ARTICLE

Time-dependent effects of clinical predictors in unrelated hematopoietic stem cell transplantation

Daniel Fuerst,[1,2] Carlheinz Mueller,[3,4] Dietrich W Beelen,[5,6] Christine Neuchel,[1,2] Chrysanthi Tsamadou,[1,2] Hubert Schrezenmeier,[1,2] and Joannis Mytilineos[1,2,3]

[1]Institute of Clinical Transfusion Medicine and Immunogenetics Ulm, German, Red Cross Blood Transfusion Service, Baden-Wuerttemberg – Hessen, Ulm; [2]Institute of Transfusion Medicine, University of Ulm; [3]DRST – German Registry for Stem Cell Transplantation, Ulm Office; [4]Zentrales Knochenmarkspender-Register Deutschland (ZKRD), Ulm; [5]DRST – German Registry for Stem Cell Transplantation, Essen Office; and [6]Department of Bone Marrow Transplantation, University Hospital, University of Duisburg-Essen, Essen, Germany

Haematologica 2016
Volume 101(2):241-247

ABSTRACT

Hematopoietic stem cell transplantation is a multifactorial process. Some of the predictors exhibit time-dependent effects. We present a systematic analysis and description of selected clinical predictors influencing outcome in a time-dependent manner based on an analysis of registry data from the German Registry for Stem Cell Transplantation. A total of 14,951 patients with acute myeloid leukemia, acute lymphocytic leukemia, myelodysplastic syndrome and non-Hodgkin lymphoma transplanted with peripheral blood stem cells or bone marrow grafts were included. Multivariate Cox regression models were tested for time-dependent effects within each diagnosis group. Predictors not satisfying the proportional hazards assumption were modeled in a time-dependent manner, extending the Cox regression models. Similar patterns occurred in all diagnosis groups. Patients with a poor Karnofsky performance score (<80) had a high risk for early mortality until day 139 following transplantation (HR 2.42, CI: 2.19-2.68; P<0.001) compared to patients with a good Karnofsky performance score (80-100). Afterwards the risk reduced to HR 1.43, CI: 1.25-1.63; P<0.001. A lower mortality risk was found for patients after conditioning treatment with reduced intensity until day 120 post transplant (HR: 0.81 CI: 0.75-0.88; P<0.001). After this, a slightly higher risk could be shown for these patients. Similarly, patients who had received a PBSC graft exhibited a significantly lower mortality risk until day 388 post transplantation (HR 0.79, CI: 0.73-0.85; P<0.001), reversing to a significantly higher risk afterwards (HR 1.23, CI: 1.08-1.40; P=0.002). Integrating time dependency in regression models allows a more accurate description and quantification of clinical predictors to be made, which may help in risk assessment and patient counseling.

Introduction

Outcome after hematopoietic stem cell transplantation is influenced by different factors. These include disease stage[1] or cytogenetic risk,[2] but also pre-transplant treatment-related variables, such as conditioning toxicity and selection of therapeutic agents.[3] Furthermore, donor or graft properties may affect outcome.[4] Of these, for example HLA-matching, cytomegalovirus (CMV) status or graft source are usually considered, as they were found to impact transplant-related outcome.[5] Finally,

Correspondence:
d.fuerst@blutspende.de

Received: 11/05/2015.
Accepted: 25/11/2015.
Pre-published: 26/11/2015.

doi:10.3324/haematol.2015.130401

post-transplant treatment in the form of graft-*versus*-host disease (GvHD) prophylaxis and therapy with growth factors or steroids may influence morbidity and mortality.[6] Predictors may be categorized as modifiable, such as type and dosage of therapeutic agent, or as invariable, e.g. genetic factors.

In survival analysis, different groups are compared and effects may vary in size, also in a time-dependent manner, as the outcomes are time-to-event end points. Some effects may change in intensity over time or may be present only for a limited period, e.g. therapeutic interventions. From a clinical perspective, intensity of conditioning treatment is likely to show time-dependent effects as toxicity resolves with time. In addition, graft source may show time dependency as the kinetics of immune reconstitution is quite different in bone marrow (BM) and in peripheral blood stem cell (PBSC) grafts.[7] Karnofsky performance score (KPS) has been shown to associate with a high early mortality.[8,9] Finally, disease stage might be of interest, as relapse-associated and transplantation-associated events may have a greater impact on advancing disease stage early after transplantation. One important tool to evaluate survival is Kaplan-Meier analysis,[10] in which time-dependent effects may manifest as crossing or diverging/converging survival curves (*Online Supplementary Figures S1 and S2*). However, Kaplan-Meier analysis is a univariate method and does not allow examination of multiple effects in combination. For multivariate analysis, the standard methodology is Cox regression analysis,[11] which is limited by the "Proportional Hazards Assumption" (PHA). This means that all effects in the model are assumed to remain constant over time. This is often not the case, and would lead to false regression estimates, and, therefore, false hazard ratios for such effects if ignored. In such situations, the Cox regression model has been extended to allow for adjustment of time-dependent effects.[12] To visualize such effects over time, dynamic regression modeling has been proposed.[13] Models including time-dependent variables have been used in important clinical studies.[14,15] However, such time-dependent effects have not been studied in a dedicated analysis of HSCT data before, as their interpretation is not as intuitive as that of relative risk estimates obtained from variables fulfilling the PHA. In this study, we investigated selected clinical predictors of unrelated stem cell transplantation (HSCT) for time-varying effects and aimed to identify, describe and quantify these effects in such a way as to facilitate their interpretation by the clinician. For this reason, we analyzed a large cohort of patients based on data from the German Registry for Stem Cell Transplantation (DRST). Patterns that were observed in a similar fashion across the different disease entities are reported.

Methods

Patients

The DRST database is a subset of the European Group for Blood and Marrow Transplantation (EBMT) ProMISe database and includes patients transplanted in Germany. Clinical data of all patients receiving a first allogeneic transplant for the disease entities acute myeloid leukemia (AML), acute lymphocytic leukemia (ALL), myelodysplastic syndrome (MDS) and lymphoma were retrieved. The category lymphoma included chronic lymphocytic leukemia (CLL) and was sub-classified in aggressive and indolent. No missing values were accepted for disease stage, HLA-match type, and source of stem cells. Only adult patients and patients transplanted with BM or PBSC were included in this analysis. A total of 14,951 patients were eligible with a median age of 48 years (range 18-78). Transplantations were performed between 1976 and 2013. Median follow up of surviving patients was 49 months. Patients' characteristics are summarized in Table 1 and *Online Supplementary Table S1.*

Table 1. Patients' characteristics.

Category, n (%)		AML	ALL	MDS	Lym aggressive	Lym indolent	Total
Karnofsky Performance Score	80-100	5702 (79.9)	2095 (77.7)	2019 (84.8)	1018 (85.0)	1410 (91.3)	12,244 (81.9)
	<80	455 (6.4)	179 (6.6)	179 (7.5)	111 (9.3)	94 (6.1)	1018 (6.8)
	Missing data	976 (13.7)	422 (15.7)	182 (7.6)	68 (5.7)	41 (2.7)	1689 (11.3)
Conditioning	MAC	4654 (65.2)	2416 (89.6)	1156 (48.6)	776 (64.8)	682 (44.1)	9684 (64.8)
	RIC	2479 (34.8)	280 (10.4)	1224 (51.4)	421 (35.2)	863 (55.9)	5267 (35.2)
Graft source	PBSC	5958 (83.5)	2055 (76.2)	2099 (88.2)	1092 (91.2)	1444 (93.5)	12,648 (84.6)
	BM	1175 (16.5)	641 (23.8)	281 (11.8)	105 (8.8)	101 (6.5)	2303 (15.4)
Patient age	Median	48	36	57	47	55	48
	Range	18-77	18-74	18-78	18-75	18-76	18-78
Patient sex	Female	3477 (48.7)	1022 (37.9)	955 (40.1)	420 (35.1)	441 (28.5)	6315 (42.2)
	Male	3652 (51.2)	1670 (61.9)	1422 (59.7)	775 (64.7)	1101 (71.3)	8620 (57.7)
	Unknown	4 (0.1)	4 (0.1)	3 (0.1)	2 (0.2)	3 (0.2)	16 (0.1)
Donor type	mREL	2986 (41.9)	1028 (38.1)	782 (32.9)	468 (39.1)	531 (34.4)	5795 (38.8)
	mUNREL	2547 (35.7)	1072 (39.8)	1030 (43.3)	449 (37.5)	674 (43.6)	5772 (38.6)
	mmREL	378 (5.3)	125 (4.6)	67 (2.8)	62 (5.2)	44 (2.8)	676 (4.5)
	mmUNREL	1222 (17.1)	471 (17.5)	501 (21.1)	218 (18.2)	296 (19.2)	2708 (18.1)
Year of Tx	1976-2000	1472 (20.6)	712 (26.4)	311 (13.1)	140 (11.7)	92 (6.0)	2727 (18.2)
	2001-2005	1774 (24.9)	782 (29.0)	499 (21.0)	274 (22.9)	384 (24.9)	3713 (24.8)
	2006-2013	3887 (54.5)	1202 (44.6)	1570 (66.0)	783 (65.4)	1069 (69.2)	8511 (56.9)

AML: acute myeloid leukemia; ALL: acute lymphoblastic leukemia; Lym: lymphoma; PBSC: peripheral blood stem cells; BM: bone marrow; MAC: myeloablative conditioning; RIC: reduced intensity conditioning; mREL: matched related; mUNREL: matched unrelated; mmREL: mismatched related; mmUNREL: mismatched unrelated; Tx: transplantation.

Definitions

The term "time dependent" is used for covariates not satisfying the PHA, which may be caused by an effect that changes in intensity over time or an effect that is present only for a limited time period and modifies subsequent risk. Disease stage definitions were adopted from the EBMT study group defining the EBMT risk score.[14] Early disease stage was defined as transplantation in first complete remission for acute leukemia and as untreated or in first complete remission for MDS and lymphoma. Intermediate disease stage grouped for acute leukemia transplantation in second complete remission, for MDS in second complete or partial remission, and for lymphoma in second complete remission, partial remission or stable disease. Stages other than early or intermediate were classified as advanced disease stage.

Regarding conditioning regimen intensity, the terms myeloablative (MAC), non-myeloablative (NMA), and reduced intensity conditioning (RIC) have been introduced.[16] However, the EBMT ProMISe format currently classifies non-myeloablative and reduced intensity together as RIC. From retrospective data, it is sometimes difficult to distinguish between NMA and RIC. Therefore, our study only distinguished between myeloablative and conditioning with less intensity, termed RIC. According to the EBMT standards, MAC conditioning is defined as total body irradiation (TBI) of 10 Gy or more combined with cyclophosphamide or etoposide, or busulfan 16 mg/kg combined with cyclophosphamide 120-200 mg/kg. For lymphomas, also BEAM and CBV polychemotherapy were considered as MAC conditioning. Regimens with lower dosages were considered as RIC (see EBMT MED-AB forms manual Appendix III; *www.ebmt.org*).

Overall survival (OS) was defined as the fraction of surviving patients at any given time point after transplantation. Death from any cause was considered as an event. Patients alive at the last follow up were censored.

Disease-free survival (DFS) was defined as the proportion of patients alive without evidence of disease at any given time point after transplantation. Death from any cause or recurrence of diseases, whichever occurred earlier, was considered as an event. Patients alive and free from disease at the last follow up were censored. Consent for scientific data analysis was obtained upon registration in the DRST. The study was approved by the ethical committee of the University of Ulm, Germany (n. 108/15).

Statistical analysis

The Cox regression model is a multiplicative hazard model, which means that the effect of covariates is modeled on a multiplicative scale in relationship to the underlying base-line hazard. While the effects of covariates are assumed to be proportional over time, no assumptions are made on the structure of the base-line hazard (the distribution of survival times). In contrast, the Cox-Aalen model includes a multiplicative as well as an additive component. The multiplicative part incorporates covariates that modify the excess risk similarly to the Cox model, while the additive part models the base-line hazard rate allowing for time-varying effects.[17,18] For OS and DFS multivariate Cox regression models were fitted and PHA was examined by a test based on weighted Scheonfeld residuals according to the algorithm proposed by Grambsch and Therneau implemented in the "survival package" of the R statistical software (R-3.0.2).[19] The test basically checks for linearity of Schoenfeld-residuals over time. Covariates not satisfying the PHA were further examined by fitting a Cox-Aalen model and plotting cumulative hazard curves.[20] These curves show the change in hazard over time. A positive slope adds risk and therefore correlates to increased hazard rates (i.e. HR>1), a negative slope corresponds to reduced hazard rates (i.e. HR<1), and a slope 0 (a horizontal line) implies no effect on outcome. In time periods where the slope is constant or, in other words, a straight line, the PHA is satisfied for these time periods. Based on this information, it is possible to extend the Cox model by splitting the follow up for time-dependent covariates into

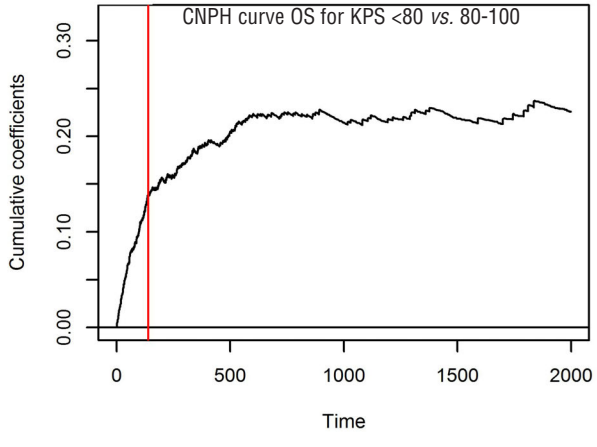

Figure 1. Cumulative non-parametric hazard (CNPH) curve, overall survival (OS) for Karnofsky Performance Score (KPS) less than 80 compared to KPS 80-100. The vertical red line marks day 139 after transplantation.

Table 2. Overall survival.

	Time period	HR	CI	*P*	PHA-test *P*
Karnofsky Performance Score <80 *vs.* 80-100	Before d139	2.42	2.19-2.68	<0.001	0.255
	After d139	1.43	1.25-1.63	<0.001	0.794
	Unadjusted	2.01	1.85-2.18	<0.001	<0.001
Reduced intensity conditioning *vs.* myeloablative conditioning	Before d120	0.81	0.75-0.88	<0.001	0.626
	After d120	1.11	1.04-1.18	0.003	0.310
	Unadjusted	0.97	0.92-1.02	0.220	<0.001
PBSC *vs.* BM	Before d388	0.79	0.73-0.85	<0.001	0.161
	After d388	1.23	1.08-1.40	0.002	0.436
	Unadjusted	0.94	0.87-1.01	0.069	<0.001

HR: hazard ratio; CI: confidence interval; PHA: proportional hazards assumption; PBSC: peripheral blood stem cells; BM: bone marrow.

Table 3. Disease-free survival.

	Time period	HR	CI	*P*	PHA-test *P*
Karnofsky Performance Score <80 *vs.* 80-100	Before d130	2.26	2.05-2.49	<0.001	0.071
	After d130	1.30	1.13-1.50	<0.001	0.910
	Unadjusted	1.91	1.76-2.07	<0.001	<0.001
Reduced intensity conditioning *vs.* myeloablative conditioning	Before d140	0.88	0.82-0.95	0.001	0.835
	After d140	1.12	1.04-1.21	0.002	0.991
	Unadjusted	0.98	0.93-1.03	0.458	<0.001
PBSC *vs.* BM	Before d242	0.86	0.79-0.93	<0.001	0.994
	After d242	1.14	1.01-1.29	0.039	0.317
	Unadjusted	0.97	0.90-1.04	0.382	0.001
Intermediate disease stage *vs.* early disease stage	Before d309	1.82	1.69-1.95	<0.001	0.231
	After d309	1.52	1.36-1.70	<0.001	0.748
	Unadjusted	1.73	1.62-1.84	<0.001	0.004
Advanced disease stage *vs.* early disease stage	Before d242	2.43	2.27-2.60	<0.001	0.214
	After d242	1.79	1.63-1.98	<0.001	0.473
	Unadjusted	2.23	2.10-2.36	<0.001	<0.001

HR: hazard ratio; CI: confidence interval; PHA: proportional hazards assumption; PBSC: peripheral blood stem cells; BM: bone marrow.

observation periods and to estimate regression coefficients separately for these respective time intervals. Cut-off points for these time intervals were determined by fitting models from ranges of cut-off points. Selection of cut-off points was based on optimal log likelihood.

Multivariate models included risk factors defined by the EBMT risk score: age, disease stage, time from diagnosis to transplant, donor type, and recipient-donor sex combination. In addition, year of transplantation, graft source and conditioning therapy and KPS (<80=poor *vs.* 80-100=good) were included. Missing data for KPS (11.3%) were included in the models as separate category. Stratification was performed for diagnosis and donor-type; a center effect was adjusted using a gamma frailty term. Final models were tested for PHA and after adjustment for time-dependent effects all models satisfied PHA. P=0.01 was considered statistically significant.

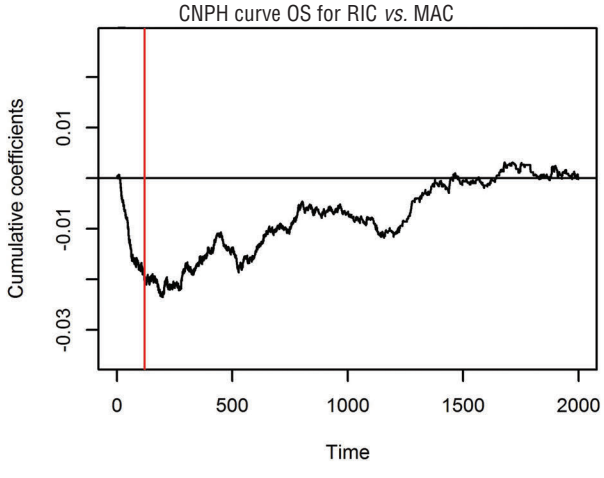

Figure 2. Cumulative non-parametric hazard curve (CNPH) curve, overall survival (OS) for reduced intensity conditioning (RIC) compared to myeloablative conditioning (MAC). The vertical red line marks day 120 after transplantation.

Results

As predictors of primary interest, we defined disease stage, graft source, conditioning regimen intensity, and KPS. Cox regression modeling for OS and testing of PHA indicated a strong non-proportionality of KPS (<80 *vs.* 80-100; P<0.001), conditioning regimen (RIC *vs.* base-line MAC; P<0.001), and graft source (PBSC *vs.* BM; P<0.001) (Table 2). The estimates for disease stage showed no time-dependent effect in OS analysis. For DFS, KPS (P<0.001), RIC *vs.* MAC (P<0.001), and PBSC *vs.* BM (P=0.001), as well as both the estimates for intermediate disease stage (intermediate disease stage *vs.* base-line early disease stage; P=0.004) and advanced disease stage (advanced disease stage *vs.* base-line early disease stage; P<0.001) indicated strong time-dependent effects. Based on these results, cumulative hazard curves were plotted to visualize the change in hazards over time [OS: Figures 1-3; DFS: *Online Supplementary Figures S3-S7*]. After selection of optimal cut-off points, final multivariate models were adjusted in a time-dependent manner. For the estimates of all time periods, the PHA assumption held; the PHA test was not significant (Tables 2 and 3), which proved that the specified cut-off points were adequate for our dataset. For OS analysis, these time periods were before and after day 139 for KPS, before and after day 120 for conditioning intensity, and before and after day 388 for graft source (Figures 1-3). For DFS, these time periods were before and after day 130 for KPS, before and after day 140 for conditioning intensity, before and after day 242 for graft source, before and after day 309 after transplantation for intermediate disease stage, and before and after day 242 for advanced disease stage (*Online Supplementary Figures S3-S7*). Follow up of the cohort was split into observation intervals, and separate hazard ratios were obtained for time-dependent covariates within the respective time periods. For analysis of KPS, the estimates obtained were considerably higher in the early period after transplantation as compared to the period later on. Unadjusted values lay in between (Tables 2 and 3). For conditioning therapy, no significant effect on OS and DFS was found in the unadjusted model, but modeling of time-dependent effects showed opposing significant effects before and after the cut-off time points for OS and DFS. In the early post-transplant period, reduced intensity conditioning appears to have a protective effect, whereas later on a higher hazard ratio was found for patients who had

undergone conditioning with a reduced intensity. The observed effects correlate with lower early mortality and higher late mortality in this patient cohort. In the unadjusted models for OS and DFS, graft source did not correlate with significantly different outcomes. Interestingly, the effect of graft source could be better described by modeling separate estimates for the early and for the later period after transplantation. Similarly for OS and DFS, significant effects were found in the early period after transplantation, indicating lower probability for adverse events for patients transplanted with PBSC, while in the later period these patients seemed to do less well, as higher hazard ratios were observed (Tables 2 and 3). Other predictive covariates in the final model were: age (OS: HR 1.013, CI: 1.011-1.015, $P<0.001$; DFS: HR 1.011, CI: 1.01-1.013, $P<0.001$), year of transplantation 2001-2005 (OS: HR 0.89, CI: 0.82-0.96, $P<0.001$; DFS: HR 0.94, CI: 0.87-1.01, $P=0.092$), year of transplantation 2006-2013 (OS: HR 0.73, CI: 0.67-0.79, $P<0.001$, DFS: HR 0.81, CI: 0.75-0.88, $P<0.001$), and time to transplantation from initial diagnosis of more than 12 months (OS: HR 1.18, CI: 1.12-1.25, $P<0.001$; DFS: HR 1.21, CI: 1.16-1.29, $P<0.001$). A subset analysis was performed for transplantations using PBSC grafts performed between 2006 and 2013. For OS analysis, KPS and conditioning intensity showed time-dependent characteristics. For DFS analysis, KPS, conditioning intensity as well as disease stage (intermediate and advanced stage) could be confirmed as time-dependent variables. For both end points, similar risk estimates such as those seen in the analysis of the complete dataset were obtained in the subset analysis (*Online Supplementary Tables S2 and S3*).

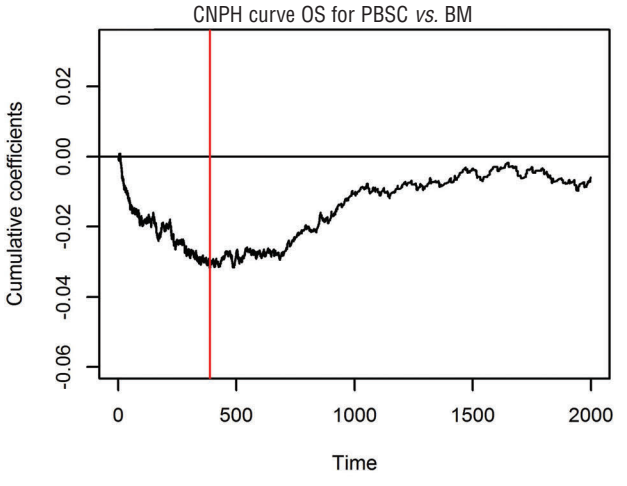

Figure 3. Cumulative non-parametric hazard curve (CNPH) curve, overall survival (OS) for peripheral blood stem cells (PBSC) compared to bone marrow (BM) as graft source. The vertical red line marks day 388 after transplantation.

Discussion

We show in a large German patient cohort that important clinical predictors (KPS, disease stage, conditioning regimen and graft source) exhibit significant time-dependent effects. These effects may be quantified in a sense that regression coefficients and relative risk can be calculated for follow-up periods, which satisfy the PHA, by extending the Cox regression model. We describe patterns that were valid for all disease entities included in this analysis. Our data contain a large proportion of transplantations performed in recent years (2006-2013, 56.9%) and may, therefore, represent current approaches in conditioning treatment and supportive care. The selection of diagnoses aimed to include a large number of patients, while, on the other hand, restricting disease-associated heterogeneity to a limited number of disease entities. Considering time-dependent effects may reveal relationships that could remain undetected in standard Cox regression models. Such effects can be shown for KPS in analysis of OS and DFS. In the early phase after transplantation, the hazard ratio is substantially higher as compared to the later phase after day 139 (OS) and 130 (DFS), respectively. The biological explanation is that the higher risk in the early phase may be attributed to a substantially higher early mortality. This indicates that patients with a poor KPS tolerate transplantation-associated morbidity/toxicity less well. In the later phase, the effect of transplantation morbidity/toxicity disappears (according to our data around 139) and the higher risk in this group diminishes. It is, therefore, no coincidence that the estimated cut-off time points are similar for KPS and conditioning treatment (between days 120-140 post transplantation) as they most likely both correlate to conditioning-associated toxicity.

Perhaps even more interesting is the observation that the intensity of conditioning regimen, which did not show a statistically significant effect in standard regression modeling, showed clearly significant effects depending on the time interval after transplantations. Our data show that, in the early phase after transplantation, risk is significantly reduced for patients who had undergone a dose-reduced conditioning regimen when compared to patients who had received a standard myeloablative conditioning therapy. This observation was obtained consistently for OS- and DFS-associated risk (OS: HR 0.81; DFS: HR: 0.88). However, later on and presumably after resolution of acute conditioning toxicity, the slope of the cumulative hazard curve changes from negative to positive and the risk is higher in the patient group with reduced intensity conditioning. It has been shown that non-relapse mortality is lower in patients undergoing RIC as compared to patients treated with MAC.[21] As non-relapse mortality is dominated by treatment-associated mortality soon after transplantation, this effect reflects a potentially protective effect of RIC in the early phase. However, later on, and as conditioning toxicity resolves, a differential effect on relapse remains. Lower relapse rates are observed in patients treated with MAC compared to patients treated with RIC, which explains the adverse effect of RIC in the later time period after HSCT.[22] The time-dependent effect of conditioning treatment may, therefore, be explained by opposing effects with different time kinetics on treatment-related mortality and relapse incidence. The cumulative hazard curves are very similar for OS and DFS (Figure 2 and *Online Supplementary Figure S4*). The differences in cut-off time points are probably a result of optimal model selection rather than a correlate of clearly different time kinetics. We could show an almost 20% risk reduction in early mortality for patients with RIC in the OS analysis, which is of clinical significance. The above-mentioned differences were less intense in the DFS analysis, which may be attributed to the fact that higher relapse rates observed

in RIC patients tend to partly attenuate the beneficial effect of RIC on mortality for this end point. Without considering time dependency, these effects would not have been noticed. In a Kaplan-Meier analysis, such a relationship is shown as overlapping survival curves (*Online Supplementary Figure S1*).

Peripheral blood stem cells are considered by many transplant physicians as graft source of choice for adult patients with malignant diseases of the hematopoietic system.[23,24] One advantage of PBSCs is the faster engraftment compared to bone marrow, leading to a shorter aplasia time, which may reduce early post-transplant morbidity.[25] Our data do not support this practice, as patients who were transplanted with allogeneic PBSCs showed similar hazard ratios compared to patients receiving BM grafts (OS: HR 0.94, CI: 0.87-1.01; *P*=0.069). However, regarding both end points, a significant time-dependent effect was detected, which, similarly to the analysis of conditioning regimen, allowed a distinction between a phase with risk reduction and a phase with increased risk. In the early phase (before day 388 in OS and before day 242 in DFS), PBSC had a protective effect, possibly because of the faster engraftment time, which reduces aplasia time and in turn has a beneficial influence on transplant-related mortality early after transplantation.[26] In the later phase, a risk increase for PBSC was observed, which might be attributed to late complications, e.g. chronic GvHD.[27-29] The earlier cut-off time point for DFS is caused by the addition of relapse events, which accumulate early after transplantation when compared to OS. For analysis of OS, a 21% risk reduction in the early phase after transplantation was seen, while the risk afterwards was increased by 23%. These differences are in the magnitude of a single HLA-mismatch and are of clinical relevance. DFS analysis revealed weaker differences compared to those observed in OS analysis, underscoring the stronger influence of relapse events in the analysis of DFS. In general, PBSC might be preferred in cases where high early post-transplant mortality is expected, e.g. in elderly patients or in patients with relevant comorbidity.

In DFS analysis, the variables "intermediate disease stage" and "advanced disease stage" showed time-dependent effects when compared with early disease stage, which was not seen in the OS analysis. The reason for this observation is that occurrence of relapse is counted as an event in DFS analysis, whereas for OS only death from relapse accounts for an event. Since relapse does not necessarily lead to death from relapse, and as many relapse events occur early after transplantation, such events accumulate for DFS in the early phase after transplantation, particularly for patients with more advanced disease stages. The difference in relative risk seen between advanced and early disease stage patients (HR 2.43 early after transplantation *vs.* HR 1.79 in the later phase after transplantation) highlights the problem of early morbidity/mortality in advanced disease stage patients, while on the other hand predicting a more favorable outcome for such patients who managed to survive these early complications.

Other commonly used approaches for inclusion of time-dependent variables are stratification on non-proportional variables as well as landmark analysis with refitting of separate models. Stratification, however, allows no quantification or comparison of the effects of various levels of the stratified variable. For the landmark approach, different models are fitted based on landmark time points with inclusion of case subsets with a survival time, which is at least as long as the landmark time points.[30] It is often difficult to interpret the results of landmark analysis from a clinical perspective, as subsets of patients are analyzed in each landmark model and multiple models might be necessary. Generally speaking, our approach could be considered as a variant of a landmark approach using the cut-off time points we describe as landmarks. However, only one statistical model is necessary to include all time-dependent variables, which facilitates interpretation. Adjustment of variables for time dependency, in the way we describe here, does not affect the estimates for the other covariates satisfying the PHA in the multivariate Cox regression model. To address the problem of confounding by older transplantations and over-representation of BM grafts in this group, a subset analysis was performed for patients transplanted with PBSC grafts between 2006 and 2013. Very similar risk estimates to those seen in the analysis of the complete dataset were found (*Online Supplementary Tables S2 and S3*), indicating that our approach did not introduce a relevant bias, but instead allowed us to evaluate the effect of graft source in the context of the other predictors mentioned.

A limitation of our analysis is that exact HLA-matching patterns could not be included, as definitions of HLA-matching changed over time, and detailed information about the number and resolution of HLA-mismatches is currently not available from the DRST database. Therefore, stratification had to be performed for the covariate "donor type". In addition, the dataset included also historical transplantations, which leads to heterogeneity with regard to changes in transplant protocols, graft source, donor selection, or supportive care. This heterogeneity is probably only partly reflected by including the time period of transplantation as covariate. Another important clinical covariate is cytogenetic risk.[31,32] However, cytogenetic information is currently not available in the majority of patients in the DRST database, which precluded inclusion in our analysis.

Hematopoietic stem cell transplantation is a highly complex and multifactorial process. Understanding and evaluating time-dependent effects allows more sophisticated risk quantification in HSCT to be made, particularly when a predictor has differential impact on outcome, as is the case for conditioning regimen intensity and graft source. Such information may help clinicians choose treatment according to the individual patient.

Acknowledgments

The authors would like to thank the DRST Data administrators F. Strehle and H. Neidlinger for providing the clinical data for this analysis.

References

1. Finke J, Schmoor C, Bethge WA, et al. Prognostic factors affecting outcome after allogeneic transplantation for hematological malignancies from unrelated donors: results from a randomized trial. Biol Blood Marrow Transplant. 2012;18(11):1716-1726.
2. Grimwade D, Hills RK. Independent prognostic factors for AML outcome. Hematology Am Soc Hematol Educ Program. 2009;385-395.
3. Blaise D, Castagna L. Do different conditioning regimens really make a difference? Hematology Am Soc Hematol Educ Program. 2012;2012:237-245.
4. Spellman SR, Eapen M, Logan BR, et al. A perspective on the selection of unrelated donors and cord blood units for transplantation. Blood. 2012;120(2):259-265.
5. Lee SJ, Klein J, Haagenson M, et al. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. Blood. 2007;110(13):4576-4583.
6. Perez L, Anasetti C, Pidala J. Have we improved in preventing and treating acute graft-versus-host disease? Curr Opin Hematol. 2011;18(6):408-413.
7. Holtick U, Albrecht M, Chemnitz JM, et al. Comparison of bone marrow versus peripheral blood allogeneic hematopoietic stem cell transplantation for hematological malignancies in adults - a systematic review and meta-analysis. Crit Rev Oncol Hematol. 2015;94(2):179-188.
8. Algara M, Valls A, Marrugat J, Granena A. Early mortality in bone marrow transplantation for acute lymphocytic leukaemia a multivariate analysis of risk factors. Eur J Med. 1993;2(7):386-392.
9. Valls A, Algara M, Marrugat J, Carreras E, Sierra J, Granena A. Risk factors for early mortality in allogeneic bone marrow transplantation. A multivariate analysis on 174 leukaemia patients. Eur J Cancer. 1993;29A(11):1523-1528.
10. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53(282):457-481.
11. Cox D. Regression Models and Life-Tables. J Am Stat Assoc. Series B. 1972;34(2):187-220.
12. Therneau T, Grambsch P. Modeling Survival Data: Extending the Cox Model. Springer-Verlag; 2000.
13. Martinussen T, Scheike TH. Dynamic Regression Models for Survival Data. Springer; 2006.
14. Gratwohl A, Stern M, Brand R, et al. Risk score for outcome after allogeneic hematopoietic stem cell transplantation: a retrospective analysis. Cancer. 2009;115(20):4715-4726.
15. Furst D, Muller C, Vucinic V, et al. High-resolution HLA matching in hematopoietic stem cell transplantation: a retrospective collaborative analysis. Blood. 2013; 122(18):3220-3229.
16. 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.
17. Scheike TH, Zhang MJ. Extensions and applications of the Cox-Aalen survival model. Biometrics. 2003;59(4):1036-1045.
18. Zahl PH. Regression analysis with multiplicative and time-varying additive regression coefficients with examples from breast and colon cancer. Stat Med. 2003; 22(7):1113-1127.
19. Grambsch P, Therneau T. Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. Biometrika. 1994; 81(3):515-526.
20. Cortese G, Scheike TH, Martinussen T. Flexible survival regression modelling. Stat Methods Med Res. 2010;19(1):5-28.
21. Ringden O, Labopin M, Ehninger G, et al. Reduced intensity conditioning compared with myeloablative conditioning using unrelated donor transplants in patients with acute myeloid leukemia. J Clin Oncol. 2009;27(27):4570-4577.
22. Mohty M, Labopin M, Volin L, et al. Reduced-intensity versus conventional myeloablative conditioning allogeneic stem cell transplantation for patients with acute lymphoblastic leukemia: a retrospective study from the European Group for Blood and Marrow Transplantation. Blood. 2010; 116(22):4439-4443.
23. Stem Cell Trialists Collaborative Group. Allogeneic peripheral blood stem-cell compared with bone marrow transplantation in the management of hematologic malignancies: an individual patient data meta-analysis of nine randomized trials. J Clin Oncol. 2005;23(22):5074-5087.
24. Bertz H, Spyridonidis A, Ihorst G, et al. Marrow versus blood-derived stem cell grafts for allogeneic transplantation from unrelated donors in patients with active myeloid leukemia or myelodysplasia. Biol Blood Marrow Transplant. 2012;18(6):894-902.
25. Ringden O, Remberger M, Runde V, et al. Faster engraftment of neutrophils and platelets with peripheral blood stem cells from unrelated donors: a comparison with marrow transplantation. Bone Marrow Transplant. 2000;25 Suppl 2:S6-S8.
26. Champlin RE, Schmitz N, Horowitz MM, et al. Blood stem cells compared with bone marrow as a source of hematopoietic cells for allogeneic transplantation. IBMTR Histocompatibility and Stem Cell Sources Working Committee and the European Group for Blood and Marrow Transplantation (EBMT). Blood. 2000; 95(12):3702-3709.
27. Mohty M, Kuentz M, Michallet M, et al. Chronic graft-versus-host disease after allogeneic blood stem cell transplantation: long-term results of a randomized study. Blood. 2002;100(9):3128-3134.
28. Schmitz N, Beksac M, Bacigalupo A, et al. Filgrastim-mobilized peripheral blood progenitor cells versus bone marrow transplantation for treating leukemia: 3-year results from the EBMT randomized trial. Haematologica. 2005;90(5):643-648.
29. Schmitz N, Eapen M, Horowitz MM, et al. Long-term outcome of patients given transplants of mobilized blood or bone marrow: A report from the International Bone Marrow Transplant Registry and the European Group for Blood and Marrow Transplantation. Blood. 2006;108(13):4288-4290.
30. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response and other comparisons of time-to-event by outcome variables. J Clin Oncol. 2008; 26(24):3913-3915.
31. Tallman MS, Dewald GW, Gandham S, et al. Impact of cytogenetics on outcome of matched unrelated donor hematopoietic stem cell transplantation for acute myeloid leukemia in first or second complete remission. Blood. 2007;110(1):409-417.
32. Armand P, Gibson CJ, Cutler C, et al. A disease risk index for patients undergoing allogeneic stem cell transplantation. Blood. 2012;120(4):905-913.