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

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

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

Abstract
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 re-evaluate 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 (FEV1), 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.
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]. DLEO was adjusted for hemoglobin concentration according to the Dinkacka equation [12]. FEV1 and DLEO 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, 50th, 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.

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 .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
patients who received myeloablative conditioning in the current cohort, the risk of death by 2 years post-transplant increased by 10% for each 1-point increase in PAM score (HR, 1.10; 95% CI, 1.08 to 1.12; *P* < .0001). For patients who received RIC, the risk of death by 2 years increased by 6% for each 1-point increase in PAM score (HR, 1.06; 95% CI, 1.03 to 1.08; *P* < .0001). These data suggest the magnitude of association of the PAM score with mortality is larger in the myeloablative group than in the RIC group, with a test of statistical interaction yielding *P* = .002. Kaplan-Meier survival curves by conditioning regimen for specified PAM groupings are available in Supplemental Figure 1. The concordance of PAM with mortality was higher in the myeloablative group than in the RIC group (c = .64 [95% CI, .62 to 0.67] and c = .57 [95% CI, .54 to .60], respectively).

**Association of PAM with Outcome and Performance of PAM with Myeloablative Conditioning, Current Cohort versus Previous Validation Cohort**

Differences in performance of the PAM score and reduction in the magnitude of association with outcome were also seen when the comparison of the current cohort with the previous validation cohort was restricted to patients who received myeloablative conditioning. We observed a 13% increase in 2-year mortality for each 1-point increase in PAM score (HR, 1.13; 95% CI, 1.11 to 1.14; *P* < .0001) in the previous validation cohort and 10% (HR, 1.10; 95% CI, 1.08 to 1.12; *P* < .0001) in the current cohort. A test of interaction between the PAM score and cohort yields *P* = .02, and Supplemental Figure 2 shows the association when the PAM score was modeled as a cubic spline. These results indicate that the magnitude of the association of PAM with mortality differed in the 2 cohorts. Among patients who received myeloablative conditioning, the c-statistic was .69 (95% CI, .67 to 0.71) for the previous validation cohort and .64 (95% CI, .62 to .67) for the current cohort.

**Individual PAM Variables Compared Between the Previous Validation and Current Cohorts among Patients Who Received Myeloablative Conditioning**

Given the decrease in magnitude of the association of PAM with outcome, we examined the association of each PAM component with outcome after HCT with myeloablative conditioning in both the current cohort and previous validation cohort. The adjusted association of each of the continuous factors with the risk of 2-year mortality is summarized in Figure 3, where the continuous factors not in question were modeled as linear functions for adjustment purposes. In the previous validation cohort, the associations between mortality and all continuous variables were statistically significant. In the current cohort, the associations between 2-year mortality and creatinine (linear *P* = .76, nonlinear versus linear *P* = .38, cubic spline *P* = .40), ALT (linear *P* = .78, nonlinear versus linear *P* = .32, cubic spline *P* = .47), and DLCO (linear *P* = .20, nonlinear versus linear *P* = .51, cubic spline *P* = .42) were not statistically significant.

The increased mortality risk associated with age appears to occur later in the current cohort than in the previous cohort. Results in the current cohort show a statistically significant positive association (linear *P* = .02, nonlinear versus linear *P* = .03, cubic spline *P* = .007), but the data suggest the association is nonlinear. The association between FEV1 and outcome was statistically significant in the current cohort (linear *P* = .006, nonlinear versus linear *P* = .77, cubic spline *P* = .009), similar to results in the previous validation
cohort. Supplemental Table 2 shows the association of the PAM components as originally categorized for both the current cohort and previous validation cohort, where notably both disease risk and type of donor were associated with outcome in the current cohort.

Development and Testing of a Revised PAM Model

Data from 914 patients in the current cohort who were diagnosed with a hematologic malignancy, received myeloablative conditioning, and had cytogenetic data available were used to develop a revised PAM model, where all factors considered for the original PAM score were re-examined [1]. Whereas the previous PAM model used an older disease risk classification, we reorganized overall risk groups according to a more updated risk index developed by Armand et al. [16]. In the original PAM model, unrelated donors were considered as a single group regardless of HLA matching or stem cell source. In the new model, we stratified unrelated donors using HLA matching (HR 1.40, \( P = .007 \) for 10/10 HLA-matched unrelated donors and HR 2.07, \( P < .0001 \) for 9/10 HLA-matched unrelated donors compared with HLA-matched related donors) and separated unrelated cord blood donors as a distinct group (HR 2.19, \( P = .002 \)). Patient age, donor type, disease risk, FEV1, and patient and donor cytomegalovirus (CMV) serology (\( P = .0005 \)) were included in the revised PAM, with FEV1 modeled as a continuous linear variable. Even though the association with age appeared to be nonlinear, a model with age dichotomized provided a better fit to the data as compared with the more complex model with a nonlinear function of age. Table 2 summarizes the scores and associations for the categorical and continuous factors for the revised PAM.

A separate cohort of 401 patients from DFCI was used to validate the revised PAM score, using the same selection criteria applied in the model-building cohort. Except for CMV (and age, where there were no patients older than 65), each of the revised PAM components was associated with 2-year mortality in the DFCI cohort (Supplemental Table 3). Modeling PAM as a continuous linear variable, each increase in PAM by 1 point was associated with an 8% increase in 2-year mortality (HR, 1.08; 95% CI, 1.06 to 1.10; \( P < .0001 \)). The assumption of linearity was consistent with the data when modeling PAM as a cubic spline (Supplemental Figure 3). Figure 4 displays Kaplan-Meier survival curves for the external validation cohort with PAM scores divided into 5 groupings of roughly equal interval lengths.

The bias-corrected Akaike information criteria for the revised PAM model was smaller compared with the original
DISCUSSION

The PAM score was originally published in 2006 as a simple and effective clinical scoring system and predictor of mortality [1]. Our follow-up analysis of the PAM score shows the performance of the original PAM score has diminished over time, as has the strength of the association with 2-year mortality. Similar changes are likely to occur with other predictive models [17-20].

Several factors likely contributed to the decreased discriminatory capacity of the PAM score and the loss of association of ALT, creatinine, and DLCO with mortality. Data have demonstrated improved outcomes after HCT over time [21]. In comparing patients who had HCT from 1993 to 1997 and from 2003 to 2007, Gooley et al. [21] showed significant reductions in nonrelapse mortality, all-cause mortality, liver dysfunction, acute kidney injury, and pulmonary complications through day 100. It is likely that changes in transplant practice and supportive care and the ability to manage comorbidities have all contributed to improved outcomes after HCT. Abnormalities in ALT, creatinine, and DLCO in the current study cohort of patients who received myeloablative conditioning regimens were relatively mild and do not extend to the range that can be allowed with RIC regimens. Patients with ALT, creatinine, or DLCO abnormalities more severe than observed in this study are very likely to experience higher risks of mortality after HCT with myeloablative conditioning regimens.

In updating the PAM model, we used a recently refined disease risk classification, stratified unrelated donors according to HLA matching, and analyzed cord blood recipients as a separate group. The differences in HRs assigned to each unrelated donor group now reflect a more accurately weighted contribution to the revised PAM model. The association between CMV status and mortality in the current cohort is consistent with a recently published large multicenter European analysis [22]. The sample size of the validation cohort may have been too small to detect this association, especially because the CMV effect is relatively small. Of note, the HRs for the CMV effect in both the current PAM study and the European study were within the 95% CI around the HR in the validation cohort.

Table 2
Factors Included in Revised PAM Model and Their Association with 2-Year Mortality

<table>
<thead>
<tr>
<th>Variable</th>
<th>HR (95% CI)</th>
<th>P</th>
<th>Score*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient age, yr</td>
<td>Reference</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>&lt;65</td>
<td>2.13</td>
<td>.0004</td>
<td>7.6</td>
</tr>
<tr>
<td>≥65</td>
<td>—</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Donor type</td>
<td>Global P &lt; .0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-matched, related</td>
<td>Reference</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Unrelated, 10/10</td>
<td>1.40 (1.10-1.79)</td>
<td>.007</td>
<td>3.4</td>
</tr>
<tr>
<td>Unrelated, &lt;9/10</td>
<td>2.07 (1.57-2.47)</td>
<td>&lt;.0001</td>
<td>7.3</td>
</tr>
<tr>
<td>Unrelated, cord blood</td>
<td>2.19 (1.35-3.57)</td>
<td>.002</td>
<td>7.9</td>
</tr>
<tr>
<td>HLA-mismatched, related</td>
<td>2.00 (1.10-3.64)</td>
<td>.02</td>
<td>6.9</td>
</tr>
<tr>
<td>Disease risk*</td>
<td>Global P &lt; .0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Reference</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>Intermediate</td>
<td>1.71 (1.07-2.75)</td>
<td>.03</td>
<td>5.4</td>
</tr>
<tr>
<td>High</td>
<td>3.75 (2.30-6.12)</td>
<td>&lt;.0001</td>
<td>13.2</td>
</tr>
<tr>
<td>Very high</td>
<td>5.49 (3.19-9.44)</td>
<td>&lt;.0001</td>
<td>17.0</td>
</tr>
<tr>
<td>FEV1</td>
<td>1.20 (1.09-1.28)</td>
<td>&lt;.0001</td>
<td>0.181 × (100 – %FEV1)</td>
</tr>
<tr>
<td>Patient/donor CMV</td>
<td>Global P = .0003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>—/—</td>
<td>Reference</td>
<td>—</td>
<td>0</td>
</tr>
<tr>
<td>—/+</td>
<td>1.71 (1.20-2.44)</td>
<td>.003</td>
<td>5.4</td>
</tr>
<tr>
<td>+/—</td>
<td>1.70 (1.31-2.21)</td>
<td>&lt;.0001</td>
<td>5.3</td>
</tr>
<tr>
<td>+/+</td>
<td>1.41 (1.07-1.86)</td>
<td>.01</td>
<td>3.5</td>
</tr>
</tbody>
</table>

* The score for each group of the categorical components was obtained by multiplying the appropriate regression coefficient by 10. For the continuous variable FEV1, the regression coefficient was multiplied by 10 to yield .168. Moreover, to ensure that all PAM scores generated were greater than zero, FEV1 was transformed to 100% and positive when FEV1 was less than 100%.

† HLA matching includes HLA-A, -B, -C, -DR, and -DQ alleles.

Overall risk groups were determined according to the risk index developed by Armand et al. [16], which includes disease risk, stage risk, and cytogenetic data for acute myeloid leukemia and myelodysplastic syndromes (MDS). The poor and very poor MDS cytogenetic risk categories defined by Deeg et al. [27] were grouped as high-risk disease, and all other categories were grouped as intermediate-risk disease.

* Represents the relative change in HR for each decrease in FEV1 by 10%.

Figure 4. Two-year survival for the external validation cohort (n = 401) from the Dana Farber Cancer Institute, grouped into 5 separate PAM categorizations that were divided into intervals of similar length. The stratification of groupings by equal interval lengths is supported by Supplemental Figure 3, because the association appears relatively linear. Category 1 (black): PAM 0-9; category 2 (red): PAM 9-14; category 3 (blue): PAM 14-19; category 4 (green): PAM 19-24; category 5 (pink): PAM 24-41.
The PAM score performs less well in patients treated with RIC regimens than in those treated with myeloablative regimens, as indicated by the c-statistic estimates and the statistical models (both cubic spline modeling and categorization of PAM score). Comorbidities not included in the PAM score are more likely to occur in patients treated with RIC regimens than in those treated with myeloablative regimens, and the omission of these comorbidities contributes to the relatively poor performance of the PAM score in predicting outcomes after RIC HCT. As expected, older age and comorbidities have lesser effects after RIC HCT as compared with myeloablative regimens. Additionally, differences in outcome according to disease risk groups used in the original PAM report (and duplicated here) were smaller in patients who received RIC as compared with myeloablative conditioning. After RIC, the intermediate- and high-risk groups were 1.34 and 1.61, respectively, times more likely to die by 2 years compared with patients in the low-risk group [23]. After myeloablative conditioning, these HRs were 1.97 and 4.61, respectively.

The PAM score and the HCT-specific comorbidity index (HCT-CI) [24] represent different tools for prognostication after HCT. The PAM score was developed as a simple global prognostic tool that incorporates patient age, selected comorbidities, disease risk, and donor type to predict overall survival at 2 years. In contrast, the HCT-CI is a comorbidity index designed specifically to capture the burden of multiple organ dysfunctions (n = 17 comorbidities) before allogeneic HCT to predict the risk of nonrelapse mortality. The HCT-CI was recently modified to incorporate patient age as an additional risk factor for nonrelapse mortality [25]. Even though this HCT-CI/age score does not consider disease risk or donor type, its performance in predicting survival is similar to its performance in predicting nonrelapse mortality [25]. The PAM and HCT-CI scores should complement each other, because each has components the other does not. Previous results have shown that consideration of disease severity led to a statistically significantly improved model when added to a model containing HCT-CI [26]. Similarly, we expect that consideration of certain components of HCT-CI would lead to a statistically significantly improved model when added to PAM. Both HCT-CI and PAM scores therefore provide useful information.

Our study has several limitations. Although the data clearly showed a change in association between PAM and mortality, it is impossible within the scope of this study to prove the underlying reason for these changes. The cubic spline graphs clearly demonstrate the relationship between PAM variables and mortality is complex and dynamic as opposed to simple and static, highlighting the potential limitations of categorical modeling. Results from these models could be used to select the most appropriate cutoffs for categorical analyses or, ideally, to develop more sophisticated noncategorical algorithms for estimating survival probabilities. The revised PAM score might not apply to patients with diseases other than hematologic malignancies and does not apply to patients treated with RIC regimens.

In summary, the association between PAM and mortality has changed over time. Its performance and the strength of association with outcome have diminished, and the risk factors for mortality have also changed. PAM provides better prediction for patients treated with myeloablative conditioning regimens than for those treated with RIC regimens. All components in the PAM score can be easily ascertained by referring physicians, although the FEV1i is seldom measured before patients arrive at the transplant center, unless they have pulmonary symptoms. If this information is available, then the updated correlation between PAM score and survival after HCT (available at www.pamscore.org) may assist clinicians in counseling patients during the initial discussion about the benefits and risks of HCT. The PAM score could also be used in balancing cohorts of patients involved in clinical trials. Our results indicate that prognostic tools should be re-evaluated and refined periodically, especially when clinical practices or patient characteristics change over time.

ACKNOWLEDGMENTS
Financial disclosure: Supported by grants CA18029, CA78902, and CA15704 from the National Cancer Institute.
Conflict of interest statement: There are no conflicts of interest to report.


SUPPLEMENTARY DATA
Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.bbmt.2015.01.011

REFERENCES


