nicolaus Kröger,⁴ Massimiliano Gambella,⁵ Thomas Schroeder,⁶ Marie Robin,⁷ Katja Sockel,⁸ Jakob Passweg,⁹ Igor Wolfgang Blau,¹⁰ Ibrahim Yakoub-Agha,¹¹ Ruben Van Dijk,¹² Mattias Stelljes,¹³ Henrik Sengeloev,¹⁴ Jan Vydra,¹⁵ Uwe Platzbecker,¹⁶ Moniek de Witte,¹⁷ Frédéric Baron,¹⁸ Kristina Carlson,¹⁹ Javier Rojas,² Carlos Pérez Míguez,² Davide Crucitti,² Kavita Raj,²⁰ Joanna Drozd-Sokolowska,²¹ Giorgia Battipaglia,²² Nicola Polverelli,²³ Tomasz Czerw,²⁴ and Donal P. McLoman,²⁰ on behalf of the Chronic Malignancies Working Party of the EBMT

¹Hematology Department, Hospital Clínico Universitario, Instituto de Investigación Sanitaria del Hospital Clínico de Valencia, University of Valencia, Valencia, Spain; ²Hematology Department, University Hospital of Santiago de Compostela, Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago de Compostela, Spain; ³European Group for Blood and Marrow Transplantation Leiden Study Unit, Leiden, The Netherlands; ⁴Hematology Department, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; ⁵Ematologia e Terapie Cellulari, IRCCS Ospedale Policlinico San Martino, Genova, Italy; ⁶Hematology Department, University Hospital Essen, Duesseldorf, Germany; ⁷Hematology Department, Saint-Louis Hospital, Bone Marrow Transplantation Unit, Paris, France; ⁸Hematology Department, University Hospital Technische Universität Dresden, Dresden, Germany; ⁹Hematology Department, University Hospital Basel, Basel, Switzerland; ¹⁰Medizinische Klinik m. S. Hämatologie, Onkologie und Tumorummunologie, Berlin, Germany; ¹¹Hematology Department, Centre Hospitalier Universitaire de Lille, INSERM U1286, InFINITE, Lille, France; ¹²Hematology Department, Erasmus MC Cancer Institute, Rotterdam, The Netherlands; ¹³Hematology Department, University of Muenster, Muenster, Germany; ¹⁴Hematology Department, Rigshospitalet, Copenhagen, Denmark; ¹⁵Institute of Hematology and Blood Transfusion, Prague, Czech Republic; ¹⁶Hematology Department, Medical Clinic and Policlinic, Leipzig, Germany; ¹⁷Hematology Department, University Medical Centre, Utrecht, The Netherlands; ¹⁸Hematology Department, University of Liege, Liege, Belgium; ¹⁹Hematology Department, University Hospital Uppsala, Uppsala, Sweden; ²⁰Hematology Department, University College London Hospitals National Health Service Trust, London, United Kingdom; ²¹Hematology Department, Central Clinical Hospital, The Medical University of Warsaw, Warsaw, Poland; ²²Hematology Department, Federico II University of Naples, Naples, Italy; ²³Unit of Bone Marrow Transplantation, Division of Hematology, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy; and ²⁴Hematology Department, Maria Skłodowska-Curie National Research Institute of Oncology, Gliwice, Poland

KEY POINTS

- A ML model, available as an interactive web application, was created to predict survival after transplant in MF.
- This tool is a step toward personalized medicine, enabling the identification of 25% of patients with poor transplantation outcomes.

With the incorporation of effective therapies for myelofibrosis (MF), accurately predicting outcomes after allogeneic hematopoietic cell transplantation (allo-HCT) is crucial for determining the optimal timing for this procedure. Using data from 5183 patients with MF who underwent first allo-HCT between 2005 and 2020 at European Society for Blood and Marrow Transplantation centers, we examined different machine learning (ML) models to predict overall survival after transplant. The cohort was divided into a training set (75%) and a test set (25%) for model validation. A random survival forests (RSF) model was developed based on 10 variables: patient age, comorbidity index, performance status, blood blasts, hemoglobin, leukocytes, platelets, donor type, conditioning intensity, and graft-versus-host disease prophylaxis. Its performance was compared with a 4-level Cox regression-based score and other ML-based models derived from the same data set, and with the Center for International Blood and Marrow Transplant Research score. The RSF outperformed all comparators, achieving better concordance indices across both primary

and postessential thrombocythemia/polycythemia vera MF subgroups. The robustness and generalizability of the RSF model was confirmed by Akaike information criterion and time-dependent receiver operating characteristic area under the curve metrics in both sets. Although all models were prognostic for nonrelapse mortality, the RSF provided better curve separation, effectively identifying a high-risk group comprising 25% of patients. In conclusion, ML enhances risk stratification in patients with MF undergoing allo-HCT, paving the way for personalized medicine. A web application (<https://gemfin.click/ebmt>) based on the RSF model offers a practical tool to identify patients at high risk for poor transplantation outcomes, supporting informed treatment decisions and advancing individualized care.

Introduction

Myelofibrosis (MF) is a chronic myeloproliferative neoplasm that appears *de novo* (primary MF [PMF]) or after a diagnosis of essential thrombocythemia or polycythemia vera (secondary MF [SMF]). Managing MF is complex because of its diverse clinical manifestations including an inherent risk of progression to acute myeloid leukemia (AML). Although the median overall survival (OS) is ~6 years, it varies significantly among patients. Medical treatment focuses on symptom control and quality of life but is not curative and does not reduce the risk of AML.^{1,2}

Allogeneic hematopoietic cell transplantation (allo-HCT) remains the only curative option for MF.³ However, its significant morbidity and mortality require a careful risk-benefit analysis to identify appropriate candidates. This has become particularly critical recently, as several effective therapies for MF have been incorporated into clinical practice,⁴⁻⁶ with others showing promising results in clinical trials.⁴ Existing prognostic models for OS after allo-HCT have been instrumental in guiding clinical decisions.^{7,8} However, they do not account for key factors such as patient comorbidities⁷ and emerging transplant strategies, such as haploidentical transplants or posttransplant cyclophosphamide use.^{9,10} Furthermore, there is significant room for improving their ability to accurately identify patients at high risk of posttransplant mortality, who may benefit more from alternative treatments or clinical trials.

Machine learning (ML) is a field of artificial intelligence in which prediction is based on modeling of outcomes considering the complex interactions among multiple variables, rather than relying on predefined human-made rules. These techniques have demonstrated their utility to provide accurate personalized survival predictions for patients with MF undergoing conventional drug treatment.^{11,12} In this study, we aim to assess whether ML techniques can similarly improve prognostication of OS in the setting of allo-HCT for MF. The ultimate goal is to enhance transplant decision-making by providing more precise and individualized survival predictions.

Methods

Data source

We retrieved data from adult patients with MF (PMF or SMF) who underwent first allo-HCT between 2005 and 2020 in European Society for Blood and Marrow Transplantation (EBMT) centers. Patients who received transplantation from umbilical cord blood and those with AML transformation were excluded. Myeloablative or reduced intensity conditioning regimens were defined by standard EBMT criteria.¹³ Matched unrelated transplants were matched at allele level for HLA-A, -B, -C, -DRB1, and -DQB1. Centers ensured informed consent in compliance with local regulations to report pseudonymized data to the EBMT. The study was approved by the Chronic Malignancies Working Party of EBMT and conducted in accordance with the Declaration of Helsinki. Informed consent for inclusion in the EBMT registry was obtained for all patients. The study database included 52 variables selected for their prognostic significance based on previous studies.^{7,14,15}

Main study outcomes

The primary goal was to develop a prognostic model for OS using ML techniques and compare its performance with that of

a Cox regression-based score developed in the same data set, and with the Center for International Blood and Marrow Transplant Research (CIBMTR) model.⁸ The models were also applied for predicting the secondary outcome of nonrelapse mortality (NRM). Progression-free survival (PFS) and cumulative incidence of relapse were estimated for descriptive purposes. Median follow-up was determined using the reverse Kaplan-Meier method. In patients who died after disease relapse, relapse was considered the primary cause of death.¹⁶

Statistical analysis

OS and PFS were estimated by the Kaplan-Meier method. NRM was defined as the time from the date of transplantation to the date of death (uncensored) or to the date of disease relapse (censored). The cumulative incidences of relapse/NRM (as competing risk for each other) were analyzed separately in a competing risks framework.¹⁷

Two independent statisticians applied distinct methodologies to evaluate the factors influencing OS. One used a conventional multivariate Cox proportional hazards regression model, whereas the other used a range of ML techniques. Both approaches were based on the same random distribution of the patient cohort into a training set (75% of the cohort, $n = 3887$) and a test set (25% of the cohort, $n = 1296$). Each statistician independently determined the optimal cutoff points for the newly derived risk scores to stratify patients into distinct prognostic groups. The resulting risk classifications were subsequently compared and contextualized to assess their clinical relevance for treatment decision-making.

Statistical analyses were performed in R version 4.3,¹⁸ with the "survival,"¹⁹ "prodlm,"²⁰ and "cmprsk"²¹ packages.

Multivariate Cox regression model

Factors potentially associated with OS were entered into a Cox proportional hazards model. Selection of variables in the final model was based on expert clinical advice and data availability, to assess the independent effect of each covariate. Variables with a high degree of missingness (defined as >50%) were not considered for inclusion in the model, whereas a missing category was created for those variables with a low degree of missingness. Hazard ratios (HRs) were provided, and corresponding *P* values were calculated using the Wald test. A score of 1 or 2 was assigned to each significant variable for OS based on the HRs obtained from multivariable analysis. The cutoff values were arbitrarily defined as follows: HR of <1.25 = 0 points, HR of 1.25 to 1.50 = 1 point, HR of >1.50 = 2 points. A prognostic scoring system was subsequently developed considering the sum of risk points to discriminate 4 patient risk groups with significant differences in OS. All *P* values were 2-sided and $P < .05$ was considered significant.

RSF model

Random survival forests (RSF) were created with 1000 trees. For cross-validation, sampling was performed without replacement, which, by default, takes 0.632 times the sample size. Missing variables were imputed in the training and test cohorts separately using a missing data algorithm developed by Ishwaran et al.²² Predictions were cross-validated in the training set

(using the out-of-bag method) and then validated in the test cohort. This was done to rule out overfitting of performance metrics in the training set related to either variable selection or the imputation process. Dimensionality reduction was performed using variable-importance estimation analysis. Redundant and dependent variables were discarded applying variable-importance estimations and clinical knowledge, achieving a minimally dimensioned yet effective model.

Hyperparameter tuning was explored to optimize the performance of the RSF model. Specifically, a grid search was used to explore the impact of key hyperparameters, including “mtry” (the number of variables randomly selected at each split) and “node size” (the minimum number of samples required in terminal nodes). Partial dependence plots were used to evaluate the marginal effect of individual covariates on survival estimates, while accounting for the average influence of all other variables. These plots help visualize nonlinear relationships and potential threshold effects, providing insight into the contribution of specific predictors to survival probability over time. This approach enhances the model’s interpretability by isolating the independent effect of each covariate on the predicted survival outcomes.

Comparison of different ML techniques for survival analysis

In addition to the baseline RSF model, we evaluated the performance of 3 complementary methods using the 10 variables included in the final model: oblique RSF (ORSF), gradient-boosted survival trees using XGBoost, and a deep neural network–based survival model (DeepSurv).²³⁻²⁵

A detailed explanation of the different ML techniques is outlined in the supplemental Data, available on the *Blood* website. All 3 approaches, ORSF, XGBoost-based survival modeling, and DeepSurv, were evaluated using cross-validation for performance estimation. Early stopping was applied when computationally practical to limit overfitting. The performance of each model was assessed using standard survival metrics (eg, C-index) to facilitate a rigorous comparative analysis.

Comparison of the discriminative capacity between the ML and the Cox-derived models for survival prediction

The discriminative capacity of the ML and Cox OS models was compared in the training and test sets using the Harrell C-index.²⁶ Time-dependent receiver operating characteristic (ROC) area under the curves (AUCs) for OS were calculated using the timeROC package.²⁷ Quantitative scores (continuous risk estimates) and categorical risk groups were analyzed separately for each model. For risk groups, categorical labels were converted into numeric values for compatibility. Time-dependent ROC-AUCs assessed model performance at specific time points, allowing a robust evaluation of prognostic accuracy over time of the RSF and Cox models in terms of their quantitative and group-based predictions.

Because the C-index is not an optimal metric for competing risk models, we also included the Akaike information criterion (AIC) scores for both NRM and OS.²⁸

Results

Patient and transplant characteristics

A total of 5183 patients from 288 centers fulfilling the selection criteria were included. Baseline characteristics of all patients, along with the training ($n = 3887$) and validation ($n = 1296$) cohorts, are presented in Table 1. Median follow-up was 58.2 months (95% confidence interval [CI], 55.6-59.8) in the training set and 60.0 months (95% CI, 55.7-63.2) in the test set. Median OS was 79.4 months (95% CI, 69.2-89.6) in the training set and 73.7 months (95% CI, 54.7-92.7) in the test set. No significant differences in characteristics were observed between the cohorts, apart from a higher platelet count at allo-HCT and less antithymocyte globulin use in the test cohort.

Transplantation outcomes

The estimated OS rate at 1, 5, and 10 years was 70% (95% CI, 69-71), 53% (95% CI, 51-54), and 43% (95% CI, 41-45), respectively.

The probability of PFS after 1, 5, and 10 years was 62% (95% CI, 60-63), 44% (95% CI, 43-46), and 35% (95% CI, 33-37), respectively. The estimated NRM rate at 1, 5, and 10 years was 23% (95% CI, 22-24), 32% (95% CI, 31-33), and 36% (95% CI, 35-38), respectively. Cumulative incidence of relapse at 1, 5, and 10 years was 15% (95% CI, 14-16), 24% (95% CI, 23-25), and 29% (95% CI, 27-31), respectively.

The graphical representation of the main study outcomes in the training and test cohorts can be seen in supplemental Figure 1.

Risk model for OS using Cox regression analysis

Factors associated with OS in the univariable analysis are shown in supplemental Table 1. In the multivariable analysis, 7 independent factors significantly predicted reduced OS: older patient age, HLA-mismatched donor type, lower Karnofsky performance status (KPS), higher HCT-specific comorbidity index (HCT-CI), JAK2/triple-negative genotype, graft from a female donor to a male patient, and graft from a donor who is cytomegalovirus (CMV) positive to a recipient who is CMV negative (Table 2). Based on the HRs, a score of 2 was assigned to patient age of ≥ 60 years and the JAK2/triple-negative genotype; and a score of 1 to patient age of 50 to 59 years, haploidentical or mismatched unrelated donors, KPS of < 90 , and HCT-CI of ≥ 3 . Because the HRs for sex match and CMV serostatus were only modestly increased, no score was assigned to these factors. The total score ranged from 0 to 7 points, with 4 risk categories: low risk (0-1 points), intermediate-1 risk (2-3 points), intermediate-2 risk (4-5 points), and high risk (6-7 points). The corresponding 5-year OS of each category in the training and test set were 82% (95% CI, 74-90) and 65% (95% CI, 41-89) for low risk (6% of the cohort); 62% (95% CI, 58-66) and 65% (IC 95%, 57-73) for intermediate-1 (36% of the cohort); 52% (95% CI, 48-56) and 47% (95% CI, 40-54) for intermediate-2 (48% of the cohort); and 39% (95% CI, 30-47) and 30% (95% CI, 13-47) for high risk (10% of the cohort), respectively (Figure 1A-B).

Risk model for OS using RSF

A RSF model was created to predict OS using the 52 initial variables of the data set. This model achieved a C-index of 0.603 in the training set and 0.632 in the test set. The

Table 1. Main characteristics of a series of 5183 patients with MF undergoing allo-HCT and of the training and test cohorts

Characteristic	Group	Missing (%)	Total cohort	Training set	Test set	P value
			N (%)	n (%)	n (%)	
No. of patients			5183 (100)	3887 (100)	1296 (100)	
Age at allo-HCT, y*			58.3 (52-63.5)	58.2 (51.8-63.4)	58.6 (52.7-63.8)	.072
Age at allo-HCT	<60 y		3003 (57.9)	2274 (58.5)	729 (56.2)	.164
Patient sex	Male		3242 (62.6)	2412 (62.1)	830 (64.0)	.21
Year of MF diagnosis	<2000		615 (11.9)	455 (11.7)	160 (12.3)	.21
	2000-2010		1765 (34.1)	1310 (33.7)	455 (35.1)	
	2010-2015		1422 (27.4)	1057 (27.2)	365 (28.2)	
	≥2015		1381 (26.6)	1065 (27.4)	316 (24.4)	
MF type	PMF		3743 (72.2)	2807 (72.2)	936 (72.2)	1.00
	SMF		1440 (27.8)	1080 (27.8)	360 (27.8)	
JAK2 inhibitor treatment before allo-HCT	Yes	1503 (29)	1039 (28.2)	764 (27.9)	275 (29.2)	.459
Genotype	JAK2 ⁺	2422 (46.7)	2156 (78.1)	1625 (77.7)	531 (79.3)	.43
	MPL ⁺		94 (3.4)	68 (3.3)	26 (3.9)	
	CALR ⁺		377 (13.7)	290 (13.9)	87 (13.0)	
	Triple negative		134 (4.9)	108 (5.2)	26 (3.9)	
Constitutional symptoms at allo-HCT	Yes	3089 (59.6)	911 (43.5)	693 (44.3)	218 (41.3)	.255
Hemoglobin at allo-HCT, g/dL*		2575 (49.7)	9.2 (8.1-10.5)	9.2 (8.1-10.5)	9.2 (8.2-10.6)	.43
Leukocyte count at allo-HCT, ×10 ⁹ /L*		2600 (50.2)	6.9 (3.5-14.4)	6.8 (3.5-14.3)	7 (3.6-14.4)	.336
Blood blasts at allo-HCT, %*		3095 (59.7)	1 (0-3)	1 (0-3)	1 (0-3)	.637
Platelets at allo-HCT, ×10 ⁹ /L*		2640 (50.9)	117 (53-242.5)	114.5 (53-237)	125 (51-274)	.039
Splenectomy before allo-HCT	Yes	2536 (48.9)	324 (12.2)	231 (11.7)	93 (14.0)	.129
Spleen size below costal margin, cm, by physical examination*		3857 (74.4)	5 (0-10)	5 (1-11)	5 (0-10)	.184
Spleen span by ultrasound or CT scan, max diameter, cm*		4283 (82.6)	20 (16-23)	20 (16-23)	20 (16.2-23)	.72
HCT-CI risk group	Low risk (0 points)	1360 (26.2)	2041 (53.4)	1538 (53.7)	503 (52.4)	.745
	Intermediate risk (1-2 points)		910 (23.8)	674 (23.5)	236 (24.6)	
	High risk (≥3 points)		872 (22.8)	651 (22.7)	221 (23.0)	
KPS score at allo-HCT	90-100	489 (9.4)	3111 (66.3)	2316 (65.8)	795 (67.7)	.508
	80		1232 (26.2)	937 (26.6)	295 (25.1)	
	<80		351 (7.5)	266 (7.6)	85 (7.2)	
DIPSS risk group at allo-HCT	Low risk	2715 (52.4)	60 (2.4)	40 (2.2)	20 (3.2)	.286
	Intermediate-1		919 (37.2)	695 (37.7)	224 (36.0)	
	Intermediate-2		954 (38.7)	702 (38.0)	252 (40.4)	
	High risk		535 (21.7)	408 (22.1)	127 (20.4)	

ATG, antithymocyte globulin; Bu, busulfan; CNJ, calcineurin inhibitor; CT, computed tomography; Cy, cyclophosphamide; DIPSS, Dynamic International Prognostic Scoring System; Flu, fludarabine; MAC, myeloablative conditioning; Mel, Melphalan; MMF, mycophenolate mofetil; MMRD, mismatched related donor; max, maximum; MMUD, mismatched unrelated donor; MRD, matched related donor; MTX, methotrexate; MUD, matched unrelated donor; RIC, reduced intensity conditioning; TBI, total body irradiation; Treo, treosulfan; w/o, with/without.

*Values are median (interquartile range).

Table 1 (continued)

Characteristic	Group	Missing (%)	Total cohort	Training set	Test set	P value
			N (%)	n (%)	n (%)	
CIBMTR risk score at allo-HCT	Low Intermediate High	2640 (50.9)	1020 (40.1) 1313 (51.6) 210 (8.3)	763 (39.6) 1004 (52.2) 158 (8.2)	257 (41.6) 309 (50.0) 52 (8.4)	.641
Donor type	Identical sibling MRD (other than sibling) MMRD MUD MMUD Unrelated, number of mismatches unknown		1534 (29.6) 45 (0.9) 339 (6.5) 2175 (42.0) 673 (13.0) 417 (8.0)	1147 (29.5) 34 (0.9) 249 (6.4) 1646 (42.3) 504 (13.0) 307 (7.9)	387 (29.9) 11 (0.8) 90 (6.9) 529 (40.8) 169 (13.0) 110 (8.5)	.919
Recipient-donor match	Recipient male with donor male Recipient male with donor female Recipient female with donor male Recipient female with donor female	76 (1.5)	2297 (45.0) 895 (17.5) 1147 (22.5) 768 (15.0)	1714 (44.7) 666 (17.4) 878 (22.9) 578 (15.1)	583 (45.9) 229 (18.0) 269 (21.2) 190 (14.9)	.614
Donor age, y*		1132 (21.8)	36.6 (27.2-49.8)	36.7 (27.2-49.6)	36.1 (27.1-50.1)	.902
CMV serology in patient/donor	-/- -/+ +/- +/+	233 (4.5)	1437 (29.0) 464 (9.4) 1002 (20.2) 2047 (41.4)	1056 (28.4) 348 (9.4) 776 (20.9) 1533 (41.3)	381 (30.8) 116 (9.4) 226 (18.3) 514 (41.6)	.174
Stem cell source	Bone marrow Peripheral blood		409 (7.9) 4774 (92.1)	305 (7.8) 3582 (92.2)	104 (8.0) 1192 (92.0)	.884
Busulfan- or melphalan-based conditioning regimen	Busulfan based Melphalan based Other regimen	89 (1.7)	3522 (69.1) 718 (14.1) 854 (16.8)	2642 (69.2) 533 (14.0) 644 (16.9)	880 (69.0) 185 (14.5) 210 (16.5)	.861
Conditioning drugs	BuCy w/wo others (MAC) FluBu w/wo others (MAC) FluTreo w/wo others (MAC) TBI w/wo Cy w/wo others (MAC) Others (MAC) FluBu w/wo others (RIC) FluMel w/wo others (RIC) FluTreo w/wo others (RIC) TBI w/wo Cy w/wo Flu w/wo others (RIC) Others (RIC)	146 (2.8)	166 (3.3) 1153 (22.9) 160 (3.2) 135 (2.7) 155 (3.1) 2135 (42.4) 579 (11.5) 143 (2.8) 275 (5.5) 136 (2.7)	137 (3.6) 858 (22.7) 121 (3.2) 111 (2.9) 116 (3.1) 1588 (42.1) 428 (11.3) 105 (2.8) 206 (5.5) 105 (2.8)	29 (2.3) 295 (23.4) 39 (3.1) 24 (1.9) 39 (3.1) 547 (43.3) 151 (12.0) 38 (3.0) 69 (5.5) 31 (2.5)	.324

ATG, antithymocyte globulin; Bu, busulfan; CNI, calcineurin inhibitor; CT, computed tomography; Cy, cyclophosphamide; DIPSS, Dynamic International Prognostic Scoring System; Flu, fludarabine; MAC, myeloablative conditioning; Mel, Melphalan; MMF, mycophenolate mofetil; MMRD, mismatched related donor; max, maximum; MMUD, mismatched unrelated donor; MRD, matched related donor; MTX, methotrexate; MUD, matched unrelated donor; RIC, reduced intensity conditioning; TBI, total body irradiation; Treo, treosulfan; w/wo, with/without.

*Values are median (interquartile range).

Table 1 (continued)

Characteristic	Group	Missing (%)	Total cohort	Training set	Test set	P value
			N (%)	n (%)	n (%)	
Conditioning regimen intensity	MAC	75 (1.4)	1802 (35.3)	1366 (35.7)	436 (34.1)	.32
	RIC		3306 (64.7)	2463 (64.3)	843 (65.9)	
TBI	Yes	37 (0.7)	420 (8.2)	326 (8.4)	94 (7.3)	.219
T-cell depletion	No	148 (2.9)	1356 (26.9)	1010 (26.8)	346 (27.4)	.451
	Yes in vivo, no ex vivo		3585 (71.2)	2696 (71.4)	889 (70.5)	
	Yes ex vivo, no in vivo		17 (0.3)	10 (0.3)	7 (0.6)	
	Yes in vivo + ex vivo		77 (1.5)	58 (1.5)	19 (1.5)	
ATG	Yes	120 (2.3)	3387 (66.9)	2568 (67.7)	819 (64.5)	.038
Alemtuzumab	Yes	167 (3.2)	306 (6.1)	218 (5.8)	88 (7.0)	.146
GVHD prophylaxis group	Post-Cy	112 (2.2)	415 (8.2)	311 (8.2)	104 (8.2)	.064
	ATG + CNI + MMF		1398 (27.6)	1064 (28.0)	334 (26.3)	
	ATG + CNI + MTX		1352 (26.7)	1021 (26.9)	331 (26.0)	
	ATG + CNI		354 (7.0)	276 (7.3)	78 (6.1)	
	ATG w/wo other(s)		195 (3.8)	150 (3.9)	45 (3.5)	
	Post-Cy + ATG		88 (1.7)	57 (1.5)	31 (2.4)	
	CNI only		184 (3.6)	129 (3.4)	55 (4.3)	
	CNI + MMF w/wo other		434 (8.6)	311 (8.2)	123 (9.7)	
	CNI + MTX w/wo other		522 (10.3)	393 (10.3)	129 (10.1)	
	Other		129 (2.5)	88 (2.3)	41 (3.2)	

ATG, antithymocyte globulin; Bu, busulfan; CNI, calcineurin inhibitor; CT, computed tomography; Cy, cyclophosphamide; DIPSS, Dynamic International Prognostic Scoring System; Flu, fludarabine; MAC, myeloablative conditioning; Mel, Melphalan; MMF, mycophenolate mofetil; MMRD, mismatched related donor; max, maximum; MMUD, mismatched unrelated donor; MRD, matched related donor; MTX, methotrexate; MUD, matched unrelated donor; RIC, reduced intensity conditioning; TBI, total body irradiation; Treo, treosulfan; w/wo, with/without.

*Values are median (interquartile range).

variable-importance metrics for the model in the training set are shown in supplemental Figure 2. After dimensionality reduction, the model was refined to a smaller set of key prognostic variables: patient age, HCT-CI, KPS, blood blasts percentage, hemoglobin level, leukocyte and platelet counts, donor type, conditioning intensity, and graft-versus-host disease (GVHD) prophylaxis. This model achieved a C-index of 0.599 in the training set and 0.623 in the test set. Despite having performed hyperparameter tuning, no improvement in C-index was achieved (supplemental Table 2). To further elucidate these relationships, we have also included partial dependence plots in supplemental Figure 3.

Comparison of the RSF with other ML techniques

As shown in supplemental Table 3, RSF achieved higher concordance indices for OS and NRM predictions in both training and test sets compared with 3 alternative methods (ORSF, DeepSurv, and XGBoost). The consistent and superior performance of RSF across both data partitions justified its selection as the primary approach for downstream analyses.

Comparison of the ML model with the Cox regression-derived model

This analysis was performed on the subset of patients who had complete information on the variables included in the Cox-derived score to minimize biases (training set: $n = 1773$; test

set: $n = 566$). In mortality prediction, the ML model demonstrated modestly better performance in the training set, achieving a C-index of 0.603 compared with the Cox-derived score of 0.594. The test set results confirmed the better discriminative capacity of the ML model, with a score of 0.612, surpassing the Cox-derived score C-index of 0.587 (Table 3). These findings were corroborated by the AIC scores, with the ML model showing lower values than the Cox-derived score in the test set, indicating a better overall model fit (Table 3).

In refining our analysis, we segmented the ML score into 4 equal groups within the training set and applied the same classification thresholds to the test set (Figure 1C-D). The ML model maintained similar C-indices after segmentation, indicating that the model's prognostic accuracy is resilient to simplification (Table 3). A substantial reassignment of patients from the intermediate-2 risk group of the Cox score to other risk groups by the ML model was noted (Figure 2A-B). The time-dependent ROC AUCs comparing both models are presented in supplemental Figure 4.

Comparison of the ML model with the CIBMTR model

We compared the performance of the ML model with the CIBMTR scoring system⁸ in a comparable subset of patients with complete annotations for the CIBMTR score. By integrating

Table 2. Cox regression analysis of factors associated with OS in the training cohort

	HR (95% CI)	(Overall) P
Conditioning intensity		
MAC	1.00	
RIC	1.08 (0.97-1.20)	.17
Donor type		
MRD/MUD	1.00	
HD/MMUD	1.34 (1.19-1.51)	<.0001
Age at allo-HCT, y		
<49	1.00	
50-59	1.36 (1.17-1.58)	<.0001
≥60	1.67 (1.44-1.94)	<.0001
Sex		
Male	1.00	
Female	0.89 (0.80-0.99)	.04
KPS at allo-HCT		
90-100	1.00	
<90	1.33 (1.19-1.48)	<.0001
HCT-CI		
0-2	1.00	
≥3	1.36 (1.19-1.55)	<.0001
CMV donor/patient		
+/-	1.16 (1.03-1.32)	.01
Other	1.00	
Sex match donor/patient		
Female to male	1.15 (1.00-1.31)	.04
Other	1.00	
Genotype		
CALR ⁺ /MPL ⁺	1.00	
JAK2 ⁺ /triple negative	1.56 (1.26-1.92)	<.0001

Overall P values were obtained using the Wald test. A total of 3559 patients and 1607 events were included in the model.

HD, haploidentical donor; MAC, myeloablative conditioning; MMUD, mismatched unrelated donor; MRD, matched related donor; MUD, matched unrelated donor; RIC, reduced intensity conditioning.

patient age, hemoglobin level at transplant, and donor type, this model defined 3 risk categories in the original series, with a 3-year posttransplant OS of 69%, 51%, and 34% for low, intermediate, and high-risk groups, respectively.

The ML model achieved better performance, with a C-index of 0.608 vs 0.557 in the training set (n = 1925) and 0.654 vs 0.581 in the test set (n = 618; Table 3; supplemental Figure 5). Additionally, the lower AIC scores observed with the ML approach further validated these findings (Table 3). The time-dependent ROC AUCs comparing both models are presented in supplemental Figure 6. The difference between the CIBMTR score and ML method was mostly driven by prognostic refinement within the CIBMTR intermediate risk group, for which the

ML algorithm reclassified most patients into different risk categories (Figure 2C-D).

Notably, the ML model exhibited a consistently better discriminative performance than the Cox-derived models for patients with PMF and patients with SMF, with the advantage being more pronounced in the test set (supplemental Table 4).

Application of the ML model to predict NRM

In predicting NRM, the ML model achieved comparable AIC scores to those of the Cox-based model in the training set but substantially lower AIC scores in the test set, indicating better overall performance (Table 3). Furthermore, when compared with the CIBMTR score, the ML model demonstrated an even more pronounced improvement in overall model fit (Table 3; Figure 3).

Ability of the ML model to identify patients at high risk of posttransplant mortality

The clinical utility of the ML model was evident in its ability to stratify patients into risk groups. Notably, it assigned 25% of patients to the high-risk group, significantly more than the 10.1% in the Cox-derived score and 8.2% in the CIBMTR model (Figure 2). Moreover, the ML model not only identified a larger proportion of high-risk patients but also showed consistent and generalizable results across training and test sets (Figure 1). In the training set, the 12- and 24-month OS rates for the ML high-risk group were 58.9% and 51.5%, respectively, closely aligning with the Cox-derived scores of 58.3% and 52.7%, respectively. In the test set, the ML high-risk group had OS rates of 61.0% at 12 months and 48.1% at 24 months, closely matching the 61.8% and 50.1% of the Cox model high-risk group.

The ML model also identified a larger high-risk population for NRM compared with the Cox-derived score (Figure 3). In the training set, the ML high-risk group had 12- and 24-month NRM rates of 34.9% and 40.7%, respectively, lower than the 46.3% and 48.8% observed in the Cox model high-risk group. However, in the test set, the ML high-risk group showed 12- and 24-month NRM rates of 36.4% and 42.6%, respectively, nearly matching the Cox score rates of 36.0% and 42.8%.

The comparison of patient distribution between the Cox-derived method and the CIBMTR model is elicited in supplemental Figure 7.

To predict OS after allo-HCT in MF, we developed a prognostic tool based on the RSF model, accessible as an interactive web application (<https://gemfin.click/ebmt>). Figure 4 illustrates the web-based calculator, showing the risk score for a hypothetical transplant candidate.

Impact of modifiable key factors of the transplantation procedure on OS

We compared OS after allo-HCT in patients who received the "optimal" donor type, conditioning intensity, or GVHD prophylaxis with those who did not. Optimal strategies were defined as those that maximize the survival probability according to the model's predictions for a given patient, conditional on the other individual and disease characteristics. To

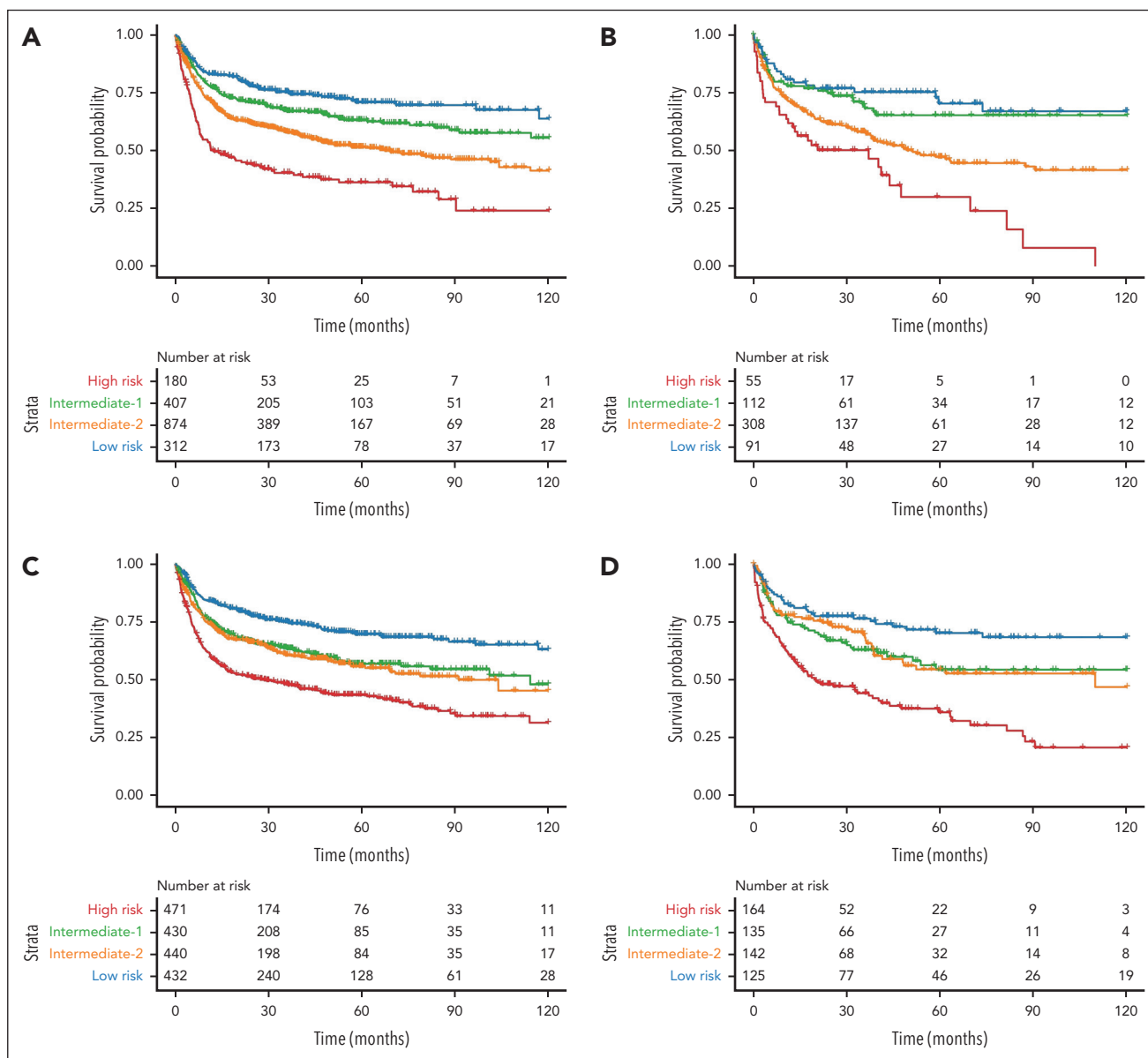


Figure 1. Kaplan-Meier curves illustrating OS after transplant based on risk groups defined by the prognostic models. (A-B) Kaplan-Meier plots showing OS according to the Cox regression statistical model in the training (A) and test (B) sets. (C-D) Kaplan-Meier plots displaying OS according to the ML model in the training (C) and test (D) sets. Patients were split according to the predicted quartile of risk. Each branch represents a quartile of patients with either low (blue), intermediate-low (green), intermediate-high (orange), and high risk (red).

ensure reliability, these predictions were evaluated exclusively on the test set.

For the optimal donor type, the univariable Cox proportional hazards model indicated that receiving transplantation from a donor type predicted as optimal by the ML model was associated with a HR of 0.76 (95% CI, 0.64-0.89; $P = .001$) compared with nonoptimal donor type, suggesting a statistically significant survival benefit. However, after adjusting for potential confounding factors in the multivariable analysis using inverse probability weighting (IPW),²⁹ the HR was 0.96 (95% CI, 0.56-1.65; $P = .89$), indicating no significant survival advantage for ML-predicted optimal donor type.

The univariable Cox model revealed that patients receiving the conditioning regimen intensity predicted as optimal by the ML

model had a HR of 0.89 (95% CI, 0.76-1.06; $P = .19$) compared with those receiving nonoptimal intensities. However, the IPW-adjusted analysis showed a HR of 1.025 (95% CI, 0.07-14.80; $P = .99$), indicating the lack of predictive value of the ML model to select the optimal conditioning regimen intensity.

Regarding GVHD prophylaxis, neither the unadjusted Cox model nor the IPW-adjusted analysis showed any significant difference in survival between patients receiving the ML-predicted optimal GVHD prophylaxis and those who did not. The unadjusted Cox model yielded a HR of 0.95 (95% CI, 0.79-1.15; $P = .62$), whereas the IPW-adjusted analysis resulted in a HR of 0.95 (95% CI, 0.43-2.09; $P = .90$). These results indicate no discernible impact of the ML-predicted optimal GVHD prophylaxis on patient survival.

Table 3. Comparison of the performance of the ML model, the Cox-derived score, and the CIBMTR model

	ML model	ML model	Cox-derived score	Cox-derived score
		4 groups		4 groups
OS risk score for mortality prediction (Harrell C-index)				
Training set (n = 1773)*	0.603	0.596	0.594	0.589
Test set (n = 566)*	0.612	0.608	0.587	0.580
	ML model	ML model		CIBMTR model
		4 groups		3 groups
Training set (n = 1925)†	0.608	0.599		0.557
Test set (n = 618)†	0.654	0.650		0.581
OS risk score for mortality prediction (AIC)				
	ML model	ML model	Cox-derived score	Cox-derived score
		4 groups		4 groups
Training set (n = 1773)*	9 944	9 954	9945	9 954
Test set (n = 566)*	2 757	2 759	2765	2 767
	ML model	ML model		CIBMTR model
		4 groups		3 groups
Training set (n = 1925)†	12 573	12 588		12 647
Test set (n = 618)†	3 318	3 322		3 374
OS risk score for NRM prediction (AIC)				
	ML model	ML model	Cox-derived score	Cox-derived score
		4 groups		4 groups
Training set (n = 1763)*	7 210	7 210	7208	7 214
Test set (n = 566)*	2 006	2 002	2008	2 006
	ML model	ML model		CIBMTR model
		4 groups		3 groups
Training set (n = 1925)†	8 738	8 738		8 762
Test set (n = 618)†	2 154	2 160		2 182

Analyses performed on the subset of patients with complete data on the variables included in the Cox-derived score and the CIBMTR model to minimize biases.

Higher C-index values reflect better model performance in ranking predictions whereas lower AIC values indicate a better fit to the data.

*Cox-derived score.

†CIBMTR model.

Discussion

In this study, we have developed a ML model to enhance risk stratification for patients with MF undergoing allo-HCT, using a large database of 5183 patients with MF from the EBMT registry. Notably, this model is particularly comprehensive, because it considers the broad spectrum of current transplant practices, including diverse conditioning regimens, GVHD prophylaxis approaches, and donor types, such as haploidentical transplants. After dimensionality reduction, the model was simplified to a set of 10 key variables maintaining a

notable discriminative capacity for both OS and NRM in both training and test sets.

Comparative analyses demonstrated a better performance of the ML model over a risk score developed within the same cohort using Cox regression methods. It also showed better discriminative capacity than the CIBMTR score,⁸ along with improved generalizability and an enhanced ability to identify a larger group of patients at high risk for posttransplant mortality. Notably, the improved performance of the ML model was evident in both patients with PMF and those with SMF. The ML model's

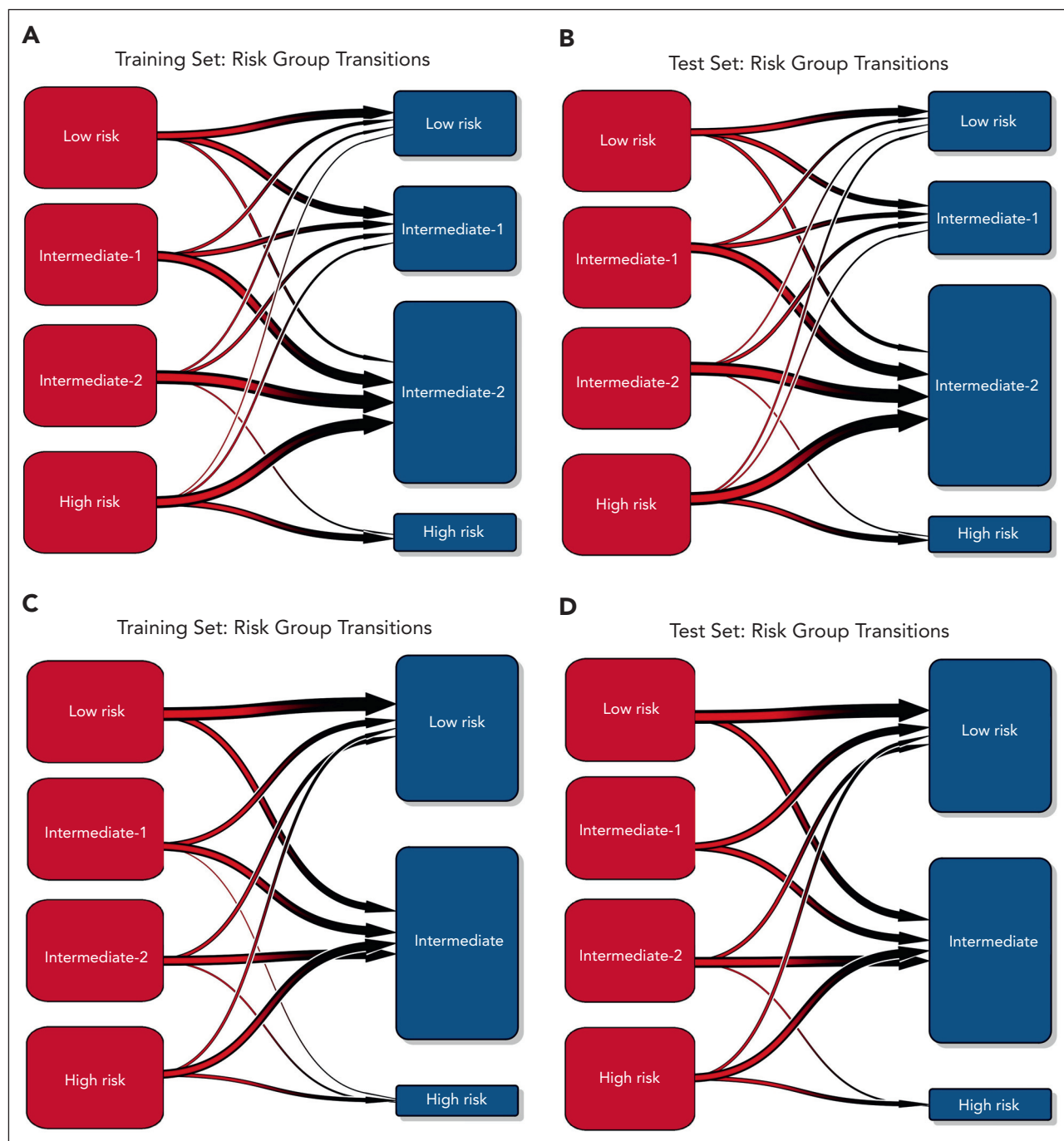


Figure 2. Transition plots illustrating the flow of patients between the ML model and Cox-derived scoring systems. (A-B) Flow of patients between the ML model (red) and the Cox-derived score (blue) in the training (A) and test (B) cohorts. (C-D) Flow of patients between the ML model (red) and the CIBMTR model (blue) in the training (C) and test (D) cohorts.

discriminative capacity remained higher after dividing patients into equally sized risk groups based on individual risk predictions, making the method comparable with traditional risk grouping strategies used in clinical practice. However, the intermediate risk categories identified by the ML model had similar OS and should be consolidated into a single, broader intermediate category, because the model lacks sufficient discrimination within this range. Although the C-indices of the ML model may be deemed moderate in terms of discriminative capacity, our

data support its integration into clinical prognostics, offering a more refined and nuanced approach to managing the complexities of patient risk assessment before allo-HCT.

The clinical relevance of our ML model is evident in its ability to stratify allo-HCT candidates into well-defined risk categories. Notably, it identifies 25% of the cohort as high-risk with poor outcome after allo-HCT (~35% NRM rate and 40% overall mortality at 1 year). Moreover, the web-based calculator permits the

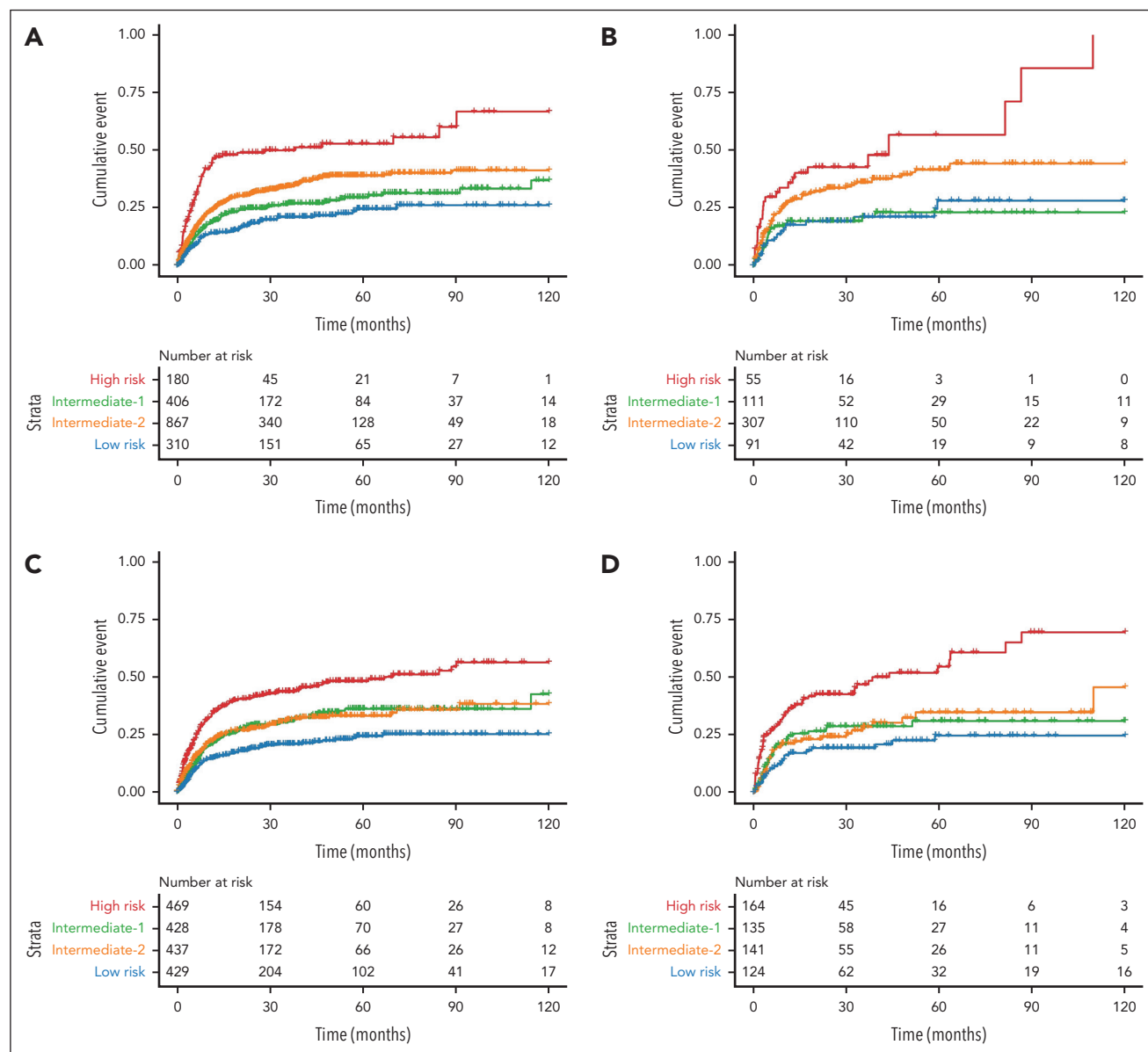


Figure 3. Cumulative incidence of NRM after transplant based on risk groups defined by the prognostic models. (A-B) Cumulative incidence of NRM according to the Cox-derived score in the training (A) and test (B) sets, and according to the ML model in the training (C) and test (D) sets. Patients were divided into 4 quartile groups according to their risk.

identification of a subset of very high-risk patients with a predicted 1-year OS of <50%, allowing for more tailored therapeutic interventions where needed. Accurate risk stratification is essential for optimizing allo-HCT outcomes, enabling physicians to select candidates who are most likely to benefit from transplantation, thereby improving treatment efficacy and patient survival.³ The ML model's robust capability to classify these patients enhances its utility in clinical decision-making, ultimately fostering more personalized and effective patient management strategies in the complex landscape of allo-HCT.³⁰

Although we have proved the feasibility of modeling risk using certain prognostic variables in the context of allo-HCT, the observational nature of the data caution against using these

variables to guide the optimal transplantation procedure. The complexity of allo-HCT, characterized by numerous interacting factors that were not all included in our models, limits the practical utility of predictive models for determining the most effective treatment approaches.³¹ This underscores the need for continued research, randomized clinical studies, as well as sophisticated modeling techniques that can account for the dynamic and multifactorial nature of such medical interventions before they can be reliably implemented in clinical decision-making processes.

Our study has several limitations that warrant consideration. Firstly, allo-HCT is a multifaceted procedure influenced by numerous interrelated and independent variables, including differences in patient and disease characteristics and center-specific

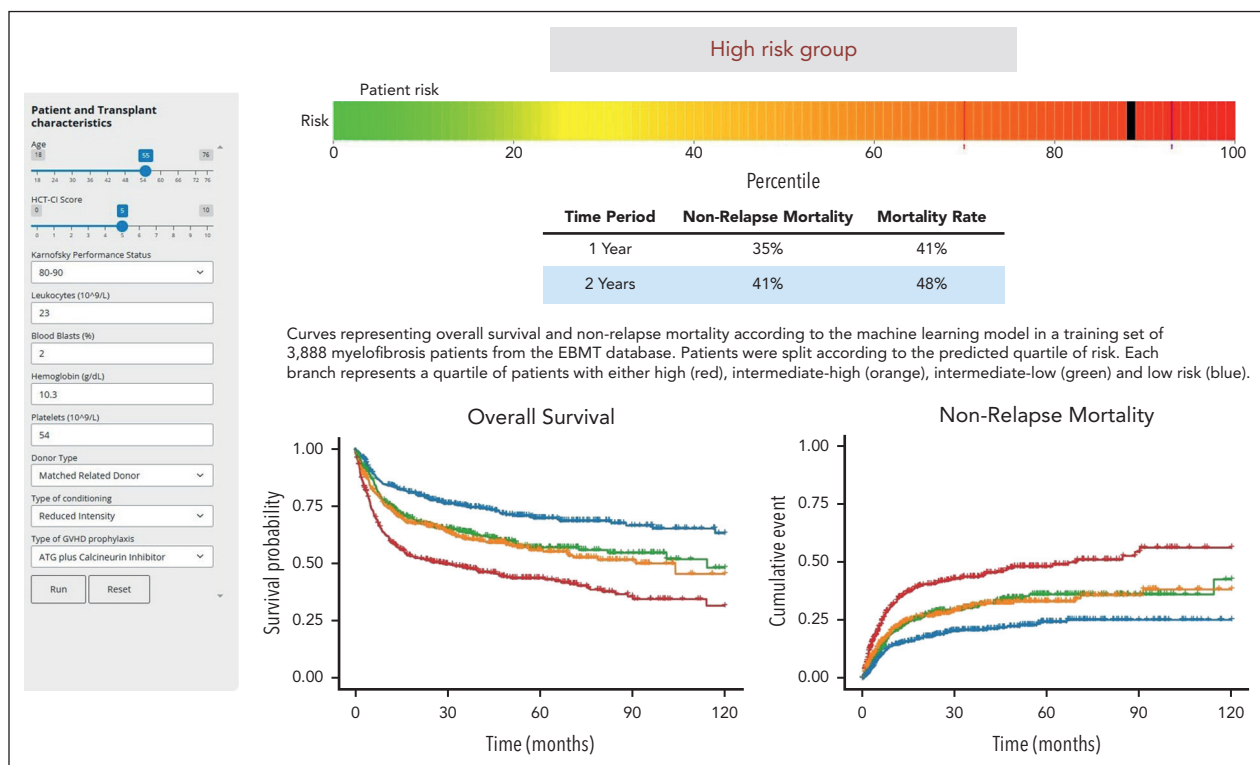


Figure 4. Illustration of the web-based calculator for the ML model. Risk score for a hypothetical MF patient candidate for transplantation. ATG, antithymocyte globulin.

protocols. This complexity can significantly constrain the power of prognostic models, especially for detecting early posttransplant mortality, which may be influenced by acute and unforeseeable clinical events. We recognize that further investigation into potential, less obvious correlations among pretransplant risk factors could provide additional insights. However, our primary focus was on constructing a clinically actionable prognostic tool rather than conducting an in-depth mechanistic analysis of risk determinants. Additionally, the EBMT data set had a substantial rate of missing data for some variables. The ML method addressed this through data imputation, whereas our Cox model used the missing indicator method for variables with missing values, excluding those with a high degree of missingness (eg, hematologic parameters and spleen size) from score development. Although cross-validation was used to reduce overfitting in the ML model, it was not applied to the risk score developed using Cox regression. The lack of molecular annotation regarding additional somatic mutations, which has been shown to provide prognostic information after allo-HCT in some studies^{7,32} but not others,^{33,34} prevented a comparison with the Myelofibrosis Transplant Scoring System.⁷ Furthermore, data on the grade of bone marrow fibrosis and the variant allele frequency of driver mutations at transplant were not available. Future research could benefit from enhancing data completeness to potentially refine the model's prognostic accuracy.

In conclusion, this investigation compared the effectiveness of 2 different Cox regression-derived models with a ML-driven approach for stratifying risk in patients with MF undergoing allo-HCT. The results demonstrate that the ML-driven model outperformed traditional statistical approaches by providing enhanced generalizability and identifying a broader subset of

patients at high risk for adverse outcomes. ML methods facilitate the modeling of complex interactions and nonlinear associations more effectively than traditional statistical methods. However, our findings also underscore the challenges in predicting early posttransplant mortality based on conventional baseline characteristics, which remain difficult to anticipate with current prognostic tools. To improve clinical decision-making, we have developed a novel prognostic tool using ML techniques that can identify 25% of patients at high risk for mortality after transplantation. The web-based calculator (<https://gemfin.click/ebmt>) represents a significant advance toward personalized medicine for patients with MF, enabling better strategic planning and potentially improving outcomes. As we move forward, refining this tool through the integration of more comprehensive data and ongoing validation will be crucial to fully realize its clinical potential.

Acknowledgments

The authors are grateful to all the centers and patients contributing to the European Society for Blood and Marrow Transplantation database.

Authorship

Contribution: J.C.H.-B. and A.M.-O. conceived the idea and developed the project proposal; A.M.-O. performed the machine learning analysis, created figures and tables, and cowrote the first draft of the manuscript with J.C.H.-B.; L.G. and J.C.H.-B. developed the Cox regression-based prognostic model for survival; L.K. and J.T. managed the study data; J.R. contributed to the tables and cowrote part of the first draft; C.P.M. and D.C. designed the online calculator; and all other coauthors contributed data to the study, critically revised the paper, and approved the final version.

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

A complete list of the centers from the Chronic Malignancies Working Party of the EBMT that participated in this study appears in the supplemental Appendix.

ORCID profiles: J.C.H.-B., [0000-0002-4289-3113](https://orcid.org/0000-0002-4289-3113); A.M.-O., [0000-0003-4838-6750](https://orcid.org/0000-0003-4838-6750); N.K., [0000-0001-5103-9966](https://orcid.org/0000-0001-5103-9966); M.R., [0000-0003-1388-9876](https://orcid.org/0000-0003-1388-9876); J.P., [0000-0001-7092-3351](https://orcid.org/0000-0001-7092-3351); I.Y.-A., [0000-0003-4524-8782](https://orcid.org/0000-0003-4524-8782); R.V.D., [0000-0003-4748-7648](https://orcid.org/0000-0003-4748-7648); M.S., [0000-0002-9331-5145](https://orcid.org/0000-0002-9331-5145); J.V., [0000-0002-4274-3895](https://orcid.org/0000-0002-4274-3895); U.P., [0000-0003-1863-3239](https://orcid.org/0000-0003-1863-3239); F.B., [0000-0002-2944-3812](https://orcid.org/0000-0002-2944-3812); K.C., [0000-0002-0303-1350](https://orcid.org/0000-0002-0303-1350); C.P.M., [0009-0004-7972-9431](https://orcid.org/0009-0004-7972-9431); D.C., [0009-0006-9458-9793](https://orcid.org/0009-0006-9458-9793); K.R., [0000-0002-8258-354X](https://orcid.org/0000-0002-8258-354X); J.D.-S., [0000-0002-4562-6264](https://orcid.org/0000-0002-4562-6264); G.B., [0000-0002-5201-1786](https://orcid.org/0000-0002-5201-1786); N.P., [0000-0001-6297-9697](https://orcid.org/0000-0001-6297-9697); T.C., [0000-0003-3108-4035](https://orcid.org/0000-0003-3108-4035).

Correspondence: Juan Carlos Hernández-Boluda, Hematology Department, Hospital Clínico Universitario, Avd Blasco Ibáñez 17, 46010 Valencia, Spain; email: hernandez_jca@gva.es; and Donal P. McLornan, Department of Haematology and Stem Cell Transplantation, University College London Hospitals NHS Trust, 3rd Floor W Wing, 250 Euston Rd, London NW1 2PG, United Kingdom; email: donal.mclornan@nhs.net.

REFERENCES

1. Passamonti F, Mora B. Myelofibrosis. *Blood*. 2023;141(16):1954-1970.
2. Tefferi A. Primary myelofibrosis: 2023 update on diagnosis, risk-stratification, and management. *Am J Hematol*. 2023;98(5):801-821.
3. Kröger N, Bacigalupo A, Barbui T, et al. Indication and management of allogeneic haematopoietic stem-cell transplantation in myelofibrosis: updated recommendations by the EBMT/ELN International Working Group. *Lancet Haematol*. 2024;11(1):e62-e74.
4. Vachhani P, Verstovsek S, Bose P. Disease modification in myelofibrosis: an elusive goal? *J Clin Oncol*. 2022;40(11):1147-1154.
5. England J, Gupta V. Novel therapies vs hematopoietic cell transplantation in myelofibrosis: who, when, how? *Hematol Am Soc Hematol Educ Program*. 2021;2021(1):453-462.
6. Maze D, Arcasoy MO, Henrie R, et al. Upfront allogeneic transplantation versus JAK inhibitor therapy for patients with myelofibrosis: a North American collaborative study [published correction appears in *Bone Marrow Transplant*. 2024;59(2):196-202]. *Bone Marrow Transpl*. 2024;59(2):196-202.
7. Gagelmann N, Ditschkowski M, Bogdanov R, et al. Comprehensive clinical-molecular transplant scoring system for myelofibrosis undergoing stem cell transplantation. *Blood*. 2019;133(20):2233-2242.
8. Tamari R, McLornan DP, Ahn KW, et al. A simple prognostic system in patients with myelofibrosis undergoing allogeneic stem cell transplantation: a CIBMTR/EBMT analysis. *Blood Adv*. 2023;7(15):3993-4002.
9. McLornan D, Eikema DJ, Czerw T, et al. Trends in allogeneic haematopoietic cell transplantation for myelofibrosis in Europe between 1995 and 2018: a CMWP of EBMT retrospective analysis. *Bone Marrow Transpl*. 2021;56(9):2160-2172.

10. Hernández-Boluda JC, Pereira A, Alvarez-Larran A, et al. Predicting survival after allogeneic hematopoietic cell transplantation in myelofibrosis: performance of the myelofibrosis transplant scoring system (MTSS) and development of a new prognostic model. *Biol Blood Marrow Transpl*. 2020;26(12):2237-2244.
11. Mosquera-Orqueira A, Pérez-Encinas M, Hernández-Sánchez A, et al. Machine learning improves risk stratification in myelofibrosis: an analysis of the Spanish Registry of Myelofibrosis. *Hemasphere*. 2023;7(1):e818.
12. Mosquera-Orqueira A, Arellano-Rodrigo E, Garrote M, et al. Integrating AIPSS-MF and molecular predictors: a comparative analysis of prognostic models for myelofibrosis. *Hemasphere*. 2024;8(3):e60.
13. Bacigalupo A, Ballen K, Rizzo D, et al. Defining the intensity of conditioning regimens: working definitions. *Biol Blood Marrow Transpl*. 2009;15(12):1628-1633.
14. Hernández-Boluda JC, Pereira A, Kröger N, et al. Determinants of survival in myelofibrosis patients undergoing allogeneic hematopoietic cell transplantation. *Leukemia*. 2021;35(1):215-224.
15. Polverelli N, Bonneville EF, de Wreede LC, et al. Impact of comorbidities and body mass index on the outcomes of allogeneic hematopoietic cell transplantation in myelofibrosis: a study on behalf of the Chronic Malignancies Working Party of EBMT. *Am J Hematol*. 2024;99(5):993-996.
16. Copelan E, Casper JT, Carter SL, et al. A scheme for defining cause of death and its application in the T cell depletion trial. *Biol Blood Marrow Transpl*. 2007;13(12):1469-1476.
17. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496-509.
18. R Foundation. The R Project for Statistical Computing. Accessed 23 April 2024. <https://www.r-project.org/>

19. Therneau TM, Lumley T, Elizabeth A, Cynthia C. A package for survival analysis in R. Accessed 23 April 2024. <https://CRAN.R-project.org/package=survival>
20. Gerds TA. prodlm: product-limit estimation for censored event history analysis. Accessed 23 April 2024. <https://CRAN.R-project.org/package=prodlm>
21. Gray B. cmprsk: subdistribution analysis of competing risks. Accessed 23 April 2024. <https://CRAN.R-project.org/package=cmprsk>
22. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat*. 2008;2(3):841-860.
23. Jaeger BC, Long DL, Long DM, et al. Oblique random survival forests. *Ann Appl Stat*. 2019;13(3):1847-1883.
24. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery; 2016:785-794.
25. Chen Bingshu E. A package of deep neural network tools for probability models. Accessed 1 May 2024. <https://CRAN.R-project.org/package=dnns>
26. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387.
27. 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(30):5381-5397.
28. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19(6):716-723.

Footnotes

Submitted 15 October 2024; accepted 11 February 2025; prepublished online on *Blood* First Edition 27 March 2025. <https://doi.org/10.1182/blood.2024027287>.

*J.C.H.-B. and A.M.-O. contributed equally to this study.

The data that support the findings of this study are available on request from the corresponding authors, Juan Carlos Hernández-Boluda (hernandez_jca@gva.es) and Donal P. McLornan (donal.mclornan@nhs.net). The data are not publicly available because of privacy or ethical restrictions.

The online version of this article contains a data supplement.

There is a [Blood Commentary](#) on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

29. Wal WMvd, Geskus RB. ipw: an R package for inverse probability weighting. *J Stat Softw.* 2011;43(13):1-23.
30. Kröger N, Wolschke C, Gagelmann N. How I treat transplant-eligible patients with myelofibrosis. *Blood.* 2023;142(20):1683-1696.
31. Mussetti A, Rius-Sansalvador B, Moreno V, et al. Artificial intelligence methods to estimate overall mortality and non-relapse mortality following allogeneic HCT in the modern era: an EBMT-TCWP study. *Bone Marrow Transpl.* 2024;59(2):232-238.
32. Kröger N, Panagiota V, Badbaran A, et al. Impact of molecular genetics on outcome in myelofibrosis patients after allogeneic stem cell transplantation. *Biol Blood Marrow Transpl.* 2017;23(7):1095-1101.
33. Ali H, Aldoss I, Yang D, et al. MIPSS70+ v2.0 predicts long-term survival in myelofibrosis after allogeneic HCT with the Flu/Mel conditioning regimen. *Blood Adv.* 2019;3(1):83-95.
34. Tamari R, Rapaport F, Zhang N, et al. Impact of high-molecular-risk mutations on transplantation outcomes in patients with myelofibrosis. *Biol Blood Marrow Transpl.* 2019;25(6):1142-1151.

© 2025 American Society of Hematology. Published by Elsevier Inc. Licensed under [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International \(CC BY-NC-ND 4.0\)](#), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.