blood

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Prediction of Poor Survival after Hematopoietic Cell Transplantation in Myelofibrosis Using Machine Learning Techniques

Tracking no: BLD-2024-027287R2

Juan Carlos Hernandez-Boluda (Hospital Clínico Valencia, Spain) Adrian Mosquera Orgueira (University Hospital of Santiago de Compostela, Spain) Luuk Gras (EBMT Statistical Unit, Netherlands) Linda Koster (EBMT Data Office, Netherlands) Joe Tuffnell (EBMT Leiden Study Unit, Netherlands) Nicolaus Kröger (University Medical Center Hamburg_Eppendorf, Germany) Massimiliano Gambella (IRCCS Ospedale Policlinico San Martino, Italy) Thomas Schroeder (University of Duesseldorf, Germany) Marie Robin (Hopital Saint Louis, France) Katja Sockel (University Hospital Dresden, Germany) Jakob Passweg (University Hospital Basel, Switzerland) Igor-Wolfgang Blau (Charité, Germany) Ibrahim Yakoub-Agha (Cellular therapy unit, France) Ruben Van Dijck (Erasmus MC Cancer Institute, Netherlands) Matthias Stelljes (University of Muenster, Germany) Henrik Sengeloev (Rigshospitalet,) Jan Vydra (Institute of Hematology and Blood Transfusion, Czech Republic) Uwe Platzbecker (University Hospital Leipzig, Department of Hematology and Cell Therapy, Germany) Moniek Dewitte (UMC Utrecht, Netherlands) Frédéric Baron (University of Liege, Belgium) Kristina Carlson (Uppsala University Hospital, Sweden) Javier Rojas Martínez (Hospital de Santiago de Compostela, Spain) Carlos Pérez Míguez (Instituto de Investigación Sanitaria de Santiago, Spain) Davide Crucitti (Health Research Institute of Santiago de Compostela (IDIS), Spain) Kavita Raj (University College London Hospitals NHS Foundation Trust, United Kingdom) Joanna Drozd-Sokolowska (Medical University of Warsaw, Poland) Giorgia Battipaglia (Department of Clinical Medicine and Surgery, Hematology Unit, Federico II University, Naples, Italy, Italy) Nicola Polverelli (Fondazione IRCCS Policlinico San Matteo, Italy) Tomasz Czerw (Maria Skłodowska-Curie National Research Institute of Oncology, Gliwice, Poland, Poland) Donal McLornan (University College Hospital, United Kingdom)

Abstract:
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 5,183 MF patients who underwent first allo-HCT between 2005 and 2020 at EBMT centers, we examined different machine learning (ML) models to predict overall survival (OS) 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 four-level Cox regression-based score and other ML-based models derived from the same dataset, and with the CIBMTR score. The RSF outperformed all comparators, achieving better concordance indices across both primary and post-essential thrombocythemia/polycythemia vera MF subgroups. The robustness and generalizability of the RSF model was confirmed by Akaike's Information Criterion and time-dependent Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) metrics in both sets. While all models were prognostic for non-relapse mortality, the RSF provided better curve separation, effectively identifying a high-risk group comprising 25% of patients. In conclusion, ML enhances risk stratification in MF patients 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.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: JCHB and AMO conceived the idea and developed the project proposal. AMO performed the machine learning analysis, created figures and tables, and co-wrote the first draft of the manuscript with JCHB. LG and JCHB developed the Cox regression-based prognostic model for survival. LK and JT managed the study data. JR contributed to the tables and co-wrote part of the first draft. CPM and DC designed the online calculator. All other co-authors contributed data to the study, critically revised the paper, and approved the final version.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: No prior presentation of the data. The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Clinical trial registration information (if any):

A Machine Learning Model To Predict Survival After Allogeneic Hematopoietic Cell Transplantation in Patients with Myelofibrosis

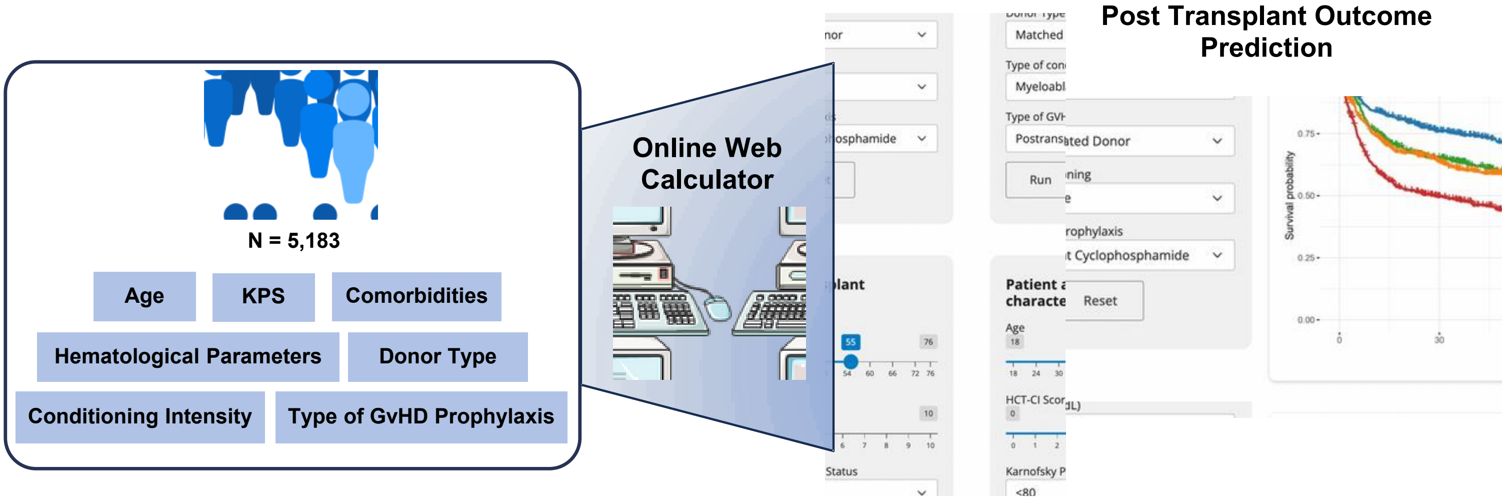
Context of Research

Allogeneic transplantation remains the only curative treatment for patients with MF. However, its significant morbidity and mortality require a careful risk-benefit analysis to identify suitable candidates.

Aim of This Study

Using data from 5,183 MF patients who underwent allogeneic transplantation at EBMT centers, we examined different machine learning models to predict overall survival after transplant.

Findings



Images are in part from Servier Medical Art, which is licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/)

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

*Hernández-Boluda et al. DOI: 10.xxxx/**blood**.2024xxxxxx*

blood Visual Abstract

Prediction of Poor Survival after Hematopoietic Cell Transplantation in Myelofibrosis Using Machine Learning Techniques

Running head: Predicting survival after allo-HCT in MF

Juan Carlos Hernández-Boluda[1a,b], Adrián Mosquera-Orgueira[2a], 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 Dijck[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. McLornan[20b], on behalf of the Chronic Malignancies Working Party of EBMT.

1. Hospital Clínico Universitario-INCLIVA, University of Valencia, Valencia, Spain
2. University Hospital of Santiago de Compostela, IDIS, Santiago de Compostela, Spain
3. EBMT Leiden Study Unit, Leiden, the Netherlands
4. University Medical Center Hamburg-Eppendorf, Hamburg, Germany
5. Ematologia e Terapie Cellulari. IRCCS Ospedale Policlinico San Martino, Genova, Italy
6. University Hospital Essen, Duesseldorf, Germany
7. Saint-Louis Hospital, BMT Unit, Paris, France
8. University Hospital TU Dresden, Dresden, Germany
9. University Hospital Basel, Basel, Switzerland
10. Medizinische Klinik m. S. Hämatologie , Onkologie und Tumorimmunologie, Berlin, Germany
11. CHU de Lille, INSERM U1286, Infinite, 59000, Lille, France
12. Erasmus MC Cancer Institute, Rotterdam, the Netherlands
13. University of Muenster, Muenster, Germany
14. Rigshospitalet, Copenhagen, Denmark
15. Institute of Hematology and Blood Transfusion, Prague, Czech Republic
16. Medical Clinic and Policlinic, Leipzig, Germany

17. University Medical Centre, Utrecht, the Netherlands
18. University of Liege, Liege, Belgium
19. University Hospital Uppsala, Uppsala, Sweden
20. University College London Hospitals NHS Trust, London, United Kingdom
21. Central Clinical Hospital, The Medical University of Warsaw, Warsaw, Poland
22. Hematology Department, Federico II University of Naples, Naples, Italy
23. Unit of Bone Marrow Transplantation – Division of Hematology, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy
24. Maria Skłodowska-Curie National Research Institute of Oncology, Gliwice, Poland

[a]Contributed equally

[b]Juan Carlos Hernández-Boluda and Donal P McLornan are corresponding authors.

Corresponding authors:

Juan Carlos Hernandez-Boluda, MD, Ph.D. Hematology Department, Hospital Clínico Universitario, INCLIVA Research Institute, Department of Medicine, University of Valencia, Avd. Blasco Ibáñez 17, 46010, Valencia, Spain. Phone: +34-96-1973842. Fax: +34-96-1973839. E-mail: hernandez_jca@gva.es

Donal P McLornan, MD, Ph.D. Department of Haematology and Stem cell Transplantation, University College London Hospitals NHS Trust, 3rd Floor West Wing, 250 Euston Road, London NW1 2PG, United Kingdom. Phone: +44- 203 447 8835. Fax: NA. Email: donal.mclornan@nhs.net

Text word count: 3980. Abstract word count: 249. Tables: 3. Figures: 4. References: 34. Supplementary material: 1
Scientific category: Transplantation

All authors declare no competing financial interests. This study was not economically funded.

Data sharing statement: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

This retrospective study was approved by the Chronic Malignancies Working Party (CMWP) of EBMT. Informed consent for inclusion in the EBMT registry was obtained in all patients.

Key points

- A machine learning model, available as an interactive web application, was created to predict survival after transplant in myelofibrosis.

- This tool is a step toward personalized medicine, enabling the identification of 25% of patients with poor transplantation outcomes.

Abstract

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 5,183 MF patients who underwent first allo-HCT between 2005 and 2020 at EBMT centers, we examined different machine learning (ML) models to predict overall survival (OS) 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 four-level Cox regression-based score and other ML-based models derived from the same dataset, and with the CIBMTR score. The RSF outperformed all comparators, achieving better concordance indices across both primary and post-essential thrombocythemia/polycythemia vera MF subgroups. The robustness and generalizability of the RSF model was confirmed by Akaike's Information Criterion and time-dependent Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) metrics in both sets. While all models were prognostic for non-relapse mortality, the RSF provided better curve separation, effectively identifying a high-risk group comprising 25% of patients. In conclusion, ML enhances risk stratification in MF patients 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.

Keywords: myelofibrosis, survival, transplantation, prognostication, machine learning.

Introduction

Myelofibrosis (MF) is a chronic myeloproliferative neoplasm (MPN) that appear *de novo* (primary, PMF) or following a diagnosis of essential thrombocythemia or polycythemia vera (secondary, SMF). Managing MF is complex due to its diverse clinical manifestations including an inherent risk of progression to acute myeloid leukemia (AML). While the median overall survival (OS) is around 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 comorbidities[7] and emerging transplant strategies, such as haploidentical transplants or post-transplant cyclophosphamide use.[9,10] Furthermore, there is significant room for improving their ability to accurately identify patients at high risk of post-transplant 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 MF patients 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 MF (PMF or SMF) patients who underwent first allo-HCT between 2005 and 2020 in EBMT centers. Patients transplanted from umbilical cord blood or with AML transformation were excluded. Myeloablative (MAC) or reduced intensity conditioning (RIC) 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 (CMWP) of EBMT and conducted in accordance with the Declaration of Helsinki. 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 dataset, and with the CIBMTR model.[8] The models were also applied for predicting the secondary outcome of non-relapse 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 multivariable Cox proportional hazards regression model, while the other employed 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=3,887) and a test set (25% of the cohort, n=1,296). Each statistician independently determined the optimal cut-off 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] *prodlim*,[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, whilst 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 cut-off values were arbitrarily defined as follows: HR <1.25 = 0 points, HR 1.25-1.50 = 1 point, HR >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 two-sided and $p<0.05$ was considered significant.

Random survival forest model

Random survival forests (RSF) were created with 1,000 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 employed 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 (PDPs) 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 non-linear 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 three complementary methods using the 10 variables included in the final model: Oblique Random Survival Forests (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 three approaches—ORSF, XGBoost-based survival modeling, and DeepSurv—were evaluated using cross-validation for performance estimation. Early stopping was applied where computationally practical to limit overfitting. The performance of each model was assessed using standard survival metrics (e.g., 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 Harrell's 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.

Since the c-index is not an optimal metric for competing risk models, we also included Akaike's information criterion (AIC) scores for both NRM and OS.[28]

Results

Patient and transplant characteristics

A total of 5,183 patients from 288 centers fulfilling the selection criteria were included. Baseline characteristics of all patients, along with the training (n=3,887) and validation (n=1,296) 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 ATG 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 S1.**

Risk model for OS using Cox regression analysis

Factors associated with OS in the univariable analysis are shown in **Supplemental Table S1**. In the multivariable analysis, 7 independent factors significantly predicted reduced OS: older patient age, HLA mismatched donor type, lower Karnofsky performance status (KPS), higher Hematopoietic Cell Transplantation Comorbidity Index (HCT-CI), JAK2/triple genotype, graft from a female donor to a male patient and graft from a CMV-positive donor to a CMV-negative patient (**Table 2**). Based on the HRs, a score of 2 was assigned to patient age ≥60 years and the JAK2/triple negative genotype, and a score of 1 to patient age 50-59 years, haploidentical or mismatched unrelated donors, KPS <90, and HCT-CI ≥3. As 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 (2-3 points), intermediate-2 (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 1 A-B**).

Risk model for OS using Random Survival Forests

A RSF model was created to predict OS using the 52 initial variables of the dataset. This model achieved a c-index of 0.603 in the training set and 0.632 in the test set. The variable importance metrics for the model in the training set are shown in **Supplemental Figure S2**. After dimensionality reduction, the model was refined to a smaller set of key prognostic variables: patient age, HCT-CI, KPS, blood blasts %, hemoglobin level, leukocyte and platelet counts, donor type, conditioning intensity and 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 S2**). To further elucidate these relationships, we have also included partial dependence plots in **Supplemental Figure S3**.

Comparison of the RSF with other ML techniques

As shown in **Supplemental Table S3**, RSF achieved higher concordance indices for OS and NRM predictions in both training and test sets compared to three alternative methods (ORSF, DeepSurv, 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=1,773; 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 to 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 four equal groups within the training set and applied the same classification thresholds to the test set **(Figure 1 C-D)**. The ML model maintained similar c-indices post-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 2 A-B**). The time-dependent ROC AUCs comparing both models are presented in **Supplementary Figure S4**.

Comparison of the ML model with the CIBMTR model

We compared the performance of the ML model with the CIBMTR scoring system[8] in a comparable subset of patients with complete annotations for the CIBMTR score. By integrating patient age, Hb level at transplant, and donor type, this model defined three risk categories in the original series, with a 3-year post-transplant 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=1,925) and 0.654 vs. 0.581 in the test set (n=618) (**Table 3**, **Supplemental Figure S5**). 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 **Supplementary Figure S6**. The difference between the CIBMTR score and ML method was mostly driven by prognostic refinement within the CIBMTR intermediate risk group, where the ML algorithm reclassified most patients into different risk categories (**Figure 2 C-D**).

Notably, the ML model exhibited a consistently better discriminative performance than the Cox-derived models for both PMF and SMF patients, with the advantage being more pronounced in the test set **(Supplementary Table S4**).

Application of the ML model to predict NRM

In predicting NRM, the ML model achieved comparable AIC scores to 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 to 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 post-transplant 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 to 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%, 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%, nearly matching the Cox score's 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 S7**.

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

For the optimal donor type, the univariable Cox proportional hazards model indicated that being transplanted 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=0.001) compared to non-optimal 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=0.89), 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=0.19) compared to those receiving non-optimal intensities. However, the IPW-adjusted analysis showed a HR of 1.025 (95% CI, 0.07-14.80, p=0.99), 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=0.62), while the IPW-adjusted analysis resulted in a HR of 0.95 (95% CI: 0.43-2.09, p=0.90). These results indicate no discernible impact of the ML-predicted optimal GVHD prophylaxis on patient survival.

Discussion

In the present study, we have developed a ML model to enhance risk stratification for MF patients undergoing allo-HCT, utilizing a large database of 5,183 MF patients from the EBMT registry. Notably, this model is particularly comprehensive, as 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 post-transplant mortality. Notably, the improved performance of the ML model was evident in both PMF and SMF patients. The ML model's 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, as the model lacks sufficient discrimination within this range. While 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 (around 35% NRM rate and 40% overall mortality at 1 year). Moreover, the web-based calculator permits the identification of a subset of very high-risk patients with a predicted 1-year OS of less than 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]

While 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 which 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 protocols. This complexity can significantly constrain the power of prognostic models, especially for detecting early post-transplant mortality, which may be influenced by acute and unforeseeable clinical events. We recognize that further investigation into potential, less obvious correlations among pre-transplant 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 dataset had a substantial rate of missing data for some variables. The ML method addressed this through data imputation, while our Cox model used the missing indicator method for variables with missing values, excluding those with a high degree of missingness (e.g., 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 MTSS model.[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 two different Cox regression-derived models with a ML-driven approach for stratifying risk in MF patients 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 modelling of complex interactions and non-linear associations more effectively than traditional statistical methods. However, our findings also underscore the challenges in predicting early post-transplant 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 towards personalized medicine for MF patients, 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.

Acknowledgements

We are grateful to all the centers and patients contributing to the EBMT database. The list of centers that participated in this study can be found in the **Appendix** of the Supplemental document.

Authorship Contributions

JCHB and AMO conceived the idea and developed the project proposal. AMO performed the machine learning analysis, created figures and tables, and co-wrote the first draft of the manuscript with JCHB. LG and JCHB developed the Cox regression-based prognostic model for survival. LK and JT managed the study data. JR contributed to the tables and co-wrote part of the first draft. CPM and DC designed the online calculator. All other co-authors contributed data to the study, critically revised the paper, and approved the final version.

Disclosure of Conflicts of Interest

The authors declare no competing interests.

References

1. Passamonti F, Mora B. Myelofibrosis. Blood. 2023;141(16):1954-1970. doi:10.1182/blood.2022017423
2. Tefferi A. Primary myelofibrosis: 2023 update on diagnosis, risk-stratification, and management. Am J Hematol. 2023;98(5):801-821. doi:10.1002/ajh.26857
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. doi:10.1016/S2352-3026(23)00305-8
4. Vachhani P, Verstovsek S, Bose P. Disease Modification in Myelofibrosis: An Elusive Goal?. J Clin Oncol. 2022;40(11):1147-1154. doi:10.1200/JCO.21.02246
5. England J, Gupta V. Novel therapies vs hematopoietic cell transplantation in myelofibrosis: who, when, how?. Hematology Am Soc Hematol Educ Program. 2021;2021(1):453-462. doi:10.1182/hematology.2021000279
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 Jan 26;:]. Bone Marrow Transplant. 2024;59(2):196-202. doi:10.1038/s41409-023-02146-6
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. doi:10.1182/blood-2018-12-890889

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. doi:10.1182/bloodadvances.2023009886
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 Transplant. 2021;56(9):2160-2172. doi:10.1038/s41409-021-01305-x
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 Transplant. 2020;26(12):2237-2244. doi:10.1016/j.bbmt.2020.07.022
11. Mosquera-Orgueira 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. 2022;7(1):e818. Published 2022 Dec 20. doi:10.1097/HS9.0000000000000818
12. Mosquera-Orgueira 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. Published 2024 Mar 20. doi:10.1002/hem3.60
13. Bacigalupo A, Ballen K, Rizzo D, et al. Defining the intensity of conditioning regimens: working definitions. Biol Blood Marrow Transplant. 2009;15(12):1628-1633. doi:10.1016/j.bbmt.2009.07.004

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. doi:10.1038/s41375-020-0815-z

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. doi:10.1002/ajh.27262

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 Transplant. 2007;13(12):1469-1476. doi:10.1016/j.bbmt.2007.08.047

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. doi:10.1080/01621459.1999.10474144.

18. R: The R Project for Statistical Computing. Accessed April 23, 2024. https://www.r-project.org/

19. Therneau T. A Package for Survival Analysis in R. Version 3.5-7. 2023. Accessed April 23, 2024. https://CRAN.R-project.org/package=survival

20. Gerds TA. prodlim: Product-Limit Estimation for Censored Event History Analysis. Version 2023.08.28. 2023. Accessed April 23, 2024. https://CRAN.R-project.org/package=prodlim

21. Gray B. cmprsk: Subdistribution Analysis of Competing Risks. Version 2.2-12. 2024. Accessed April 23, 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. doi:10.1214/08-AOAS169

23. Jaeger BC, Long DL, Long DM, et al. OBLIQUE RANDOM SURVIVAL FORESTS. Ann Appl Stat. 2019;13(3):1847-1883. doi:10.1214/19-aoas1261

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. KDD '16. New York, NY: Association for Computing Machinery; 2016:785-794.

25. Bingshu E Chen (2022). A Package of deep nerual network tools for probability models . R package version 0.0.6, https://CRAN.R-project.org/package=dnn.

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 Feb 28;15(4):361-87. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4. PMID: 8668867.

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. doi:10.1002/sim.5958

28. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control.* 1974;19(6):716-723. doi:10.1109/TAC.1974.1100705.

29. Wal WM van der, Geskus RB. ipw: An R Package for Inverse Probability Weighting. J Stat Softw. 2011;43:1-23. doi:10.18637/jss.v043.i13

30. Kröger N, Wolschke C, Gagelmann N. How I treat transplant-eligible patients with myelofibrosis. Blood. 2023 Nov 16;142(20):1683-1696. doi: 10.1182/blood.2023021218. PMID: 37647853.

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 Transplant. 2024;59(2):232-238. doi:10.1038/s41409-023-02147-5

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 Transplant. 2017;23(7):1095-1101. doi:10.1016/j.bbmt.2017.03.034

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. doi:10.1182/bloodadvances.2018026658

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 Transplant. 2019;25(6):1142-1151. doi:10.1016/j.bbmt.2019.01.002

35. **Table 1.** Main characteristics of a series of 5183 patients with myelofibrosis undergoing allogeneic hematopoietic cell transplantation and of the training and test cohorts

Characteristic	Group	Missing (%)	Total cohort N (%)	Training set N (%)	Test set N (%)	P
No. of patients			5183 (100%)	3887 (100%)	1296 (100%)	
Age at allo-HCT, years[1]			58.3 (52-63.5)	58.2 (51.8-63.4)	58.6 (52.7-63.8)	0.072
Age at allo-HCT	<60 years		3003 (57.9%)	2274 (58.5%)	729 (56.2%)	0.164
Patient sex	Male		3242 (62.6%)	2412 (62.1%)	830 (64.0%)	0.21
Year of myelofibrosis diagnosis	<2000		615 (11.9%)	455 (11.7%)	160 (12.3%)	0.21
	2000-2010		1765 (34.1%)	1310 (33.7%)	455 (35.1%)	
	2010-2015		1422 (27.4%)	1057 (27.2%)	365 (28.2%)	
	2015 or later		1381 (26.6%)	1065 (27.4%)	316 (24.4%)	
Myelofibrosis type	Primary myelofibrosis		3743 (72.2%)	2807 (72.2%)	936 (72.2%)	1.00
	Secondary myelofibrosis		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%)	0.459
Genotype	JAK2+	2422 (46.7%)	2156 (78.1%)	1625 (77.7%)	531 (79.3%)	0.43
	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%)	0.255
Hemoglobin at allo-HCT, g/dL[1]		2575 (49.7%)	9.2 (8.1-10.5)	9.2 (8.1-10.5)	9.2 (8.2-10.6)	0.43
Leukocyte count at allo-HCT, $x10^9/L$[1]		2600 (50.2%)	6.9 (3.5-14.4)	6.8 (3.5-14.3)	7 (3.6-14.4)	0.336
Blood blasts at allo-HCT, %[1]		3095 (59.7%)	1 (0-3)	1 (0-3)	1 (0-3)	0.637
Platelets at allo-HCT, $x10^9/L$[1]		2640 (50.9%)	117 (53-242.5)	114.5 (53-237)	125 (51-274)	0.039
Splenectomy prior to allo-HCT	Yes	2536 (48.9%)	324 (12.2%)	231 (11.7%)	93 (14.0%)	0.129
Spleen size below costal margin, cm by physical examination[1]		3857 (74.4%)	5 (0-10)	5 (1-11)	5 (0-10)	0.184
Spleen span by ultrasound or CT scan, max diameter cm[1]		4283 (82.6%)	20 (16-23)	20 (16-23)	20 (16.2-23)	0.72
HCT-CI risk group	Low risk (0 points)	1360 (26.2%)	2041 (53.4%)	1538 (53.7%)	503 (52.4%)	0.745
	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%)	
Karnofsky score at allo-HCT	90-100	489 (9.4%)	3111 (66.3%)	2316 (65.8%)	795 (67.7%)	0.508
	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%)	0.286
	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%)	

CIBMTR risk score at allo-HCT	Low	2640 (50.9%)	1020 (40.1%)	763 (39.6%)	257 (41.6%)	0.641
	Intermediate		1313 (51.6%)	1004 (52.2%)	309 (50.0%)	
	High		210 (8.3%)	158 (8.2%)	52 (8.4%)	
Donor type	Identical sibling		1534 (29.6%)	1147 (29.5%)	387 (29.9%)	0.919
	MRD (other than sibling)		45 (0.9%)	34 (0.9%)	11 (0.8%)	
	MMRD		339 (6.5%)	249 (6.4%)	90 (6.9%)	
	MUD		2175 (42.0%)	1646 (42.3%)	529 (40.8%)	
	MMUD		673 (13.0%)	504 (13.0%)	169 (13.0%)	
	Unrelated, number of mismatches unknown		417 (8.0%)	307 (7.9%)	110 (8.5%)	
Recipient-donor match	Recipient male - donor male	76 (1.5%)	2297 (45.0%)	1714 (44.7%)	583 (45.9%)	0.614
	Recipient male - donor female		895 (17.5%)	666 (17.4%)	229 (18.0%)	
	Recipient female - donor male		1147 (22.5%)	878 (22.9%)	269 (21.2%)	
	Recipient female - donor female		768 (15.0%)	578 (15.1%)	190 (14.9%)	
Donor age, years[1]		1132 (21.8%)	36.6 (27.2-49.8)	36.7 (27.2-49.6)	36.1 (27.1-50.1)	0.902
CMV serology in patient/donor	-/-	233 (4.5%)	1437 (29.0%)	1056 (28.4%)	381 (30.8%)	0.174
	-/+		464 (9.4%)	348 (9.4%)	116 (9.4%)	
	+/-		1002 (20.2%)	776 (20.9%)	226 (18.3%)	
	+/+		2047 (41.4%)	1533 (41.3%)	514 (41.6%)	
Stem cell source	Bone marrow		409 (7.9%)	305 (7.8%)	104 (8.0%)	0.884
	Peripheral blood		4774 (92.1%)	3582 (92.2%)	1192 (92.0%)	
Busulfan or melphalan based conditioning regimen	Busulfan based	89 (1.7%)	3522 (69.1%)	2642 (69.2%)	880 (69.0%)	0.861
	Melphalan based		718 (14.1%)	533 (14.0%)	185 (14.5%)	
	Other regimen		854 (16.8%)	644 (16.9%)	210 (16.5%)	
Conditioning drugs	BuCy +/- others (MAC)	146 (2.8%)	166 (3.3%)	137 (3.6%)	29 (2.3%)	0.324
	FluBu +/- others (MAC)		1153 (22.9%)	858 (22.7%)	295 (23.4%)	
	FluTreo +/- others (MAC)		160 (3.2%)	121 (3.2%)	39 (3.1%)	
	TBI +/- Cy +/- others (MAC)		135 (2.7%)	111 (2.9%)	24 (1.9%)	
	Others (MAC)		155 (3.1%)	116 (3.1%)	39 (3.1%)	
	FluBu +/- others (RIC)		2135 (42.4%)	1588 (42.1%)	547 (43.3%)	
	FluMel +/- others (RIC)		579 (11.5%)	428 (11.3%)	151 (12.0%)	
	FluTreo +/- others (RIC)		143 (2.8%)	105 (2.8%)	38 (3.0%)	
	TBI +/- Cy +/- Flu +/- others (RIC)		275 (5.5%)	206 (5.5%)	69 (5.5%)	
	Others (RIC)		136 (2.7%)	105 (2.8%)	31 (2.5%)	
Conditioning regimen intensity	MAC	75 (1.4%)	1802 (35.3%)	1366 (35.7%)	436 (34.1%)	0.32
	RIC		3306 (64.7%)	2463 (64.3%)	843 (65.9%)	
TBI	Yes	37 (0.7%)	420 (8.2%)	326 (8.4%)	94 (7.3%)	0.219
T-cel depletion	No	148 (2.9%)	1356 (26.9%)	1010 (26.8%)	346 (27.4%)	0.451

	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%)	0.038
Alemtuzumab	Yes	167 (3.2%)	306 (6.1%)	218 (5.8%)	88 (7.0%)	0.146
GvHD prophylaxis group	Post-Cy	112 (2.2%)	415 (8.2%)	311 (8.2%)	104 (8.2%)	0.064
	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 +/- 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+/- other		434 (8.6%)	311 (8.2%)	123 (9.7%)	
	CNI + MTX +/- other		522 (10.3%)	393 (10.3%)	129 (10.1%)	
	Other		129 (2.5%)	88 (2.3%)	41 (3.2%)	

[1]Median (Interquartile range)
Allo-HCT: allogeneic hematopoietic cell transplantation; HCT-CI: Hematopoietic Cell Transplantation Comorbidity Index;
MAC: myeloablative conditioning; RIC: reduced intensity conditioning;
Bu: busulfan; Cy: cyclophosphamide; Flu: fludarabine; Treo: treosulfan; TBI: total body irradiation; CNI: calcineurin inhibitor;
MMF: mycophenolate mofetil; MTX: methotrexate; ATG: antithymocyte globulin
GvHD: graft-versus-host disease; CMV: cytomegalovirus; TBI: total body irradiation;
ATG: anti-thymocyte globulin.

37. **Table 2.** Cox regression analysis of factors associated with overall survival in the training cohort

	HR (95% CI)	(Overall) p
Conditioning intensity		
MAC	1.00	
RIC	1.08 (0.97-1.20)	0.17
Donor type		
MRD/MUD	1.00	
HD/MMUD	1.34 (1.19-1.51)	<0.0001
Age at allo-HCT		
<49	1.00	
50-59	1.36 (1.17-1.58)	<0.0001
≥60	1.67 (1.44-1.94)	<0.0001
Sex		
Male	1.00	
Female	0.89 (0.80-0.99)	0.04
KPS at allo-HCT		
90-100	1.00	
<90	1.33 (1.19-1.48)	<0.0001
HCT-CI		
0-2	1.00	
≥3	1.36 (1.19-1.55)	<0.0001
CMV donor/patient		
+/-	1.16 (1.03-1.32)	0.01
other	1.00	
Sex match donor/patient		
Female->Male	1.15 (1.00-1.31)	0.04
other	1.00	
Genotype		
CALR+/MPL+	1.00	
JAK2+/Triple negative	1.56 (1.26-1.92)	<0.0001

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

HR: hazard risk; CI: confidence interval; MAC: myeloablative conditioning; RIC: reduced intensity conditioning;
MRD: matched related donor; MUD: matched unrelated donor; MMUD: mismatched unrelated donor; HD: haploidentical donor;
allo-HCT: allogeneic hematopoietic cell transplant; KPS: Karnofsky performance score;
HCT-CI: Hematopoietic Cell Transplantation-Comorbidity Index; CMV: cytomegalovirus.

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

Overall survival risk score for mortality prediction (Harrell's c-index)				
	ML model	ML model 4 groups	Cox-derived score	Cox-derived score 4 groups
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 4 groups		CIBMTR model 3 groups
Training Set (n=1925)**	0.608	0.599		0.557
Test Set (n=618)**	0.654	0.650		0.581

Overall survival risk score for mortality prediction (AIC)				
	ML model	ML model 4 groups	Cox-derived score	Cox-derived score 4 groups
Training Set (n=1773)*	9944	9954	9945	9954
Test Set (n=566)*	2757	2759	2765	2767
	ML model	ML model 4 groups		CIBMTR model 3 groups
Training Set (n=1925)**	12573	12588		12647
Test Set (n=618)**	3318	3322		3374

Overall survival risk score for non-relapse mortality prediction (AIC)				
	ML model	ML model 4 groups	Cox-derived score	Cox-derived score 4 groups
Training Set (n=1763)*	7210	7210	7208	7214
Test Set (n=566)*	2006	2002	2008	2006
	ML model	ML model 4 groups		CIBMTR model 3 groups
Training Set (n=1925)**	8738	8738		8762
Test Set (n=618)**	2154	2160		2182

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.

AIC: Akaike's information criterion.
Higher c-index values reflect better model performance in ranking predictions while lower AIC values indicate a better fit to the data.

Figure Legends

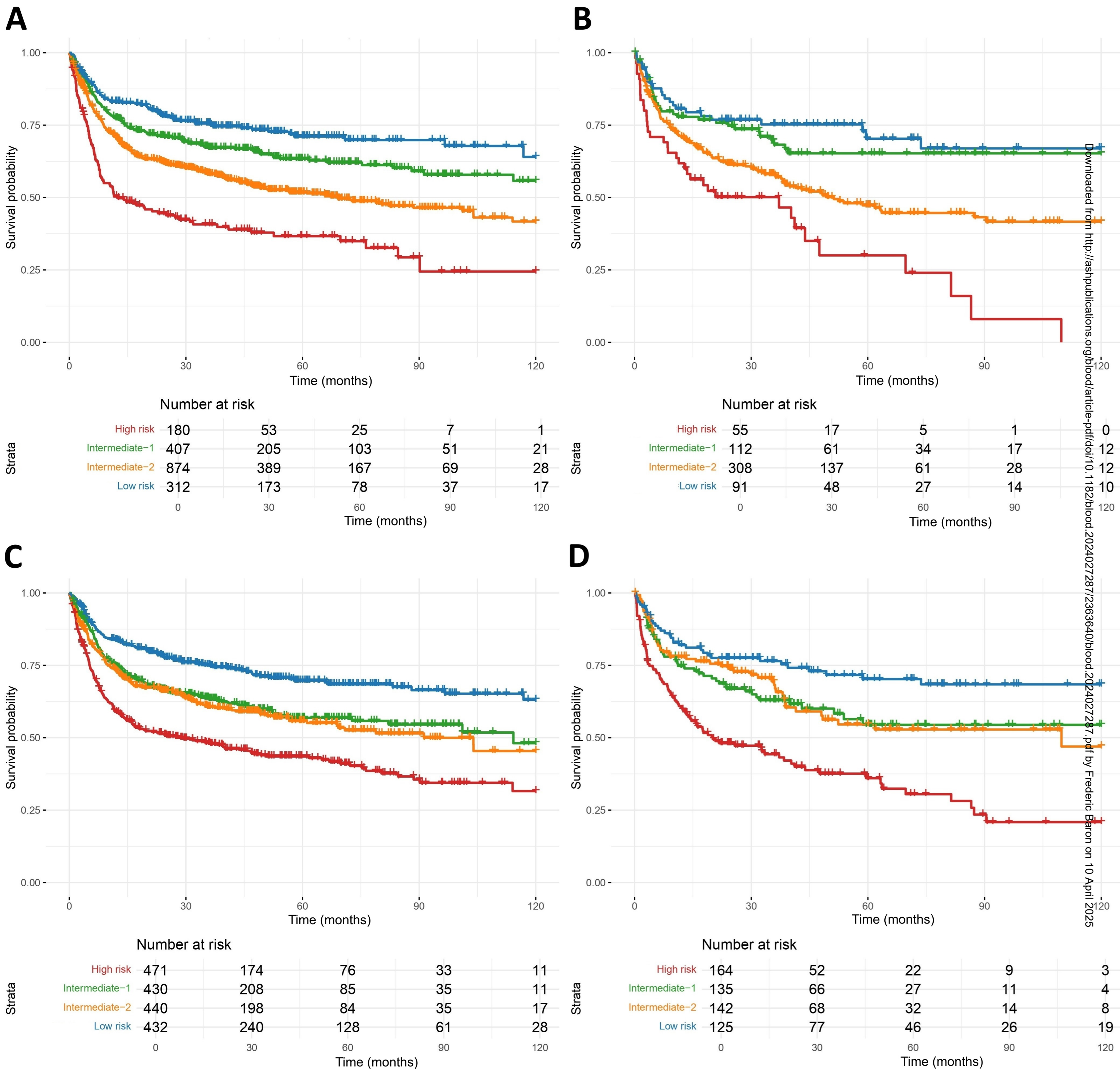
Figure 1. Kaplan-Meier curves illustrating overall survival after transplant based on risk groups defined by the prognostic models. A-B) Kaplan-Meier plots showing overall survival according to the Cox-regression statistical model in the training (A) and test (B) sets. C-D) Kaplan-Meier plots displaying overall survival 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 (yellow), and high risk (red).

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.

Figure 3. **Cumulative incidence of non-relapse mortality after transplant based on risk groups defined by the prognostic models.** A-B) Cumulative incidence of non-relapse mortality 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 four quartile groups according to their risk.

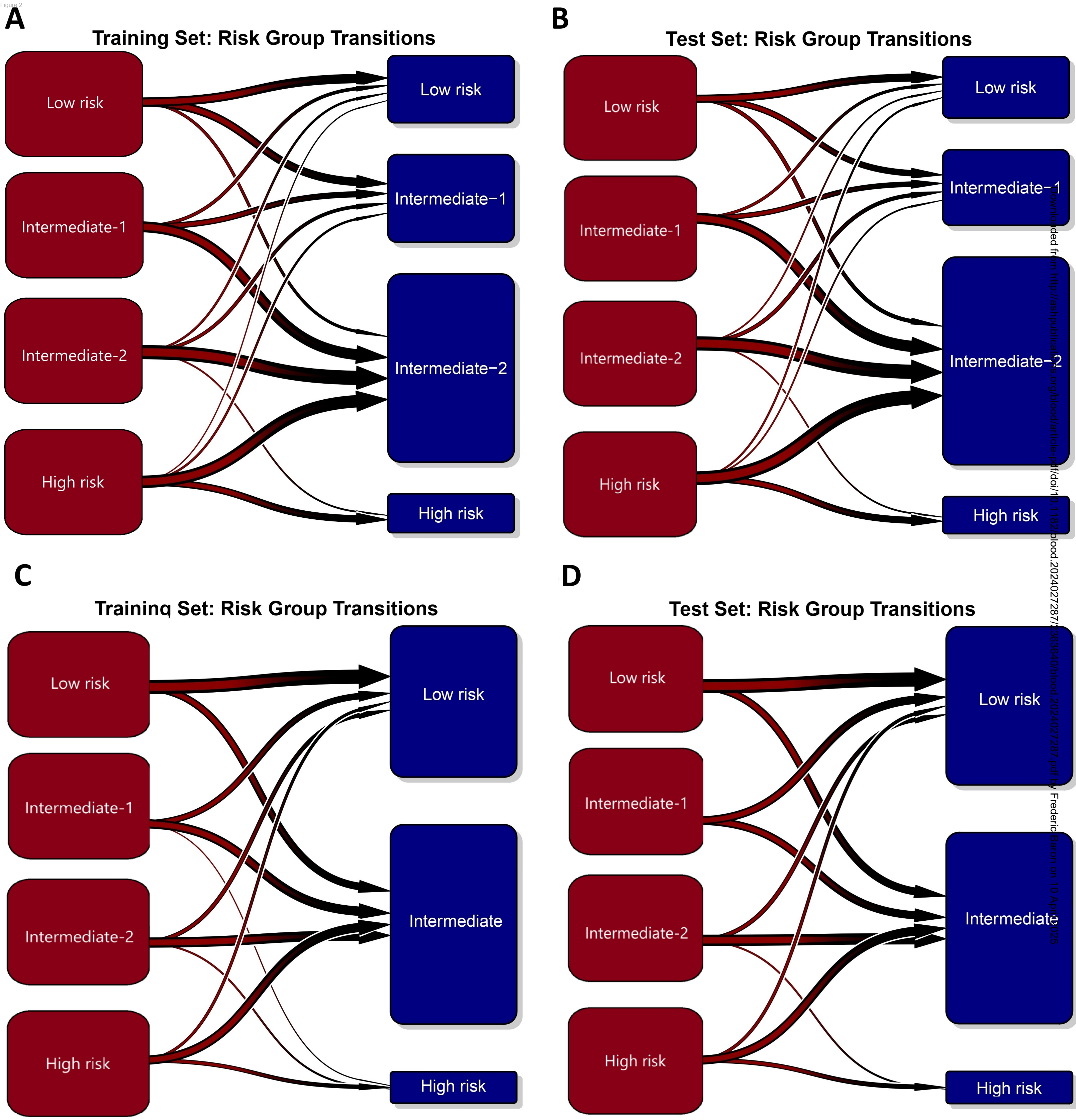
Figure 4. **Illustration of the web-based calculator for the machine learning model**. Risk score for a hypothetical myelofibrosis patient candidate for transplantation.

Figure 1



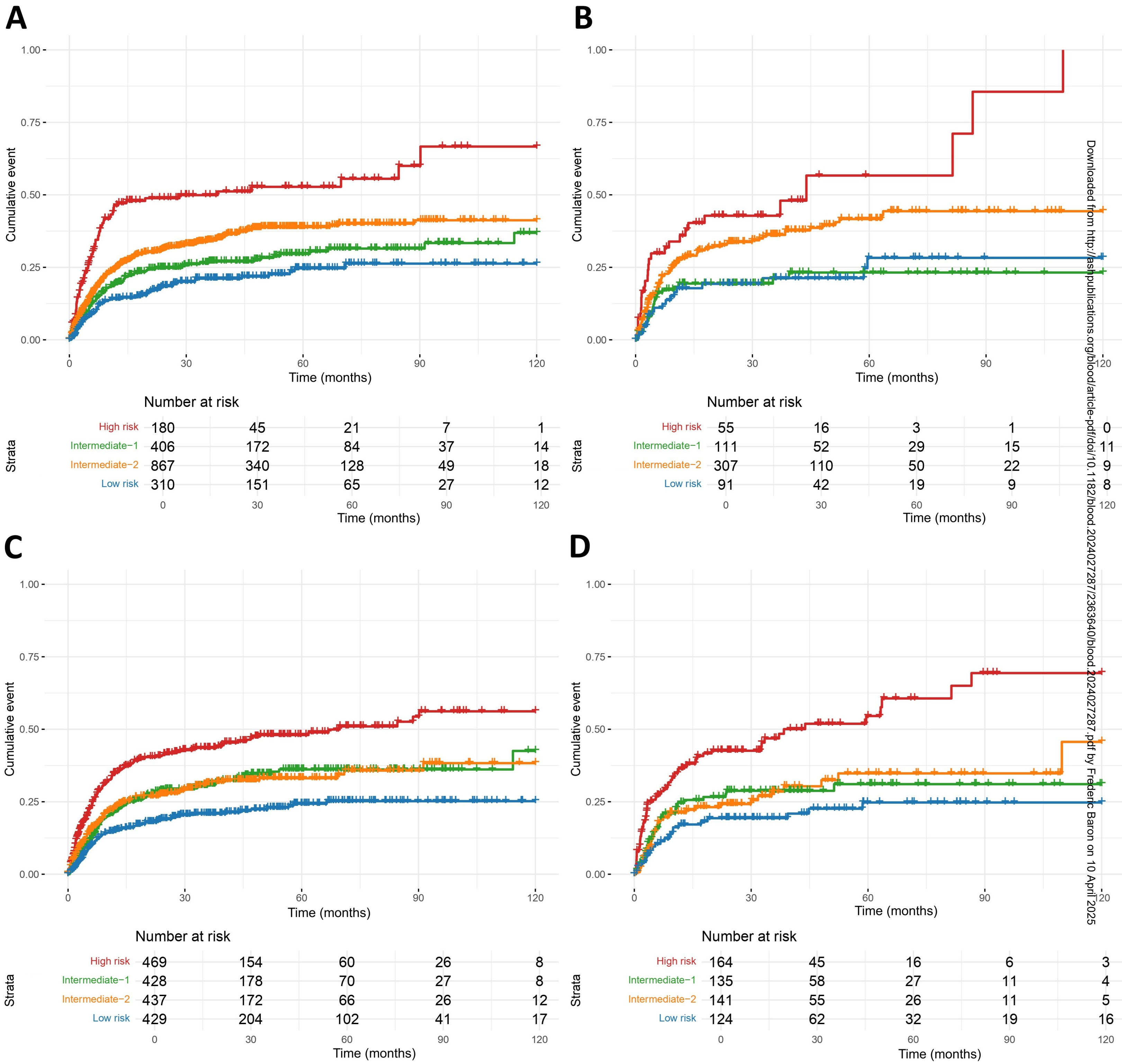
A
B
C
D
Survival probability
1.00
0.75
0.50
0.25
0.00
Time (months)
0
30
60
90
120
Number at risk
Strata
High risk
Intermediate−1
Intermediate−2
Low risk
A: High risk 180 53 25 7 1
A: Intermediate−1 407 205 103 51 21
A: Intermediate−2 874 389 167 69 28
A: Low risk 312 173 78 37 17
B: High risk 55 17 5 1 0
B: Intermediate−1 112 61 34 17 12
B: Intermediate−2 308 137 61 28 12
B: Low risk 91 48 27 14 10
C: High risk 471 174 76 33 11
C: Intermediate−1 430 208 85 35 11
C: Intermediate−2 440 198 84 35 17
C: Low risk 432 240 128 61 28
D: High risk 164 52 22 9 3
D: Intermediate−1 135 66 27 11 4
D: Intermediate−2 142 68 32 14 8
D: Low risk 125 77 46 26 19

Figure 2



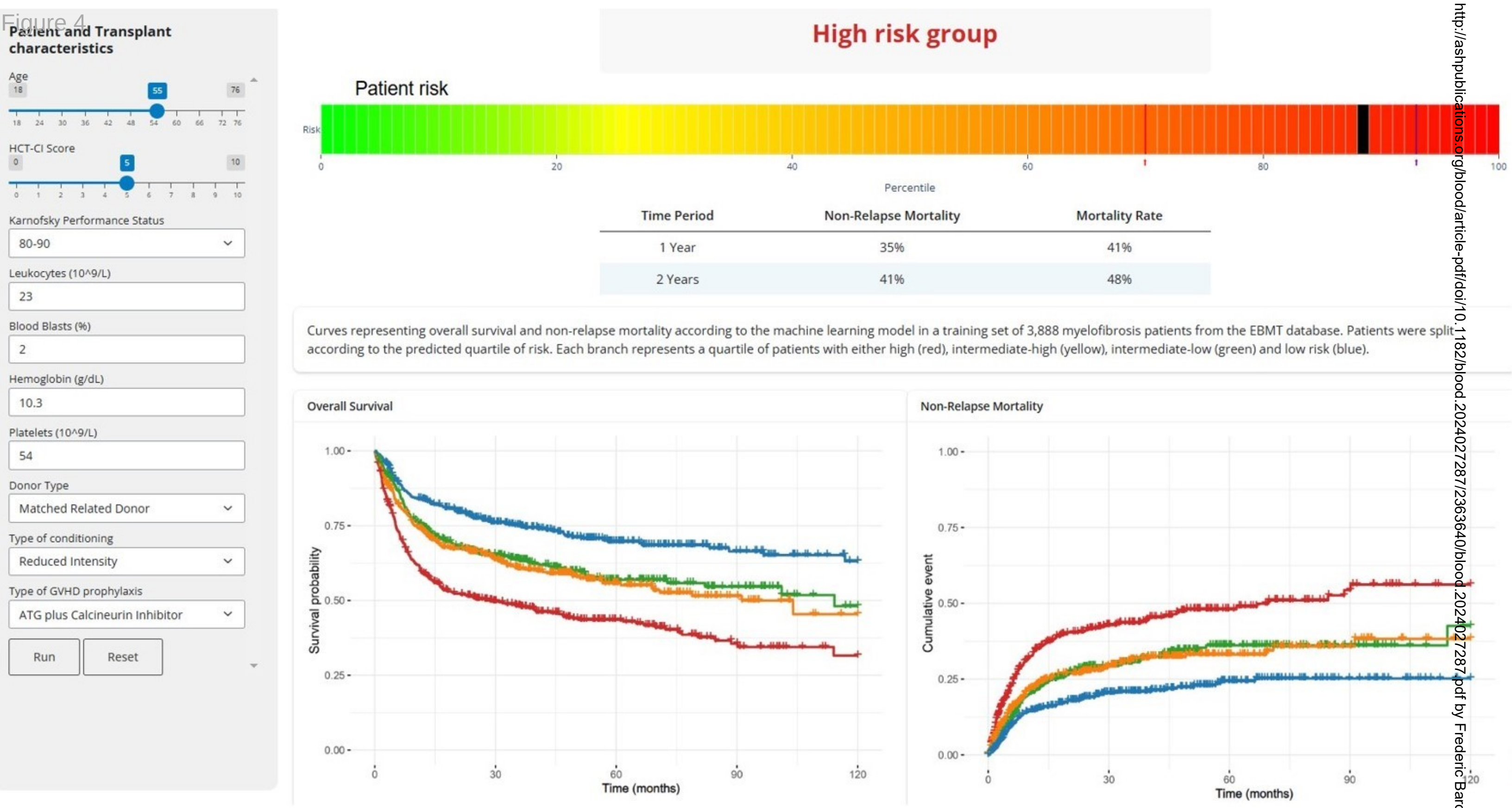
A
Training Set: Risk Group Transitions
Low risk
Intermediate-1
Intermediate-2
High risk
Low risk
Intermediate−1
Intermediate−2
High risk
B
Test Set: Risk Group Transitions
Low risk
Intermediate-1
Intermediate-2
High risk
Low risk
Intermediate−1
Intermediate−2
High risk
C
Training Set: Risk Group Transitions
Low risk
Intermediate-1
Intermediate-2
High risk
Low risk
Intermediate
High risk
D
Test Set: Risk Group Transitions
Low risk
Intermediate-1
Intermediate-2
High risk
Low risk
Intermediate
High risk

Figure 3



A
Cumulative event
1.00
0.75
0.50
0.25
0.00
0
30
60
90
120
Time (months)
Number at risk
Strata
High risk 180 45 21 7 1
Intermediate−1 406 172 84 37 14
Intermediate−2 867 340 128 49 18
Low risk 310 151 65 27 12
0
30
60
90
120
Time (months)
B
Cumulative event
1.00
0.75
0.50
0.25
0.00
0
30
60
90
120
Time (months)
Number at risk
Strata
High risk 55 16 3 1 0
Intermediate−1 111 52 29 15 11
Intermediate−2 307 110 50 22 9
Low risk 91 42 19 9 8
0
30
60
90
120
Time (months)
C
Cumulative event
1.00
0.75
0.50
0.25
0.00
0
30
60
90
120
Time (months)
Number at risk
Strata
High risk 469 154 60 26 8
Intermediate−1 428 178 70 27 8
Intermediate−2 437 172 66 26 12
Low risk 429 204 102 41 17
0
30
60
90
120
Time (months)
D
Cumulative event
1.00
0.75
0.50
0.25
0.00
0
30
60
90
120
Time (months)
Number at risk
Strata
High risk 164 45 16 6 3
Intermediate−1 135 58 27 11 4
Intermediate−2 141 55 26 11 5
Low risk 124 62 32 19 16
0
30
60
90
120
Time (months)

Figure 4


Patient and Transplant characteristics
Age
18
55
76
HCT-CI Score
0
5
10
Karnofsky Performance Status
80-90
Leukocytes (10^9/L)
23
Blood Blasts (%)
2
Hemoglobin (g/dL)
10.3
Platelets (10^9/L)
54
Donor Type
Matched Related Donor
Type of conditioning
Reduced Intensity
Type of GVHD prophylaxis
ATG plus Calcineurin Inhibitor
Run
Reset
High risk group
Patient risk
Risk
Percentile
Time Period
Non-Relapse Mortality
Mortality Rate
1 Year
35%
41%
2 Years
41%
48%
Curves representing overall survival and non-relapse mortality according to the machine learning model in a training set of 3,888 myelofibrosis patients from the EBMT database. Patients were split according to the predicted quartile of risk. Each branch represents a quartile of patients with either high (red), intermediate-high (yellow), intermediate-low (green) and low risk (blue).
Overall Survival
Survival probability
Time (months)
Non-Relapse Mortality
Cumulative event
Time (months)