Biol Blood Marrow Transplant 21 (2015) 848–854

Biology of Blood and Marrow Transplantation

journal homepage: www.bbmt.org

Reevaluation of the Pretransplant Assessment of Mortality Score after Allogeneic Hematopoietic Transplantation

Brandon K.C. Au [1], Ted A. Gooley [2], Philippe Armand [3], Min Fang [2], David K. Madtes [4], Mohamed L. Sorror [4], Michael J. Boeckh [4], Christopher J. Gibson [5], Hans Joachim Deeg [4], Rainer Storb [4], Frederick R. Appelbaum [4], Jason W. Chien [2], Paul J. Martin [4,*]

[1] *Department of Medicine, University of Washington, Seattle, Washington*
[2] *Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, Washington*
[3] *Division of Hematologic Malignancies, Dana Farber Cancer Institute, Boston, Massachusetts*
[4] *Division of Clinical Research, Fred Hutchinson Cancer Research Center; Department of Medicine, University of Washington, Seattle, Washington*
[5] *Department of Medical Oncology, Dana Farber Cancer Institute, Boston, Massachusetts*

Article history:
Received 25 November 2014
Accepted 12 January 2015

Key Words:
Allogeneic hematopoietic cell transplantation
Survival
Comorbidity
Risk assessment

A B S T R A C T

The Pretransplant Assessment of Mortality (PAM) score was developed in 2006 to predict risk of mortality after allogeneic hematopoietic cell transplantation (HCT). Transplant practices have evolved during the past decade, suggesting the need to reevaluate the performance of the PAM score. We used statistical modeling to analyze and recalibrate mortality based on overall PAM scores, its components, and conditioning regimen in a retrospective cohort of 1549 patients who had HCT from 2003 through 2009. PAM scores correlated with mortality, but the effect size was smaller in the current study than in previous studies. PAM scores also demonstrated a stronger association with mortality in patients who received myeloablative conditioning than in those who received reduced-intensity conditioning. In contrast to the original study, carbon monoxide diffusing capacity, serum alanine aminotransferase, and serum creatinine concentrations were no longer significantly associated with 2-year mortality, whereas patient and donor cytomegalovirus serology was associated with mortality in the current cohort. Based on our findings, we developed and tested a revised PAM score for clinicians to estimate survival after allogeneic HCT with myeloablative conditioning regimens for patients with hematologic malignancy. Prognostic models such as the PAM score should be updated and recalibrated periodically to accommodate changes in clinical practice.

INTRODUCTION

Allogeneic hematopoietic cell transplantation (HCT) continues to be associated with high early mortality compared with other treatments for hematologic malignancies. Clinical tools to estimate this risk include the Pretransplant Assessment for Mortality (PAM) score, which uniquely integrates patient age, disease risk, selected transplant variables and certain measures of comorbidity to predict the risk of all-cause mortality at 2 years. Transplant variables in the PAM score include donor relationship, HLA matching, and type of conditioning regimen, whereas measures of comorbidity include forced expiratory volume in 1 second (FEV_1), carbon monoxide diffusing capacity (DLCO), serum creatinine concentration, and serum alanine aminotransferase (ALT) concentration [1]. The 50-point scoring system demonstrated a strong ability to predict 2-year mortality risk (Supplemental Table 1). Subsequent attempts to validate the PAM score in other studies have had mixed results [2-6].

Transplant practices have evolved during the past decade, including the increased use of nonmyeloablative or reduced-intensity conditioning (RIC) before transplantation. These changes suggest the need to re-evaluate the performance of HCT-related prognostic models such as the PAM score. The goal of the current study therefore was to determine the extent to which the PAM score and its components continue to predict mortality after HCT and to assess the performance of the PAM model based on the type of conditioning regimen. The latter was not well defined in the original study because of the limited numbers of patients treated with RIC regimens.

Financial disclosure: See Acknowledgments on page 853

* Correspondence and reprint requests: Paul J. Martin, MD, Fred Hutchinson Cancer Research Center, P.O. Box 19024, 1100 Fairview Ave. N., Seattle, WA 98109-1024.
E-mail address: pmartin@fhcrc.org (P.J. Martin).

http://dx.doi.org/10.1016/j.bbmt.2015.01.011
1083-8791/

METHODS

Patient Cohorts

The current cohort for this study included first-time allogeneic HCT recipients at the Seattle Cancer Care Alliance/Fred Hutchinson Cancer Research Center from January 1, 2003 through December 31, 2009. All patients were followed until death or the last day of contact as of December 31, 2011. We used the validation cohort from the previous study [1] (1990 through 2002) for comparison with the current cohort (Table 1). The Institutional Review Board determined that the use of deidentified patient information was exempt from review. To create an external validation cohort, additional data were obtained from the Dana Farber Cancer Institute (DFCI) and Brigham and Women's Hospital for HCT recipients from January 1, 2005 through June 30, 2009, with approval of the DFCI Institutional Review Board.

Clinical Variables

Donor type was determined according to HLA compatibility and patient–donor relation. Conditioning regimens were classified as myeloablative or reduced intensity (nonmyeloablative). Myeloablative regimens varied but typically contained high-dose cyclophosphamide with busulfan or 12.0 to 13.2 Gy total body irradiation (TBI), busulfan or treosulfan with fludarabine, or radiolabeled CD45-specific monoclonal antibody with fludarabine and 2 Gy TBI [7]. Conditioning regimens containing radiolabeled antibody were categorized as equivalent to >12 Gy TBI. Reduced-intensity regimens included 2 to 3 Gy TBI with or without fludarabine [8]. Pulmonary function testing was performed according to American Thoracic Society guidelines [9-11]. DLCO was adjusted for hemoglobin concentration according to the Dinakara equation [12]. FEV_1 and DLCO were expressed as a percentage of predicted values [13,14] and were capped at 100%, because higher values are not known to have physiologic significance with respect to HCT.

Statistical Analysis

Cox regression was used to assess the association of PAM score and individual PAM components with 2-year all-cause mortality, with follow-up censored at 2 years. PAM score and its continuous individual components were modeled as continuous variables, both linear and nonlinear, where the nonlinear modeling was done using a cubic spline with knots at the 5th, 25th, 50th, 75th, and 95th percentiles [15]. A cubic spline provides a flexible way to model continuous associations with outcome and requires minimal assumptions regarding a particular functional form. PAM components were also modeled categorically with the same cut-points used in the original report. PAM scores were categorized into various groups, and survival curves for patients in each group were plotted as Kaplan-Meier estimates.

The associations of PAM and its components with mortality in the current cohort were compared with the associations in the validation cohort from the original PAM study. We used the validation cohort because inclusion of patients from the original development cohort would overestimate the performance of PAM and the association of PAM and its components with outcome. The performance of PAM was reassessed using a c-statistic (see Supplemental Data). The Akaike information criteria were also calculated to assess model fit, where smaller values indicate a better fit.

The interactions of PAM score with conditioning intensity and cohort were assessed by fitting the appropriate term in a Cox regression model, with PAM modeled as a continuous linear variable. A revised PAM score was developed from patients in the current cohort who received myeloablative conditioning for hematologic malignancy. The performance and fit of the revised PAM score was assessed through bootstrapping (see Supplemental Data). The external validation cohort was also used to assess the performance of the revised PAM score. All statistical analyses were performed using SAS (Version 9.3, SAS Institute, Inc., Cary, NC).

Table 1
Baseline Clinical Characteristics of the Previous Validation and Current PAM Cohorts

Factor	Previous Cohort*	Current Cohort*
Median patient age, yr (range)	41.8 (15.0-72.5)	50.7 (15.1-78.9)
Disease risk,† n (%)		
Low	382 (29)	237 (15)
Intermediate	358 (27)	885 (57)
High	574 (44)	427 (28)
Donor type, n (%)		
Related/matched	670 (51)	584 (38)
Related/mismatched	125 (10)	79 (5)
Unrelated	519 (39)	886 (57)
Conditioning, n (%)		
Reduced intensity	72 (5)	609 (39)
Myeloablative, no TBI	469 (36)	547 (35)
Myeloablative, TBI ≤ 12 Gy	326 (25)	253 (16)
Myeloablative, TBI > 12 Gy	497 (34)	140 (9)
Median creatinine, mg/dL (range)	.8 (.3-5.7)	.9 (.3-9.7)
Median ALT, U/mL (range)	27 (0-908)	23 (4-349)
Median FEV_1, % of predicted (range)	93.1 (26.4-100)	92.2 (32.0-100)
Median DLCO, % of predicted (range)	99.2 (9.7-100)	80.0 (31.2-100)

* Date range October 1, 1990 through December 31, 2002 for the previous cohort compared with January 1, 2003 through December 31, 2009 for the current cohort.

† Categorized according to reference 1.

RESULTS

Cohort Characteristics

We identified 1665 patients who received a first allogeneic HCT between January 1, 2003 and December 31, 2009. Data for all 8 PAM components were available for 1549 patients: 940 treated with myeloablative conditioning and 609 treated with RIC. Table 1 summarizes baseline clinical characteristics of the current cohort and the previous validation cohort. The overall mean PAM score was 23.1 (median 23, range 8 to 43) in the current cohort, and the distributions of PAM scores were similar for patients who received myeloablative conditioning (mean 23.3, median 24, range 11 to 43) or RIC (mean 22.9, median 22, range 8 to 41).

Association of PAM with Outcome and Performance of PAM in the Current Cohort versus Previous Validation Cohort

In the current cohort, increasing PAM score was associated with a higher risk of death. With PAM score modeled as a continuous linear variable, the risk of death from any cause increased by 8% with each 1-point increase in PAM score (hazard ratio [HR], 1.08; 95% confidence interval [CI], 1.07 to 1.10; $P < .0001$). This result compares with a relative increase of 12% in the previous PAM validation cohort (HR, 1.12; 95% CI, 1.11 to 11.14; $P < .0001$). A statistically significant interaction between score and cohort was observed ($P < .0001$), indicating the magnitude of the association of score with outcome differed between the 2 cohorts. Modeling the PAM score as a cubic spline visually showed the strength of the association was weaker in the current cohort than in the previous cohort (Figure 1). The c-statistic for PAM was .62 (95% CI, .60 to .64) for the current cohort, compared with .68 (95% CI, .67 to 0.70) in the previous validation cohort.

Figure 2 shows the association of PAM with survival to 2 years for the current cohort and the previous validation cohort. Patients with the highest PAM scores in the current cohort demonstrated improved survival compared with those in the previous validation cohort. Although increasing PAM score is still clearly associated with decreased survival, the strength of the association in the current cohort is weaker than in the previous validation cohort, and the performance of PAM has diminished.

Association of PAM with Outcome and Performance of PAM in the Current Cohort, Myeloablative versus RIC

The proportion of patients who received RIC was higher in the current cohort (39%) than in the original PAM validation cohort (5%). We hypothesized that the strength of association between PAM and risk of mortality would be greater among patients who received myeloablative conditioning as compared with RIC, thereby partially explaining the weaker association between PAM score and outcome in the current study as compared with the original report. For