ARTICLE

Differential impact of CD34+ cell dose for different age groups in allogeneic hematopoietic cell transplantation for acute leukemia: a machine learning–based discovery

Yiyang Qu[a], Hamed Shourabizadeh[a], Aravind Subramanian[a], Dionne M. Aleman[a], Louis-Martin Rousseau[b], Arjun D. Law[c], Auro Viswabandya[c], and Fotios V. Michelis[c]*

[a]*Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada;* [b]*Department of Mathematical and Industrial Engineering, Polytechnique Montreal, Montreal, Quebec, Canada;* [c]*Hans Messner Allogeneic Transplant Program, Division of Medical Oncology and Hematology, Princess Margaret Cancer Centre, University of Toronto, Toronto, Ontario, Canada*

Allogeneic hematopoietic cell transplantation (allo-HCT) presents a potentially curative treatment for hematologic malignancies yet carries associated risks and complications. Continuous research focuses on predicting outcomes and identifying risk factors. Notably, the influence of CD34+ cell dose on overall survival (OS) has been the subject of numerous studies yielding contradictory results. We developed machine learning (ML) models to predict allo-HCT outcomes and, through the application of SHapley Additive exPlanations (SHAP), an explainable artificial intelligence (XAI) technique enabled the identification of new and clinically relevant feature-outcome relationships. In particular, we identified a clear interaction between CD34+ cell dose of peripheral blood stem cell (PBSC) grafts and patient age at allo-HCT for patients with acute leukemia. Results of multivariable analysis validated the interaction effect: in young patients with acute leukemia (aged ≤45 years), low dose of CD34+ cells (<4.3 $\times$ 10^6 CD34+/kg) was associated with better OS against high dose (≥7 $\times 10^6$ CD34+/kg) (hazard ratio [HR], 0.38; $p = 0.019$), while for older patients with acute leukemia (>45 years), low CD34+ cell dose (<3.8 $\times 10^6$ CD34+/kg) was associated with worse OS against high dose (≥6.1 $\times 10^6$ CD34+/kg) (HR, 1.58; $p = 0.033$). In conclusion, our findings suggest that tailoring CD34+ cell dose by patient age may benefit patients with acute leukemia undergoing allo-HCT, while XAI showcases excellent proficiency in revealing such interactions.

HIGHLIGHTS

- We developed machine learning models to predict allo-HCT outcomes and, through the application of SHAP, identified new and clinically relevant feature-outcome relationships.
- In particular, we identified a clear interaction between CD34+ cell dose of peripheral blood stem cell grafts and patient age at allo-HCT for patients with acute leukemia.
- Through the above methodology, we determined that in young patients with acute leukemia (aged ≤45 years), a lower dose of CD34+ cells was associated with better OS, while in older patients with acute leukemia (aged >45 years), a higher cell dose correlated with improved outcomes.
- Our findings suggest that tailoring CD34+ cell dose by patient age may benefit patients with acute leukemia undergoing allo-HCT.

Allogeneic hematopoietic cell transplantation (allo-HCT) is a potentially curative treatment for patients with hematologic diseases. Despite the significant therapeutic benefits, allo-HCT is associated with many risks and complications, such as graft-versus-host disease (GVHD) and relapse. Predicting these risks before transplantation remains a challenge for oncologists, as it covers a wide range of factors that can interact with each other in complex ways. Compared with the conventional Cox proportional hazards (CPH) analysis, machine learning (ML) models exhibit a promising performance that is, at a minimum, on par with the Cox approach, which can be attributed to the ability to capture complex high-dimensional relationships [1–3]. In examining interaction effects, CPH models require a priori specification of feature pairs, whereas ML models inherently identify these interactions during the training process without prior definition. In the field of hematology, more studies have leveraged the power of ML models to predict transplant outcomes using pretransplant factors [4–7].

Address correspondence to Fotios V. Michelis, Department of Medical Oncology and Hematology, Princess Margaret Cancer Centre, 610 University Avenue, Toronto, ON M5G 2M9, Canada; E-mail: Fotios.Michelis@uhn.ca

0301-472X/

https://doi.org/10.1016/j.exphem.2024.104684

However, most ML models possess an inherent "black-box" nature, which limits the interpretability and transparency of their predictive outcomes, posing a challenge for adoption in high-stakes decision-making contexts. To address this challenge, researchers have developed explainable artificial intelligence (XAI) methods that enable the interpretation of ML models. Li et al [8] utilized an XAI approach, SHapley Additive exPlanations (SHAP), to validate, interpret, and visualize nonlinear interactions between clinical variables in prostate cancer survival. Similarly, Moncada-Torres et al [9] demonstrated the superiority of the SHAP framework over traditional Cox regression for predicting breast cancer survival and finding the interaction effects within features.

In the field of hematology, several studies have explored the impact of CD34+ cell dose on allo-HCT outcomes and yielded controversial results [10–17]. Some studies found that a higher CD34+ cell dose is associated with improved overall survival (OS) rates [15], shortened engraftment times [13], and decreased relapse rates [14]. However, other studies demonstrated that a higher CD34+ cell dose is associated with an increase in GVHD along with a decrease in OS [11,12,18]. On the other hand, Yamamoto et al [16] investigated the impact of a low CD34+ cell dose of $1-2 \times 10^6$ cells/kg and found acceptable results. However, to date, no research has conclusively demonstrated an interaction effect between CD34+ cell dose and other patient- and transplant-related variables.

In this study, we applied the SHAP XAI method to our existing ML model [7] for allo-HCT survival prediction to discover a significant interaction effect between CD34+ dose and age, which is validated through conventional statistical survival analysis.

METHODS

Patient Data

The study was retrospective, using data previously collected and stored within the Hans Messner Allogeneic Transplant Program Database of the Princess Margaret Cancer Centre (PMCC), Toronto, Canada. The single-center registry is routinely updated with new patients and follow-up data.

The initial study cohort included patients who underwent allo-HCT for acute myeloid leukemia (AML), acute lymphoblastic leukemia (ALL), lymphoma, myelodysplastic syndromes (MDS), and other hematologic diseases during the period from January 2010 to August 2019, and the last follow-up date of survivors is August 2021. This cohort included 1,153 patients with 253 clinical and laboratory pretransplant and transplant-related variables collected in the database (characteristics outlined in the Supplementary File, Table E1).

Development of ML Models

We employed the same ML model development methodology as in the study by Shourabizadeh et al [7], using an updated, more recent dataset as previously described, including several additional variables and fewer missing data. The primary disparity between the two datasets lies in the time period and patients covered: the dataset of Shourabizadeh et al [7] covered a more extended period from January 1976 to December 2017 and incorporated data from 2,697 patients, which doubled the number of patients of this study's initial cohort. After following the same process of feature selection for the ML model, the dataset consisted of 52 pretransplant features for prediction with a total missing data rate of 22%, compared with 45 features with a total missing data rate of 46% in the previous study by Shourabizadeh et al [7], including the components of the Hematopoietic Cell Transplant Comorbidity Index (HCT-CI) [19] for a subset of patients (Supplementary Table E3). Finally, the same missing data imputation method is used [7].

The Extreme Gradient Boosting model (XGBoost) was chosen for the prediction of 2-year survival outcomes in this study for its high prediction performance and efficiency in computing SHAP values due to its tree structure compared with others. Comprehensive information on the training and testing process, as well as the tuning of the model's hyperparameters, is detailed in the Supplementary File, Supplementary Text E1.

XGBoost is an extension of gradient boosting, characterized by an efficient computational speed and performance, which is used for both regression and classification problems. XGBoost operates by iteratively constructing weak prediction models (typically decision trees) to predict the target variable while minimizing the loss function, which helps to sequentially reduce errors and optimize model performance.

It is important to highlight that our model is trained using data from patients with a variety of diseases. Our goal is to enable the model to identify pertinent associations that extend across all hematologic conditions. Considering the small size of our dataset, we deliberately selected this strategy, encompassing a spectrum of diseases within the training set to maximize the model's learning effectiveness. Details regarding our ML model training and testing are elaborated in the Supplementary File, Supplementary Text E1.

ML Interpretation

The complexity of ML models has generated an upsurge in interest in XAI to augment the transparency and interpretability of black-box models. XAI methods like Anchors [20], local interpretable model-agnostic explanations (LIME) [21], and SHAP [22] have been proposed. We used the game theory-based SHAP framework in our study for model interpretation and feature importance measurement due to its easily understandable visualizations and indication of the direction of variable value impact; that is, SHAP indicates not only which variables are important, but whether high or low values of those variables correspond to survival, providing significant clinical insight, and, most importantly, its capability to investigate interaction effects, which is especially invaluable in complex clinical settings where variables rarely act in isolation. The SHAP methodology derives its foundation from the Shapley values in cooperative game theory. Shapley values allocate a payoff among players in a game depending on their marginal contribution to the total payout. Similarly, SHAP equitably apportions feature contributions and provides consistent importance measures per instance (i.e., dataset record) with the following formula:

$$\Phi_i(f,x) = \sum_{S \subseteq N\{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \left[f_{S \cup \{i\}}(x) - f_S(x) \right]$$

where x is one instance, i is the feature of investigation, N is the set of all features, S is a subset of features excluding i, $|S|$ is the size or cardinality of the set S, $|N|$ is the total number of features, and $f_{\{S \cup \{i\}\}}$ is the prediction function of the model conditioned on the subset of features S along with i. This formula calculates a weighted sum of the differences in the model's output when feature i is included and

when it is not, across all subsets of features. The sum gives the overall contribution of feature i to the prediction for instance x.

SHAP values offer comprehensive global explanations, revealing overall feature importance, identifying trends and biases, and illuminating feature interaction effects. These interaction effects, gauging the combined impact of multiple features on predictions, heighten the interpretability and discovery of high-order interactions within complex ML models, offering a deeper level of understanding. The SHAP interaction value between features i and j is computed as

$$\Phi_{ij}(f,x) = \sum_{S \subseteq N\{i,j\}} \frac{|S|!(|N| - |S| - 2)!}{|N|!} \left[f_{S\cup\{i,j\}}(x) - f_{S\cup\{i\}}(x) - f_{S\cup\{j\}}(x) + f_S(x) \right]$$

This formula extends the SHAP value calculation by considering both i and j in the prediction function and measuring the change in output due to their interaction.

SHAP values are initially computed at the instance level, providing a comprehensive breakdown of each feature's contribution to the prediction for every individual instance. These per-instance SHAP values can then be aggregated to create a higher level of interpretability. By categorizing similar instances and examining their collective SHAP values, it is possible to target specific subgroups for more in-depth analysis. This level of granular scrutiny not only enhances the interpretability of ML models but also improves their transparency.

We utilized the open-source Python package shap (version 0.41.0) to compute both SHAP and interaction values for each patient in our dataset. These computations were based on the final XGBoost model, which was trained using the entire dataset and the optimized hyperparameters. Subsequently, we produced visualizations to represent the overall feature importance, drawing on the SHAP values. To account for disease heterogeneity, we grouped instances according to their respective diagnoses, thus mitigating potential confusion arising from mixed impacts. We created SHAP dependence plots for each pair of features within each diagnosis group. Finally, we examined the most significant interaction effects, as guided by the interaction matrix for features, and discovered novel interaction findings of CD34+ dose and age for patients with acute leukemia from these dependence plots.

Cutoff Values for Age and CD34+ Dose

Two groups of cutoff values are determined for the study: cutoff age and cutoff CD34+ doses for the two age groups. From the SHAP dependence plot (Figure. 1), we focused on the range of cutoff age to 45−50 years. Young and old cohorts associated with each potential cutoff age are further divided into three groups according to their CD34+ doses: low, medium, and high. The CD34+ dose cutoff points are determined using the maximally selected rank statistics method twice [23]: the first time to find the lower cutoff point between low versus medium and high, and the second time to find the upper cutoff point between medium versus high. This method identifies cutoff points that maximize the separation between groups regarding survival outcomes while controlling for multiple testing parameters to maintain statistical validity. Applying it separately to the young and old cohorts allowed us to account for heterogeneity in the impact on outcomes of cell dose between age groups. Biological differences, such as variations in immune reconstitution and comorbidities, may influence how cell dose affects outcomes in younger versus older patients. By determining age-specific cutoff values, we attempted to reflect risk stratification within each cohort more accurately.

Lastly, we determined the appropriate age cutoff by examining the log-rank test results corresponding to all possible ages, derived from the Kaplan-Meier survival curves of both younger and older patient cohorts.

Statistical Methods

We used the Kaplan-Meier estimator to calculate univariate probabilities of OS. Log-rank test was used for univariate survival curve comparisons, although point-wise comparisons were performed using the chi-square test. Probabilities of relapse, nonrelapse mortality (NRM), acute GVHD (aGVHD), and chronic GVHD (cGVHD) were calculated with the cumulative incidence function estimator. Patients were censored at the

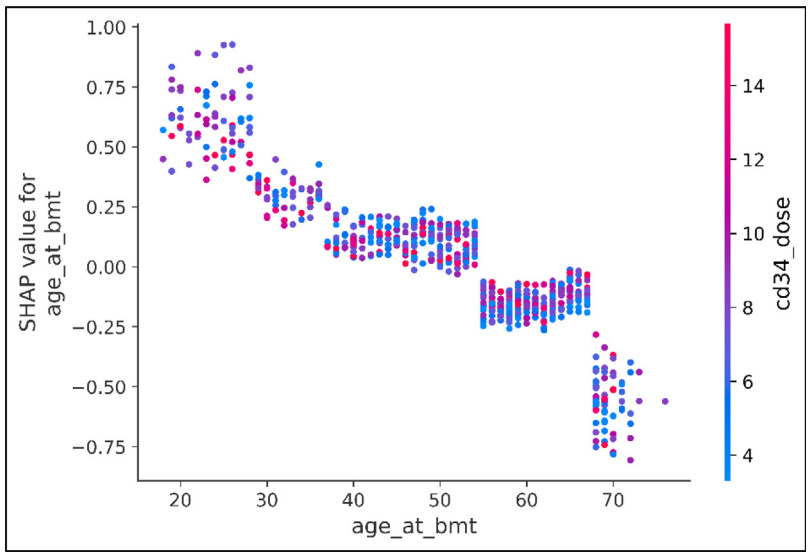


Figure 1 SHAP dependence plot for age and CD34+ for acute leukemia.

time of death or last follow-up. For NRM, relapse was the competing risk; for relapse, NRM was the competing risk. For aGVHD and cGVHD, death was the competing risk. Multivariable analysis of OS was done with CPH regression. Multivariable analyses of relapse and NRM are done using the Fine-Gray proportional hazard model. Results are shown as hazard ratio (HR) along with the 95% confidence interval (CI). Factors included in the regression model are marked by an asterisk in Supplementary Table E3. The chi-square test is used to compare categorical variables. For the cause-of-death data, we applied the chi-square test to compare each cause-of-death category against all other causes within the stratified groups for both age cohorts.

We used a stepwise selection technique with the Akaike information criterion (AIC) as the criteria to evaluate and compare the goodness of fit of the finalized regression models while considering model complexity. *p* values are two-sided.

RESULTS

SHAP Interpretation

The final model trained by the entire cohort of 1,153 patients achieved an area under the receiver operating characteristic curve (AUC) of 0.674 ± 0.012 after 5 × 10-fold repeated cross-validation.

Figure 2 presents the summary plot with the top 15 features based on mean absolute SHAP values from the entire ML model for the entire cohort. Each point in this plot represents a single patient: Its vertical location shows what feature it is depicting, its SHAP value is shown by its position on the x axis, and its specific feature value is indicated by the color of that point according to the color legend on the right. Points with the same SHAP value are clustered along the y axis. In the context of our ML model, a higher SHAP value translates to a higher probability of survival at the 2-year time point. This plot is a high-level overview of the feature importance for our ML model, where patient age (age_at_bmt), human leukocyte antigens (HLAs) mismatch donor, and donor age, which is highly correlated with the donor type feature, are the top three features that the model deems important based on SHAP values. It should be noted that the SHAP value for each feature here combines the main effect from itself and the interaction effects with other features.

Figure 1 is a SHAP dependence plot between recipient age and CD34+ cell dose specifically for the acute leukemia cohort (n = 674). The descending trend of the curve is clear and aligns with both the literature and intuition, suggesting that younger patients tend to have a better chance of survival [19]. New insight into the interaction effect of age and CD34+ dose can be visualized from the trend of colors, representing the level of CD34+ dose for each patient. On the left side of the plot, representing the younger patients, more red points appear on the bottom part of the curve, meaning young patients with higher CD34+ doses have less favorable survival outcomes than those with lower CD34+ doses in blue. However, this pattern switches roughly beyond the age of 45 years, with more red points appearing on the upper part of the curve, meaning more older patients have more favorable survival outcomes with higher CD34+ doses. This interaction effect is particularly noteworthy since physicians have a certain level of control over CD34+ cell dose (depending of course on the success of mobilization and CD34+ cell collection). In contrast to many other immutable factors, this observation about the relationship between recipient age and CD34+ dose allows for personalized adjustments, potentially enhancing survival outcomes.

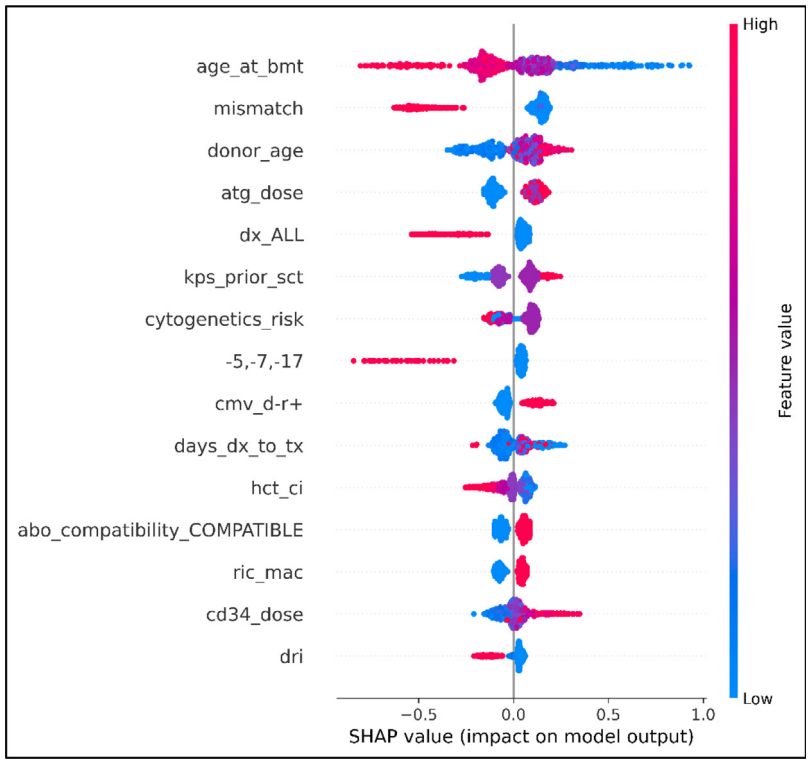


Figure 2 SHAP summary plot.

Characteristics of the Acute Leukemia Cohort

Patient characteristics of the acute leukemia cohort for developing an ML model are displayed in Supplementary Table E3. Features that are included in the multivariable analysis are noted with an asterisk. After finding the interaction effect for the acute leukemia cohort from the XAI investigation, we further analyzed other disease cohorts and found that the interaction effect only occurs within the combined AML and ALL cohort (n = 674). Tables 2 and 3 present the patient characteristics of the young and old acute leukemia cohorts, respectively, with a cutoff age of ≤45 years.

Among the 205 patients in the young group (aged ≤45 years), 26 (12.7%) received a low dose of CD34+ (<4.3 × 10^6 CD34+/kg), 56 (27.3%) received a medium dose of CD34+ (4.3−7 × 10^6 CD34+/kg), and 123 (60.0%) received a high dose of CD34+ (>7 × 10^6 CD34+/kg). The characteristics of patients receiving different doses of CD34+ were similar, except that the group with a high dose had a higher proportion of mismatched donors than the other two groups (Table 1).

Among the 469 patients in the older group (aged >45 years), 44 (9.4%) received a low dose of CD34+ (<3.8 × 10^6 CD34+/kg), 127 (27.1%) received a medium dose of CD34+ (3.8−6.1 × 10^6 CD34+/kg), and 298 (63.5%) received a high dose of CD34+ (>6.1 × 10^6 CD34+/kg). The characteristics of patients receiving different doses of CD34+ were similar, except that donor relationship and disease stage differed among the three groups. The high-dose group reported more haplo donors than the other two groups, and the low-dose group reported a higher proportion of patients at stage complete remission 2 or 3 (CR2/3) (Table 2).

OS

On univariate analysis, within the young cohort, the group with a low dose of CD34+ resulted in the highest 2-year survival: low versus high, 84.6% versus 59.1% ($p = 0.002$), and low versus medium, 84.6% versus 48.2% ($p < 0.001$) (Figure 3A). However, for the old cohort, the reverse is observed. The group with a low dose of CD34+ resulted in the lowest 2-year survival: low versus high, 33.3% versus 55.1% ($p = 0.005$), and low versus medium, 33.3% versus 48.8% ($p = 0.07$) (Figure 3B).

In the multivariable analysis, a lower dose of CD34+ is associated with superior OS in the young cohort (HR, 0.38; 95% CI, 0.17−0.85; $p = 0.019$; Table 3A). For the older cohort, a higher dose of CD34+ is associated with superior OS (HR, 1.58; 95% CI, 1.04−2.40; $p = 0.033$; Table 3B).

Relapse and NRM

On univariate analysis, the cumulative incidence of relapse at 2 years for the three groups of the young cohort is 4.0% (low), 20% (medium), and 18% (high) ($p = 0.2$). However, the low-dose group retained a lower relapse rate compared with medium and high combined (low, 4.0%; medium and high, 19%; $p = 0.059$) because there was only one case of relapse out of 26 patients in the young low-dose group. For the old cohort, the values are 26% (low), 19% (medium), and 17% (high) ($p = 0.6$).

On univariate analysis for NRM at 2 years of the young cohort, the respective values are 16% (low), 33% (medium), and 24% (high) ($p = 0.4$). For the old cohort, the values are 42% (low), 33% (medium), and 30% (high) ($p = 0.3$). Figures of cumulative incidence of relapse and NRM are in the Supplementary File, Supplementary Figures E7−E10.

Similarly, in the multivariable analysis, the low cell dose group is not significantly associated with relapse or NRM for either young or old cohorts.

Engraftment

CD34+ cell dose does not appear to impact neutrophil recovery time within the young cohort, as all three groups (low, medium, and high dose) share the same median neutrophil recovery time (at least 0.5 × 10^9/L) of 14 days, with ranges of 11−24, 11−22, and 10−32 days, respectively. In the older cohort, while the median recovery time is consistent at 16 days across the groups, there is greater variability in the range, with the low-dose group recovering in 11−40 days, the medium-dose group recovering in 9−66 days, and the high-dose group recovering in 10−109 days.

GVHD

Among the 674 patients, 373 patients (55.4%) developed aGVHD at a median of 42 days (range, 6−198 days) after transplantation, and 301 (44.6%) developed grades 2−4 aGVHD. Overall, 248 (36.8%) developed cGVHD at a median of 170 days (range, 32−2,558 days) after transplantation.

CD34+ cell dose did not affect the incidence of aGVHD in both young and old cohorts. In contrast, among the young patients, the dose of CD34+ cells had a significant influence on the development of cGVHD. The 2-year cumulative incidence of cGVHD was highest in patients receiving a low cell dose (58% [95% CI, 36%−74%]), compared with those receiving medium (32% [95% CI, 20%−45%]) or high doses (33% [95% CI, 25%−42%]) of CD34+ cells ($p = 0.045$). In the old patient cohort, the CD34+ cell dose did not significantly influence the incidence of cGVHD. Figures for the cumulative incidence of aGVHD and cGVHD are in the Supplementary File, Supplementary Figures E11−E14.

Cause of Death

Focusing on causes of death (Table 4), distinct patterns emerge when comparing the young and old age cohorts. Although the old cohort demonstrates a pronounced susceptibility to infections/sepsis, with the old-medium cell dose group leading at 34.21%, this difference across cell dose groups is not statistically significant ($p = 0.140$). The young cohorts reveal a distinct pattern concerning GVHD-related deaths. Specifically, there is a marked uptick in GVHD-related deaths with increasing CD34+ doses among the younger demographic. The younger group with the highest dose registers a significant 25.93% of GVHD-related deaths. Additionally, the young low cell dose group stands apart by displaying no deaths due to relapse, whereas the higher dose groups exhibited relapse rates of 38.71% and 35.19%, respectively, though this difference was also not statistically significant ($p = 0.138$). Notably, in the young cohort, there was a statistically significant difference in graft failure between the cell dose groups ($p < 0.001$).

DISCUSSION

In the present study, we successfully utilized the SHAP framework with an ML model we developed to investigate complex interaction

Table 1 Patient characteristic table for young patients with acute leukemia (aged ≤45years) separated by CD34+ cell dose

Variables	Missing	Overall	Low <4.3 × 10^6 CD34+/kg	Medium 4.3–7 × 10^6 CD34+/kg	High >7 × 10^6 CD34+/kg	*p* value
n		205	26	56	123	
Age at transplant (y), mean (SD)	0	32.8 (8.0)	35.9 (7.8)	33.6 (8.0)	31.8 (7.9)	0.038
Sex, n (%)	0					0.541
Female		108 (52.7)	14 (53.8)	26 (46.4)	68 (55.3)	
Male		97 (47.3)	12 (46.2)	30 (53.6)	55 (44.7)	
Donor type, n (%)	0					0.142
Haplo		25 (12.2)	1 (3.8)	3 (5.4)	21 (17.1)	
Related		64 (31.2)	9 (34.6)	19 (33.9)	36 (29.3)	
Unrelated		116 (56.6)	16 (61.5)	34 (60.7)	66 (53.7)	
Diagnosis, n (%)	0					0.438
ALL		70 (34.1)	8 (30.8)	23 (41.1)	39 (31.7)	
AML		135 (65.9)	18 (69.2)	33 (58.9)	84 (68.3)	
ABO compatibility, n (%)	0					0.598
Bidirectional		10 (4.9)		3 (5.4)	7 (5.7)	
Compatible		97 (47.3)	15 (57.7)	29 (51.8)	53 (43.1)	
Major		48 (23.4)	7 (26.9)	12 (21.4)	29 (23.6)	
Minor		50 (24.4)	4 (15.4)	12 (21.4)	34 (27.6)	
TBI dose, n (%)	0					0.468
None		11 (5.4)	1 (3.8)	1 (1.8)	9 (7.3)	
LD		61 (29.8)	6 (23.1)	15 (26.8)	40 (32.5)	
MD		91 (44.4)	15 (57.7)	26 (46.4)	50 (40.7)	
HD		42 (20.5)	4 (15.4)	14 (25.0)	24 (19.5)	
Stage, n (%)	17					0.164
Advanced		6 (3.2)		2 (3.8)	4 (3.6)	
CR1		131 (69.7)	20 (83.3)	30 (57.7)	81 (72.3)	
CR2/3		27 (24.1)	4 (16.7)	51 (27.1)	20 (38.5)	
Conditioning regimen, n (%)	0					0.417
MAC		140 (68.3)	20 (76.9)	40 (71.4)	80 (65.0)	
RIC		65 (31.7)	6 (23.1)	16 (28.6)	43 (35.0)	
HLA mismatch, n (%)	6					0.040
Negative		148 (74.4)	20 (80.0)	47 (85.5)	81 (68.1)	
Positive		51 (25.6)	5 (20.0)	8 (14.5)	38 (31.9)	
Graft, n (%)	0					0.832
Fresh		149 (72.7)	19 (73.1)	39 (69.6)	91 (74.0)	
Frozen		56 (27.3)	7 (26.9)	17 (30.4)	32 (26.0)	
Days Dx to Tx, mean (SD)	1	452.4 (672.4)	403.6 (446.8)	512.6 (815.9)	435.9 (643.6)	0.724
Donor age (y), mean (SD)	0	33.3 (11.3)	34.8 (9.5)	32.0 (12.5)	33.6 (11.2)	0.525
GVHD prophylaxis, n (%)	0					0.632
In vivo T-cell depletion		133 (64.9)	16 (61.5)	34 (60.7)	83 (67.5)	
Other		72 (35.1)	10 (38.5)	22 (39.3)	40 (32.5)	

(*continued*)

Table 1 (*Continued*)

Variables	Missing	Overall	Low <4.3 × 10^6 CD34+/kg	Medium 4.3–7 × 10^6 CD34+/kg	High >7 × 10^6 CD34+/kg	*p* value
CMV pair, n (%)	0					0.789
D+R+		83 (40.5)	13 (50.0)	25 (44.6)	45 (36.6)	
D+R−		19 (9.3)	2 (7.7)	6 (10.7)	11 (8.9)	
D−R+		54 (26.3)	7 (26.9)	13 (23.2)	34 (27.6)	
D−R−		49 (23.9)	4 (15.4)	12 (21.4)	33 (26.8)	
-5,-7,-17 cytogenetics abnormality, n (%)	12					0.762
Negative		182 (94.3)	20 (90.9)	51 (94.4)	111 (94.9)	
Positive		11 (5.7)	2 (9.1)	3 (5.6)	6 (5.1)	
DRI	26					0.546
0		13 (5.5)	7 (5.2)	1 (3.4)	5 (6.7)	
1		189 (79.4)	103 (76.9)	26 (89.7	60 (80.0)	
2		36 (15.1)	24 (17.9)	2 (6.9)	10 (13.3)	

+=Positive; −=negative; *ALL*=acute lymphoblastic leukemia; *AML*=acute lymphoblastic leukemia; *CMV*=cytomegalovirus; *CR1*=first complete remission; *CR2/3*=second or third complete remission; *Dx to Tx*=from diagnosis to transplantation; *D*=donor; *DRI*=disease risk index; *GVHD*=graft-versus-host disease; *HD*=high dose; *HLA*=human leukocyte antigens; *LD*=low dose; *MAC*=myeloablative conditioning; *MD*=medium dose; *R*=recipient; *RIC*=reduced intensity conditioning; *TBI*=total-body irradiation.

relationships between variables characterizing a large single-center allo-HCT cohort. By applying this methodology to our registry containing data from patients with acute leukemia who underwent allo-HCT over the time period of 2010–2019, we have discovered a significant interaction between patient age and the CD34+ cell dose: a lower dose of CD34+ (<4.3 × 10^6 CD34+/kg) is associated with an improved OS for adult patients aged 45 years or younger, whereas a higher dose of CD34+ (≥6.1 × 10^6 CD34+/kg) is associated with better OS for patients older than 45 years.

We showed that using SHAP values to illustrate nonlinear relationships and interactions presents potential benefits over traditional regression methods. A primary advantage is the elimination of the need to predefine pairs of interaction variables. A sophisticated ML model is capable of learning high-dimensional feature interactions during training, allowing SHAP to subsequently extract and elegantly visualize these interactions. In contrast, conventional methods usually necessitate a priori specification based on clinical assumptions, often infeasible without specific domain knowledge about the interaction in question.

The dose of CD34+ cells administered in allo-HCT for hematologic malignancies has been scrutinized in connection with transplant outcomes such as OS, relapse, NRM, and the incidence of GVHD. Yet, different studies present conflicting results. Historical studies from the early 2000s suggested that a higher CD34+ cell dose was associated with unfavorable outcomes. For instance, Zaucha et al [18] noted that doses surpassing 8.0 × 10^6 CD34+/kg were linked to a heightened risk of cGVHD. Similarly, Perez-Simon et al [12] corroborated this finding in a multicenter study. They defined their higher dose group using the 75th percentile as a threshold to account for varying site-specific thresholds. The study reported median doses of 5.68 × 10^6 CD34+/kg and 7 × 10^6 CD34+/kg in the two participating centers. Further supporting this view were findings by Urbano-Ispizua et al [24], identifying over 3 × 10^6 CD34+/kg as a high-dose threshold, and Mohty et al [11], using a threshold of 8.3 × 10^6 CD34+/kg, both associating higher doses with reduced OS and disease-free survival.

However, this perspective began to shift over the following decade, with more recent investigations presenting an opposing view. Pulsipher et al [25] highlighted the advantages of higher CD34+ cell doses over 4.5×10^6 CD34+/kg, noting faster immune recovery and improved OS. This positive association between CD34+ cell dose and OS was further reinforced by Törlén et al [15] and Remberger et al [13] with higher dose thresholds in the range of 4–6.5 × 10^6 CD34+/kg, emphasizing the potential benefits of faster engraftment and reduced risks related to post-transplantation infections.

Over the last two decades, significant advancements in transplant techniques, notably the introduction of reduced intensity conditioning (RIC) and nonmyeloablative (NMA) conditioning regimens, have broadened the demographic spectrum of eligible patients, now covering those in their advanced years, often extending into their 60s or even 70s [26]. Such transformative shifts are evident in the evolving age distributions across various study cohorts. Earlier studies document a median transplant age up to the early 50s [11,12,18,24], while more recent research cites a median age of 56 years and above [13,15]. Pulsipher et al [25] describe median age of 38 (range, 1–65) for patients undergoing myeloablative conditioning (MAC), 56 (1–75) for the RIC cohort, and 57 (17–73) for the NMA cohort.

Table 2 Patient characteristic table for older patients aged (> 45 years) separated using CD34+ cell dose

Variables	Missing	Overall	Low <3.8 × 10^6 CD34+/kg	Medium 3.8–6.1 ×10^6 CD34+/kg	High >6.1 × 10^6 CD34+/kg	*p* value
n		469	44	127	298	
Age at transplant (y), mean (SD)	0	58.6 (7.2)	57.9 (8.0)	59.1 (7.1)	58.5 (7.2)	0.574
Sex, n (%)	0					0.381
Female		220 (46.9)	19 (43.2)	54 (42.5)	147 (49.3)	
Male		249 (53.1)	25 (56.8)	73 (57.5)	151 (50.7)	
Donor type, n (%)	0					0.003
Haplo		40 (8.5)	2 (4.5)	6 (4.7)	32 (10.7)	
Related		149 (31.8)	18 (40.9)	54 (42.5)	77 (25.8)	
Unrelated		280 (59.7)	24 (54.5)	67 (52.8)	189 (63.4)	
Diagnosis, n (%)	0					0.847
ALL		55 (11.7)	4 (9.1)	15 (11.8)	36 (12.1)	
AML		414 (88.3)	40 (90.9)	112 (88.2)	262 (87.9)	
ABO compatibility, n (%)	0					0.775
Bidirectional		32 (6.8)	2 (4.5)	11 (8.7)	19 (6.4)	
Compatible		258 (55.0)	24 (54.5)	70 (55.1)	164 (55.0)	
Major		88 (18.8)	10 (22.7)	26 (20.5)	52 (17.4)	
Minor		91 (19.4)	8 (18.2)	20 (15.7)	63 (21.1)	
TBI dose, n (%)	0					0.325
None		15 (3.2)	1 (2.3)	3 (2.4)	11 (3.7)	
LD		310 (66.1)	24 (54.5)	83 (65.4)	203 (68.1)	
MD		144 (30.7)	19 (43.2)	41 (32.3)	84 (28.2)	
Stage, n (%)	14					0.010
Advanced		5 (1.1)		3 (2.4)	2 (0.7)	
CR1		367 (80.7)	26 (63.4)	100 (79.4)	241 (83.7)	
CR2/3		45 (15.6)	15 (36.6)	83 (18.2)	23 (18.3)	
Conditioning regimen, n (%)	0					0.075
MAC		155 (33.0)	21 (47.7)	43 (33.9)	91 (30.5)	
RIC		314 (67.0)	23 (52.3)	84 (66.1)	207 (69.5)	
HLA mismatch, n (%)	21					0.036
Negative		336 (75.0)	27 (65.9)	102 (82.9)	207 (72.9)	
Positive		112 (25.0)	14 (34.1)	21 (17.1)	77 (27.1)	
Graft, n (%)	0					0.824
Fresh		345 (73.6)	32 (72.7)	91 (71.7)	222 (74.5)	
Frozen		124 (26.4)	12 (27.3)	36 (28.3)	76 (25.5)	
Days Dx to Tx, mean (SD)	1	376.8 (637.0)	377.3 (375.3)	413.1 (801.1)	361.2 (588.0)	0.746
Donor age (y), mean (SD)	1	38.1 (15.4)	43.3 (18.4)	41.0 (16.1)	36.1 (14.3)	0.001
GVHD prophylaxis, n (%)	1					0.117
In vivo T-cell depletion		325 (69.4)	27 (61.4)	82 (64.6)	216 (72.7)	
Other		143 (30.6)	17 (38.6)	45 (35.4)	81 (27.3)	

(*continued*)

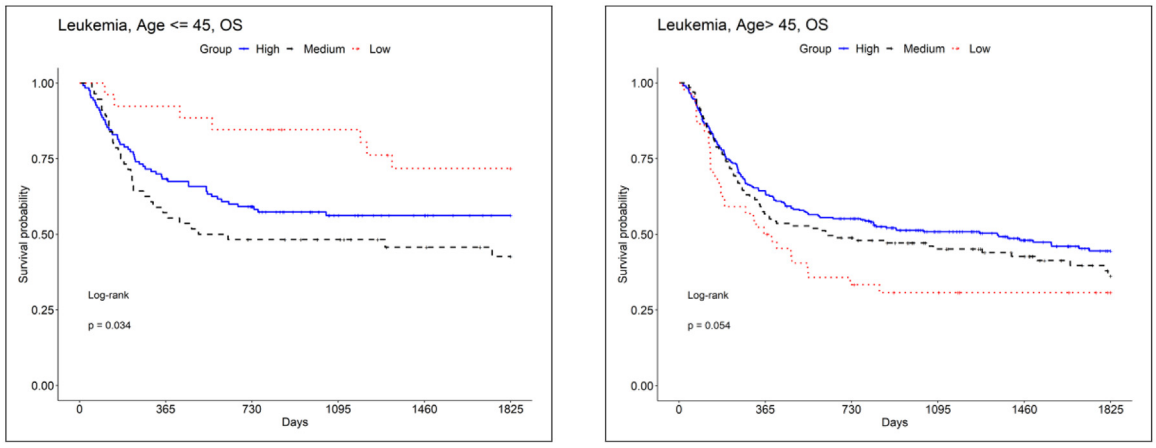
Table 2 (*Continued*)

Variables	Missing	Overall	Low <3.8 × 10^6 CD34+/kg	Medium 3.8–6.1 ×10^6 CD34+/kg	High >6.1 × 10^6 CD34+/kg	*p* value
CMV pair, n (%)	0					0.307
D+R+		211 (45.0)	125 (41.9)	23 (52.3)	63 (49.6)	
D+R−		26 (5.5)	19 (6.4)	0 (0.0)	7 (5.5)	
D−R+		162 (34.5)	110 (36.9)	12 (27.3)	40 (31.5)	
D−R−		70 (14.9)	44 (14.8)	9 (20.5)	17 (13.4)	
-5,-7,-17 cytogenetics abnormality, n (%)	16					0.547
Negative		420 (92.7)	39 (95.1)	115 (94.3)	266 (91.7)	
Positive		33 (7.3)	2 (4.9)	7 (5.7)	24 (8.3)	
DRI	12					0.379
0		22 (3.7)	14 (3.7)	4 (7.5)	4 (2.5)	
1		467 (79.6)	392 (80.1)	39 (73.6)	126 (80.3)	
2		93 (15.8)	56 (14.9)	10 (18.9)	27 (17.2)	
3		5 (0.9)	5 (1.3)			

+=Positive; −=negative; *ALL*=acute lymphoblastic leukemia; *AML*=acute myeloid leukemia; *CMV*=cytomegalovirus; *CR1*=first complete remission; *CR2/3*=second or third complete remission; *Dx to Tx*=from diagnosis to transplantation; *D*=donor; *DRI*=disease risk index; *GVHD*=graft-versus-host disease; *HLA*=human leukocyte antigens; *LD*=low dose; *MAC*=myeloablative conditioning; *MD*=medium dose; *R*=recipient; *RIC*=reduced intensity conditioning; *TBI*=total-body irradiation.

Of note, some earlier studies like Ringdén et al [14] and Singhal et al [27] posited that higher doses of CD34+ cells were correlated with superior OS; however, the threshold delineating low from high doses in the study by Singhal et al [27] was comparatively modest, standing at 2 × 10^6/kg—nearly half of the thresholds used in this study.

The shift toward an older cohort in recent studies may offer insights into the divergent observations regarding the impact of CD34+ cell dose on survival outcomes. Notably, the potential interaction between patient age and CD34+ cell dose remains an uncharted domain. Guided by XAI, our study aspired to reconcile the prevailing controversy surrounding CD34+ cell dose impact by investigating this



(a) Young cohort (b) Old cohort

Figure 3 OS for different age cohorts.

Table 3 Multivariable analysis for OS

Young cohort			
Characteristic	**HR**	**95% CI**	***p* value**
CD34 group			
High (>7 × 10^6 /kg)	—	—	
Medium (4.3–7 × 10^6/kg)	1.17	0.73–1.86	0.5
Low (<4.3 × 10^6/kg)	0.38	0.17–0.85	0.019
Donor age (y)	1.02	1.00–1.04	0.057
Diagnosis			
ALL	—	—	
AML	0.48	0.28–0.85	0.011
Conditioning regimen			
MAC	—	—	
RIC	3.72	1.02–13.6	0.047
Days Dx to Tx	1.00	1.00–1.00	0.011
BMT year	0.88	0.80–0.98	0.019
CMV pair			
D−R−	—	—	
D−R+	2.19	1.14–4.22	0.019
D+R−	0.95	0.39–2.30	> 0.9
D+R+	1.71	0.92–3.18	0.092
TBI dose			
HD	—	—	
LD	0.50	0.12–2.04	0.3
MD	1.97	0.96–4.04	0.064
None	1.56	0.39–6.34	0.5
Old cohort			
Characteristic	**HR**	**95% CI**	***p* value**
CD34 group			
High (>6.1 × 10^6/kg)	—	—	
Medium (3.8–6.1 × 10^6/kg)	1.24	0.93–1.64	0.14
Low (<3.8 × 10^6/kg)	1.58	1.04–2.40	0.033
Donor type			
Haplo	—	—	
Related	0.58	0.34–0.98	0.044
Unrelated	0.89	0.54–1.46	0.7
Diagnosis			
ALL	—	—	
AML	0.70	0.48–1.04	0.075
BMT year	0.92	0.88–0.96	< 0.001
Stage Advanced	—	—	
CR1	0.22	0.08–0.58	0.002
CR2/3	0.19	0.07–0.53	0.001

+=Positive; −=negative; *ALL*=acute lymphoblastic leukemia; *AML*=acute myeloid leukemia; *BMT*=bone marrow transplantation; CI=confidence interval; *CMV*=cytomegalovirus; *CR1*=first complete remission; *CR2/3*=second or third complete remission; *D*=donor; *Dx to Tx*=from diagnosis to transplantation; *HD*=high dose; *HR*=hazard ratio; *LD*=low dose; *MAC*=myeloablative conditioning; *MD*=medium dose; *R*=recipient; *RIC*=reduced intensity conditioning; *TBI*=total-body irradiation.

interaction. The detrimental effect of a higher cell dose in younger patients may be associated with more robust engraftment in these younger patients with a subsequent increase in the incidence of aGVHD. On the other hand, in older patients who seem to demonstrate graft failure as a more significant cause of death (Table 4), it may be that the higher cell dose associated with improved outcomes seems to overcome the increased graft failure risk these patients have, without significantly increasing GVHD severity and risk.

At this point, it is important that we raise awareness regarding responsible artificial intelligence (AI). Given the inherent limitations of single-center registry data, often characterized by its modest size, we trained our model on a diverse array of diseases to capture universal patterns, resulting in optimal performance. However, we must acknowledge an imbalanced disease distribution, with acute leukemia as the predominant class. Consequently, when utilizing SHAP across the whole cohort without considering disease heterogeneity, interaction values from this dominant class could disproportionately influence results, potentially creating a misconception that these findings are universal. It underscores the necessity of practicing AI with caution and responsibility to ensure that insights derived from SHAP are interpreted both accurately and meaningfully.

Our study demonstrates additional limitations. First, it is retrospective and registry-based from a single-center cohort with a limited number of patients, which may impact the generalizability of our findings. Because transplantation protocols and treatment management can vary significantly among different centers, it would be ideal to validate these interactions using multicenter data for further examination. Second, SHAP explanations are not causal; the underlying mechanisms causing differences in OS remain unclear. No statistically significant links were observed between the CD34+ cell dose and either relapse or NRM. However, there was a noticeable trend suggesting a reduced relapse rate in the young low-dose group. Further research is warranted for the validation of our findings in a separate cohort as well as further investigation of the biological mechanisms that drive the interaction between CD34+ cell dose, patient age, and allo-HCT outcomes.

We conclude that by applying a novel XAI method with our ML model, we were able to uncover an interaction between CD34+ cell dose and patient age for patients with acute leukemia undergoing allo-HCT. Patients with different ages at transplant may benefit from a tailored CD34+ cell dose, potentially leading to improved survival.

Table 4 Distribution of causes of death across groups

Cause	Old_low	Old_medium	Old_high	*p* value	Young_low	Young_medium	Young_high	*p* value
GVHD-related	6 (20.0%)	9 (11.84%)	32 (19.63%)	0.312	1 (14.29%)	6 (19.35%)	14 (25.93%)	0.672
Graft failure	2 (6.67%)	1 (1.32%)	1 (0.61%)	0.042	2 (28.57%)	0 (0.0%)	0 (0.0%)	<0.001
Infections/sepsis	5 (16.67%)	26 (34.21%)	41 (25.15%)	0.140	2 (28.57%)	8 (25.81%)	9 (16.67%)	0.526
Organ failure	3 (10.0%)	6 (7.89%)	15 (9.2%)	0.924	1 (14.29%)	1 (3.23%)	5 (9.26%)	0.472
Relapse	11 (36.67%)	25 (32.89%)	53 (32.52%)	0.905	0 (0.0%)	12 (38.71%)	19 (35.19%)	0.138
Other	3 (10.0%)	9 (11.84%)	21 (12.88%)	0.899	1 (14.29%)	4 (12.9%)	7 (12.96%)	0.994

GVHD=Graft-versus-host disease.

Conflict of Interest Disclosure

The authors do not have any conflicts of interest to declare in relation to this work.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.exphem.2024.104684.

REFERENCES

1. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol 2018;18:24.
2. Nicolò C, Périer C, Prague M, et al. Machine learning and mechanistic modeling for prediction of metastatic relapse in early-stage breast cancer. JCO Clin Cancer Inform 2020;4:259–74.
3. Omurlu IK, Ture M, Tokatli F. The comparisons of random survival forests and Cox regression analysis with simulation and an application related to breast cancer. Expert Syst Appl 2009;36:8582–8.
4. Choi EJ, Jun TJ, Park HS, et al. Predicting long-term survival after allogeneic hematopoietic cell transplantation in patients with hematologic malignancies: machine learning-based model development and validation. JMIR Med Inform 2022;10:e32313.
5. Eisenberg L, consortium XplOit, Brossette C, et al. Time-dependent prediction of mortality and cytomegalovirus reactivation after allogeneic hematopoietic cell transplantation using machine learning. Am J Hematol 2022;97:1309–23.
6. Shouval R, Labopin M, Bondi O, et al. Prediction of allogeneic hematopoietic stem-cell transplantation mortality 100 days after transplantation using a machine learning algorithm: a European group for blood and marrow transplantation acute leukemia working party retrospective data mining study. J Clin Oncol 2015;33:3144–51.
7. Shourabizadeh H, Aleman DM, Rousseau LM, Law AD, Viswabandya A, Michelis FV. Machine learning for the prediction of survival post-allogeneic hematopoietic cell transplantation: a single-center experience. Acta Haematol 2024;147:280–91.
8. Li R, Shinde A, Liu A, et al. Machine learning-based interpretation and visualization of nonlinear interactions in prostate cancer survival. JCO Clin Cancer Inform 2020;4:637–46.
9. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Sci Rep 2021;11:6968.
10. Maie K, Fuji S, Tajima K, et al. A higher number of infused CD34^{+} cells has a positive impact on the clinical outcome after related PBSC transplantation. Bone Marrow Transplant 2014;49:1113–5.
11. Mohty M, Bilger K, Jourdan E, et al. Higher doses of CD34+ peripheral blood stem cells are associated with increased mortality from chronic graft-versus-host disease after allogeneic HLA-identical sibling transplantation. Leukemia 2003;17:869–75.
12. Perez-Simon JA, Diez-Campelo M, Martino R, et al. Impact of CD34+ cell dose on the outcome of patients undergoing reduced-intensity-conditioning allogeneic peripheral blood stem cell transplantation. Blood 2003;102:1108–13.
13. Remberger M, Grønvold B, Ali M, et al. The CD34+ cell dose matters in hematopoietic stem cell transplantation with peripheral blood stem cells from sibling donors. Clin Hematol Int 2020;2:74–81.
14. Ringdén O, Barrett AJ, Zhang MJ, et al. Decreased treatment failure in recipients of HLA-identical bone marrow or peripheral blood stem cell transplants with high CD34 cell doses. Br J Haematol 2003;121:874–85.
15. Törlén J, Ringdén O, Le Rademacher J, et al. Low CD34 dose is associated with poor survival after reduced-intensity conditioning allogeneic transplantation for acute myeloid leukemia and myelodysplastic syndrome. Biol Blood Marrow Transplant 2014;20:1418–25.
16. Yamamoto C, Ogawa H, Fukuda T, et al. Impact of a low CD34+ cell dose on allogeneic peripheral blood stem cell transplantation. Biol Blood Marrow Transplant 2018;24:708–16.
17. Yokoyama Y, Maie K, Fukuda T, et al. A high CD34+ cell dose is associated with better disease-free survival in patients with low-risk diseases undergoing peripheral blood stem cell transplantation from HLA-matched related donors. Bone Marrow Transplant 2020;55:1726–35.
18. Zaucha JM, Gooley T, Bensinger WI, et al. CD34 cell dose in granulocyte colony-stimulating factor–mobilized peripheral blood mononuclear cell grafts affects engraftment kinetics and development of extensive chronic graft-versus-host disease after human leukocyte antigen–identical sibling transplantation. Blood 2001;98:3221–7.
19. Sorror ML, Maris MB, Storb R, et al. Hematopoietic cell transplantation (HCT)-specific comorbidity index: a new tool for risk assessment before allogeneic HCT. Blood 2005;106:2912–9.
20. Ribeiro MT, Singh S, Guestrin C. Anchors: high-precision model-agnostic explanations. In: In: Proceedings of the AAAI Conference on Artificial Intelligence; 2018. Available at: https://ojs.aaai.org/index.php/AAAI/article/view/11491 Accessed January 3, 2023.
21. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" Explaining the predictions of any classifier. HLT-NAACL Demos; 2016. p. 1135–44.
22. Lundberg S, Lee SI. A unified approach to interpreting model predictions. In: In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Red Hook, NY: Curran Associates Inc; 2017:4768–77.
23. Lausen B, Schumacher M. Maximally selected rank statistics. Biometrics 1992;48:73–85.

24. Urbano-Ispizua A, Carreras E, Marın P, et al. Allogeneic transplantation of $CD34^+$ selected cells from peripheral blood from human leukocyte antigen–identical siblings: detrimental effect of a high number of donor $CD34^+$ cells? Blood 2001;98:2352–7.
25. Pulsipher MA, Chitphakdithai P, Logan BR, et al. Donor, recipient, and transplant characteristics as risk factors after unrelated donor PBSC transplantation: beneficial effects of higher CD34+ cell dose. Blood 2009;114:2606–16.
26. McClune BL, Weisdorf DJ, Pedersen TL, et al. Effect of age on outcome of reduced-intensity hematopoietic cell transplantation for older patients with acute myeloid leukemia in first complete remission or with myelodysplastic syndrome. J Clin Oncol 2010;28:1878–87.
27. Singhal S, Powles R, Treleaven J, et al. A low CD34+ cell dose results in higher mortality and poorer survival after blood or marrow stem cell transplantation from HLA-identical siblings: should 2×10^6 CD34+ cells/kg be considered the minimum threshold? Bone Marrow Transplant 2000;26:489–96.