FEATURED PAPERS

Development and validation of a major adverse transplant event (MATE) score to predict late graft loss in pediatric heart transplantation

Christopher S. Almond, MD, MPH, Helena Hoen, MS, Joseph W. Rossano, MD, Chesney Castleberry, MD, Scott R. Auerbach, MD, Lingyao Yang, MS, Ashwin K. Lal, MD, Melanie D. Everitt, MD, Matthew Fenton, MD, Seth A. Hollander, MD, Elfriede Pahl, MD, Elizabeth Pruitt, MPH, David N. Rosenthal, MD, Doff B. McElhinney, MD, Kevin P. Daly, MD, Manisha Desai, PhD, and on behalf of the Pediatric Heart Transplant Study (PHTS) Group Registry

From the Department of Pediatrics (Cardiology), Stanford University, Palo Alto, California, USA; Department of Medicine - Quantitative Sciences Unit, Stanford University, Palo Alto, California, USA; Department of Cardiology, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; Division of Pediatric Cardiology, St. Louis Children’s Hospital, St. Louis, Missouri, USA; Division of Pediatric Cardiology, The Children’s Hospital Colorado, Aurora, Colorado, USA; Division of Pediatric Cardiology, Primary Children’s Hospital, Salt Lake City, Utah, USA; Great Ormond Street Hospital, London, UK; Pediatric Heart Transplant Study Group, University of Alabama at Birmingham, Birmingham, Alabama, USA; Department of Cardiology, Lurie Children’s Hospital, Chicago, Illinois, USA; and the Department of Cardiology, Boston Children’s Hospital, Boston, Massachusetts, USA.

KEYWORDS:
congenital heart disease;
heart failure;
heart transplantation;
pediatrics;
risk-prediction model;
study design;
surrogate end-point

BACKGROUND: There is inadequate power to perform a valid clinical trial in pediatric heart transplantation (HT) using a conventional end-point, because the disease is rare and hard end-points, such as death or graft loss, are infrequent. We sought to develop and validate a surrogate end-point involving the cumulative burden of post-transplant complications to predict death/graft loss to power a randomized clinical trial of maintenance immunosuppression in pediatric HT.

METHODS: Pediatric Heart Transplant Study (PHTS) data were used to identify all children who underwent an isolated orthotopic HT between 2005 and 2014 who survived to 6 months post-HT. A time-varying Cox model was used to develop and evaluate a surrogate end-point comprised of 6 major adverse transplant events (MATEs) (acute cellular rejection [ACR], antibody-mediated rejection [AMR], infection, cardiac allograft vasculopathy [CAV], post-transplant lymphoproliferative disease [PTLD] and chronic kidney disease [CKD]) occurring between 6 and 36 months, where individual events were defined according to international guidelines. Two thirds of the study cohort was used for score development, and one third of the cohort was used to test the score.

RESULTS: Among 2,118 children, 6.4% underwent graft loss between 6 and 36 months post-HT, whereas 39% developed CKD, 34% ACR, 34% infection, 9% AMR, 4% CAV and 2% PTLD. The best predictive score involved a simple MATE score sum, yielding a concordance probability...
At present, there are no U.S. Food and Drug Administration (FDA)-approved immunsuppressants for pediatric heart transplantation (HT), nor has there ever been a successful randomized clinical trial in pediatric HT. Although many study design challenges exist, perhaps the single greatest challenge is the lack of statistical power stemming from the rareness of the disease (small sample size) and low event rate (e.g., death or graft loss). For example, to detect an inordinately large 2-fold difference in graft survival, which may not emerge for 5 to 10 years given the favorable transplant outcomes in the current era, one would need >500 subjects to achieve adequate power. To detect a more realistic treatment difference, such as a 10% improvement in graft loss over a 3-year study period, >6,000 subjects would be required.

One established strategy to decrease sample size is to increase the event rate by using a composite end-point. In adult cardiovascular trials, the major adverse cardiac event (MACE) score is perhaps the most widely used composite end-point, and includes mortality, non-fatal myocardial infarction (MI) and non-fatal stroke. A second strategy to decrease sample size is to use an ordinal rather than a binary endpoint. The Modified Rankin Scale (MRS), a 6-point stroke severity scale, has been used successfully to reduce sample size in stroke trials. We hypothesized that, by combining both of these strategies, we could develop a novel surrogate end-point for a pediatric HT trial comprised of major adverse transplant events (MATEs), using a sample size that is sufficient, yet feasible to conduct a randomized trial in pediatric HT medicine.

Therefore, the specific aims of this study are to: (1) develop and evaluate a surrogate end-point comprised of MATEs that can predict graft loss; and (2) estimate the sample size necessary for a randomized clinical trial. The broader purpose of this study was to find better ways to measure the safety and efficacy of competing immunsuppressant regimens to improve the overall safety and survival among children undergoing HT.

Methods

Study population and data source

Pediatric Heart Transplant Study Group (PHTS) Registry data were used to identify all children <18 years of age who underwent primary orthotopic HT between January 1, 2005 and December 31, 2014 and survived to 6 months post-HT. Subjects undergoing heart re-transplant, multi-organ transplant and heterotopic HT were excluded. The PHTS database is an event-driven pediatric HT registry that collects information from 54 North American centers, 2 centers in England and 1 in Brazil (for a complete list of participating centers, refer to Table S1 in the Supplementary Material, available online at www.jhltonline.org/). The details of the PTHS Registry have been reported elsewhere. This registry was selected because it captures the vast majority of pediatric HTs in the USA while also providing the most detailed information available regarding individual adverse events, severity level and outcome. Each center maintains institutional review board approval or obtains exemption according to institutional guidelines. The data were obtained under a limited data use agreement data set provided to Stanford University for this analysis.

Study definitions and end-points

The primary aim of this study was to develop and validate a surrogate end-point comprised of major adverse transplant events occurring between 6 and 36 months post-transplant that could predict late graft loss after pediatric HT. The 30-month window from 6 to 36 months post-HT was chosen because it simulated the projected time frame of a randomized clinical trial being planned by the authors and FDA to evaluate the safety and efficacy of everolimus conversion at 6 months after pediatric HT. The baseline characteristics of the study cohort were defined at the time of transplant. The level of hemodynamic support was categorized into 5 mutually exclusive categories: extracorporeal membrane oxygenation (ECMO); ventilator; ventricular assist device (VAD); inotropes; and oral therapy, as reported previously. Cardiac diagnosis was categorized as congenital heart disease (CHD) or other. Creatinine clearance (CrCl) was estimated using the Schwartz formula. The primary end-point of the study was death or graft loss leading to re-transplantation.

A priori MATE definitions and scoring

Table S2 (see Supplementary Material online) summarizes the a priori definitions of MATE and their severity levels and scores (weights), based on clinical experience, which served as an initial framework for model development. The 6 MATEs were defined according to published guidelines from the International Society for Heart and Lung Transplantation (ISHLT), the National Kidney Foundation (NKF) and the World Health Organization (WHO). Levels of illness severity and/or disease progression were also defined using the same guidelines and assigned an initial weight of 0 for no adverse event, 1 for a mild event, 2 for a moderate event, 3 for a severe event and 4 for an event resulting in death or graft loss. For infection and post-transplant lymphoproliferative disease (PTLD), we used risk factors for mortality identified in the ISHLT and WHO guidelines to generate low-, medium- and high-risk categories that were subsequently evaluated in the model.
development phase. Patients treated for rejection without a biopsy were categorized as severe rejection if they had evidence of hemodynamic compromise treated with inotropes (a PHTS definition analyzed previously)\textsuperscript{17} and moderate rejection if they had mild hemodynamic compromise; all other patients with treated rejection were categorized as mild rejection.\textsuperscript{17} Adverse event definitions were then mapped to PHTS data fields. In most cases, adverse event definitions could be mapped with reasonable precision because of the detailed information available in PHTS on adverse events. In cases where several possibilities existed, we analyzed variables based on the closest clinical definition and then according to what optimized model performance, as outlined in what follows.

**MATE score development and evaluation**

Two thirds of study cohort was randomly assigned to the training set, with the remaining one third assigned to the test set. Summary statistics are presented as median (interquartile range [IQR]) or number (percent). A wide variety of MATE score strategies were considered, covering a range of: (1) coding strategies for adverse events; (2) weighting strategies to score the severity of adverse events, assuming a 1-unit worsening did not carry the same risk across all adverse events or severity levels; and (3) imputation strategies when data were missing. Because the data were longitudinal and event-driven, missing values for adverse events not reported when another event was reported (e.g., when clinical rejection was reported, a value of PTLD was not usually reported) were imputed using a multiple imputation method.\textsuperscript{18}

**Statistical analysis**

A Cox model with time-varying covariates and time to graft loss as the outcome was used to predict the risk of graft loss based on the cumulative burden of adverse events between 6 and 36 months post-HT, as determined by the MATE score using a forward stepwise selection technique. Model performance was evaluated based on the concordance probability estimate (CPE) statistic, a C statistic suitable for analyzing longitudinal data.\textsuperscript{19} The fully conditional specification was used for multiple imputation of missing values, assuming a non-decreasing adverse event severity function.\textsuperscript{18,20,21}

**Results**

**Patient cohort**

Of 2,671 patients who underwent HT during the study period, 2,118 met the study inclusion criteria, (494 died or had graft loss in the first 6 months post-HT, 48 were excluded for multi-organ transplant and 11 were excluded for heterotopic transplants). Of these, 1,412 subjects (two thirds) were randomly assigned to the training cohort and 706 to the test cohort. Overall, there were no significant differences in the baseline characteristics of the 2 cohorts, which are summarized in Table 1.

**Overall graft survival and association with adverse events**

Overall, the median follow-up time post-HT was 3.7 years (range 0.5 to 10 years post-transplant), during which 230 patients had a graft loss event (11%). Figure 1 depicts the Kaplan—Meier graft survival by individual adverse event using the maximum score observed in the 6- to 36-month post-HT window. Overall, the association between the maximum score and graft loss was reasonably strong for cardiac allograft vasculopathy (CAV), acute cellular rejection (ACR), antibody-mediated rejection (AMR), PTLD and infection, but weaker for chronic kidney disease (CKD). When the time frame was restricted to the 30-month window between 6 and 36 months, 6.4% of subjects had a graft loss event.

**Final MATE score**

The best-fitting or final MATE score is summarized in Table 2, after considering a wide range of possible variable coding, weighting and imputation strategies (see Table S3 in Supplementary Material online). Overall, in the test cohort, the CPE statistic for the final MATE score was 0.74 (Figure 2) and corresponded to the simple MATE sum of scores with a non-decreasing constraint accumulated between 6 and 36 months, with a modest modification to the a priori CKD categories. Whereas in the a priori model we divided CKD risk categories based on CKD stage (see Table S2 in Supplementary Material online), we found the best-fitting model assigned CKD Stage 1 and 2 to the lowest risk group, Stage 3A to the mild risk group, Stage 3B to moderate risk group and Stages 4 and 5 to the highest risk group. Last, to preclude multiple, non-fatal, low-risk adverse events summing to a total MATE score greater than death or graft loss (score of 4 in the a priori model), we assigned a maximum score of 24 to death or graft loss—equivalent to a score of 4 (graft failure) for all 6 major adverse transplant events. Figure 2 depicts the relative contributions of each adverse event to the CPE statistic, illustrating how even low-frequency events, such as PTLD, can still make an important contribution to the ability of the model to predict death/graft loss.

Figure 3 summarized the observed distribution of the final MATE score exclusive of graft loss between 6 and 36 months post transplant after imputation. The simple sum of scores (totaling 19,360 observations in 706 patients) appears normally distributed with a mean of 10 and standard deviation of 4 units, inclusive of graft loss where graft loss is assigned a score of 24, the standard deviation is 4.4 units. Using the adverse event classifications in the final MATE model, the frequency of adverse events between 6 and 36 months included CAV in 4% of visits, PTLD in 5%, CKD in 9%, ACR in 36%, AMR in 4% and infection in 45%. The association between the maximum MATE score and long-term graft survival is depicted in Figure 4A. The association between the maximum MATE score and graft loss for 3 adverse events (CKD, CAV and ACR) of special interest to the FDA for a trial involving a proliferation signal inhibitor is depicted in Figure 4B. Table 3 illustrates the clinical application of the MATE score in a hypothetical patient after imposing a non-decreasing constraint on values within a subject as described in Table S3. At 6 months post-HT, a 10-year-old girl is free of adverse events. At 7 months post-HT, she develops Grade 2R ACR for which
she is assigned a score of “2” for ACR, and “0” for the remaining adverse events because none is reported. At 14 months post-HT, she develops CMV disease treated with ganciclovir, for which she is assigned an infection score of “3,” retaining a score of “2” for ACR, and assigned “0” for the remaining advents. On her annual evaluation at 24 months post-HT, she is found to have moderate CAV and a glomerular filtration rate (GFR) of 58 ml/min/1.73 m² (down from 68 ml/min/m²), for which she is assigned a “2” for CAV and a “1” for CKD. At 30 months post-HT, her CAV is found to be severe, for which she is assigned a “3” for CAV, and then listed for HT. At 34 months post-HT, she receives a new heart (graft loss), for which she is assigned a score of “24,” corresponding to a “4” across all 6 MATE categories. This example illustrates how recurring events and events of worsening severity are factored into the
overall MATE score. Although we considered the effect of multiple adverse events (e.g., rejection) in the model development phase, the model performed best when the highest severity level alone was included in the model.

Sample size considerations

To illustrate the utility of the MATE score to reduce sample size, we estimated the sample sizes necessary for a randomized clinical trial of 2 immunosuppressant regimens between 6 and 36 months post-HT. Assuming an accrual period of 24 months, a minimum follow-up time of 30 months, a normal distribution of MATE scores with standard deviation of 4 units (Figure 3), and a within-subject correlation ranging from 0.2 to 0.5, a cohort of 200 subjects (100 in each arm) would have >80% power to detect non-inferiority, assuming a non-inferiority margin of 2 MATE units using a 1-sided, 2-sample t-test at the 0.025 level of significance. With this sample size of 200 subjects, there is only 12% power to detect a relatively large 45% difference in graft survival, assuming a baseline graft loss rate of 6% between 6 and 36 months. In response to statistical feedback from the FDA, we have also calculated the sample size required to test a superiority hypothesis for treatment efficacy, based on the MATE score derived from CAV, CKD and ACR only (MATE-3), which has a standard deviation of 2.4. Assuming a within-subject correlation of 0.5, a cohort of 208 subjects (104 in each arm) would have 85% power to detect a mean difference of 1 MATE unit in the baseline-adjusted 30-month MATE scores of two treatment groups at the 0.05 level of significance.

Discussion

In this study, we developed and evaluated a MATE risk-prediction score that predicts death or graft loss in pediatric HT patients. This risk-prediction score is comprised of 6 MATEs, CKD, CAV, ACR, AMR, infection and PTLD, and reflects the cumulative burden of major adverse transplant events. The MATE score performs well in
predicting graft loss in children followed from 6 to 36 months post-HT, and therefore may be useful as a surrogate end-point for clinical trials in pediatric HT.

Our findings are consistent with previous studies that have shown CAV,22,23 rejection,24,25 PTLD26,27 and infection28,29 to be associated with graft loss and death. To date, however, no studies have incorporated these traditional transplant complications into a risk-prediction tool that has been evaluated in its ability to predict graft loss. Previous HT risk-prediction studies have been validated;
however, they were limited to covariates present at the time of heart transplantation. Thus, these models have excluded transplant complications, such as rejection and infection, that occur after transplant and may be time-varying. To our knowledge, this is the first report to develop and evaluate a risk-prediction model involving multiple time-varying adverse events post-HT, while demonstrating that the model can predict graft loss based on the cumulative progression of events through 36 months post-HT with strong concordance.

We chose a 30-month time window for the MATE analysis beginning at 6 months post-HT to simulate the projected time frame of a randomized clinical trial involving conversion of maintenance immunosuppression at 6 months. Because the frequency of adverse events after 6 months post-HT is considerably lower during the first 6 months, we felt that a time frame of shorter than 30 months may be insufficient for a critical number of adverse events to occur. On the other hand, a longer follow-up time frame would be less feasible to conduct because of increases in trial costs and greater patient attrition.

We were somewhat surprised to discover that a simple sum of MATE scores performed better than a variety of more complex weighting strategies. We did not want to make the assumption that the risk attributable to a 1-unit increase was uniform within or across major adverse events. Ultimately, however, it turned out that the simplest weight strategy (e.g., 0, 1, 2 and 3, corresponding to none, mild, moderate and severe, respectively) performed best. It is also of interest that no one adverse event made an outsized contribution—as reflected in the CPE statistic corresponding to each MATE—suggesting that including each of the major adverse events or their synergy was important to the performance of the overall score in predicting graft failure.

This study has several major implications. First, the MATE score may be a useful surrogate end-point for designing an immunosuppression clinical trial in pediatric cardiac transplantation. In contrast to the fields of adult heart transplant, pediatric kidney and liver transplantation, there has yet to be a randomized trial of immunosuppression therapy in pediatric HT patients. Most believe the greatest challenge to designing a clinical trial is the lack of statistical power stemming from the rareness of the disease and the low end-point frequency (e.g., death and/or graft loss). This fundamental problem (Type II statistical error) has beleaguered a number of well-organized but negative pediatric cardiovascular trials where, in retrospect, the trial’s primary end-point has drawn scrutiny. The MATE score seeks to fill this gap by creating a variable that has leverage to power a trial, but also is associated with outcome. The association between adverse events and graft survival appears to hold true for the overall MATE score comprised of all 6 adverse events as well as the subset of 3 adverse events of interest to the FDA (CKD, CAV and ACR).
in evaluating the risk—benefit profile of a proliferation signal inhibitor for pediatric heart transplantation.

The use of surrogate end-points to design clinical trials has gained widespread support in recent years from academics, industry and the FDA. In recent years, the FDA has hosted a series of workshops on developing valid surrogate end-points for clinical trials in solid-organ transplantation. These workshops have emphasized that not all surrogate end-points are equally valid. To serve as a valid surrogate end-point for a regulatory trial, the FDA requires the surrogate end-point satisfy the Prentice criteria, which include: (1) that the surrogate end-point is correlated with a clinical outcome; (2) the surrogate end-point captures the net effect of treatment on the clinical outcome; (3) the pattern characterizing the effects of the surrogate end-point on the clinical outcome must be determined; and (4) the degree of deviation in that pattern is known. Herein we found that: (1) the MATE score is correlated with graft loss; (2) the MATE score captures a broad range of ways in which immunosuppressant may be linked to graft loss (rejection, infection, CKD, CAV, AMR and PTLD); (3) the pattern characterizing the effect of the MATE score on graft loss is known; and (4) the degree of deviation of the pattern is known. These characteristics suggest the MATE score may satisfy the Prentice criteria. Prospective data collection in the form of a clinical trial would provide the best opportunity to confirm that the MATE score satisfies the Prentice criteria.

Our study has several limitations. First, we identified patients retrospectively through a national registry of transplant recipients, which creates the opportunity for selection bias. However, the PHTS currently captures nearly 80% of transplant centers in the USA, and the characteristics of patients reported and not reported to the PHTS appear relatively similar based on internal comparisons with data from the United Network for Organ Sharing (UNOS), suggesting that selection bias did not play a significant role in the study’s findings. Second, PHTS data fields could not be mapped exactly to consensus definitions of adverse events in all cases, creating the potential for misclassification bias. However, in the large majority of cases, the definitions could be mapped precisely to PHTS data fields because of the PHTS’s level of coding detail. Where there was uncertainty, we created several coding options that were tested in the score development stage. These coding options performed well in the testing cohort, regardless of the coding strategy used, suggesting the overall effect on the scoring system was small. Misclassification bias of CKD stage could be caused by choice of formula to estimate the GFR. We sought to minimize this effect by using the most updated version, the modified Schwartz formula, which is also the preferred formula for children with existing renal dysfunction. Last, because the PHTS is an event-driven database, missing data could threaten the validity of the findings. To address this, we evaluated a wide variety of imputation techniques in the training cohort, as well as the sensitivity of the findings to the imputation technique, and found the results to be robust. Although this limitation remains relevant to a retrospective analysis of historical registry data, it is worth noting it would be less relevant in a prospective clinical trial where data on all adverse events could be incorporated into the data collection to minimize missing data.

In conclusion, we have developed and evaluated a surrogate end-point for graft loss in pediatric HT recipients that is comprised of 6 MATEs. The surrogate end-point, known as the MATE score, predicts graft loss well (C statistic = 0.74) and may be useful as a primary end-point for FDA clinical trials in pediatric HT, given that it appears to satisfy the Prentice criteria for surrogate end-points. This is important because the lack of a surrogate end-point has been one of the greatest challenges to designing clinical trials that evaluate promising new therapies in pediatric HT, and perhaps throughout solid-organ transplantation. Although survival after pediatric HT has improved considerably in the current era, many children still fail to survive into adulthood because of a predictable set of transplant complications. Novel clinical trial designs and end-points will afford the pediatric community the best prospect for evaluating promising medical therapies involving rare diseases.

Disclosure statement

The authors have no conflicts of interest to disclose. The interpretation and reporting of these data are the responsibility of

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Time since transplant (months)</th>
<th>CKD</th>
<th>CAV</th>
<th>ACR</th>
<th>AMR</th>
<th>Infection</th>
<th>PTLD</th>
<th>Graft loss</th>
<th>MATE sum with single imputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>24</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>34</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>24</td>
</tr>
</tbody>
</table>

The table demonstrates a non-decreasing MATE score value where the risk of earlier adverse events is carried forward contributing to the overall risk. ACR, acute cellular rejection; AMR, antibody-mediated rejection; CAV, coronary artery vasculopathy; CKD, chronic kidney disease; ID, identification; PTLD, post-transplant lymphoproliferative disease.

To avoid the scenario where multiple minor adverse events could sum to a score in a living patient that is higher than death/graft failure (4), by convention, patients with graft loss are assigned a score of 24, equivalent to a graft loss score of 4 across all 6 domains (4 × 6 = 24).
the authors and not an official policy of or interpretation by the PHTS. This study was funded in part by generous support of the Kate Marra family to the Pediatric Advanced Cardiac Therapies (PACT) Team Research Fund at Stanford University. The data were supplied by the PHTS to Stanford University under a limited data use agreement. The authors would like to thank Drs. Yulin Zhang, Ms. Miranda Zinsman and Ms. FeiFei Qin (Stanford University), Dr. Lynn Sleeper (Boston Children’s Hospital), and Dr. Charlie Canter (Washington University in St. Louis) for their thoughtful review of the manuscript. Most importantly, we thank the centers participating in the PHTS, especially the heart transplant clinical and data coordinators, who assume primary responsibility for submitting PHTS data.

Supplementary material
Supplementary materials associated with this article can be found in the online version at www.jhltonline.org/.

References