Interpretable Boosted Decision Trees for Prediction of Mortality Following Allogeneic Hematopoietic Stem Cell Transplantation

Roni Shouval1,2,*, Arnon Nagler1,4, Myriam Labopin4 and Ron Unger2

1Division of Hematology and Bone Marrow Transplantation, The Chaim Sheba Medical Center, Tel-Hashomer, Ramat-Gan, Israel
2Internal medicine 9th Department and the 2013 Pinchas Borestein Telpilot Medical Leadership Program, The Chaim Sheba Medical Center, Tel-Hashomer, Ramat-Gan, Israel
3The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel
4EBMT Paris Office, Hospital Saint Antoine, Paris, France

Allogeneic (allo) hematopoietic stem cell transplantation (HSCT) is a potentially curative procedure for selected patients with hematological disease. Despite a reduction in transplant risk in recent years, morbidity and mortality remains substantial, making the decision of whom, how and when to transplant, of great importance [1].

Numerous parameters affect transplant related risk. When indicated, clinical judgment often plays a key role in patient selection [2]. Risk scores for mortality prediction, such as the European Group for Blood and Marrow Transplantation (EBMT) risk score, the Hematopoietic Cell Transplant-Co-morbidity Index (HCT-CI) and others, may aid decision [3-5]. These risk scores were developed using a standard statistical approach and have been validated. However, their predictive accuracy is still sub-optimal [6-9].

The development of large and complex registries, incorporating biological and clinical data, and the need for improved prediction models, generate the drive to apply machine learning (ML) algorithms for clinical predictions [10,11]. ML is a field in artificial intelligence stemming from computer sciences. The underlying paradigm does not start with a pre-defined model, rather it lets the data create the model by detecting underlying patterns [11]. Thus, this approach avoids pre-assumptions about model types and variable interactions, and may complement standard statistical methods [12,13]. ML algorithms are often used as tools in the data mining approach for knowledge discovery in databases [11].

Motivated by the need for improved risk prediction of allogeneic HSCT, the potential benefits of ML algorithms and their success in other clinical scenarios, we performed a predictive data mining study on a large cohort of transplanted acute leukemia (i.e., Acute Myeloid Leukemia and Acute Lymphoblastic Leukemia) patients, developing a readily accessible prediction model for mortality following transplantation [14-18]. Methodological and clinical aspects of the model are discussed below, whereas a full description of the model is available under the following reference [19].

The study cohort consisted of 28,236 adult allogeneic HSCT recipients from the Acute Leukemia Working Party registry of the European Group for Blood and Marrow Transplantation. The primary objective was prediction of overall mortality (OM) at 100 days after HSCT. Secondary objectives were estimation of non-relapse mortality (NRM), leukemia-free survival (LFS), and overall survival (OS) at 2 years. Donor, recipient, and procedural characteristics were analyzed. The alternating decision tree (ADT) ML algorithm was applied for model development on 70% of the data set and validated on the remaining data.

Alternating decision trees are a generalization of decision trees that result from applying a variant of boosting to combine weak classifiers. Questions are asked iteratively, until a user pre-defined number of iterations are reached. The ADT Tree structure consists of alternating levels of prediction and decision nodes. Each prediction node is associated with a weight, representing its contribution to the final prediction score, while each decision node contains a binary single question regarding a certain attribute. In contrast to standard decision trees, where classification is achieved by following a unique path from the root to a leaf for a given unknown instance, prediction with ADT involves pursuing multiple paths, corresponding to the instance features. The cumulative score gathered by an instance (i.e., a patient being evaluated before transplant) is the sum of the prediction values along all paths that the patient traverses in the decision tree. A positive score implies membership of one class and a negative sum membership of the other. The absolute score value is directly correlated with the classification confidence [20,21]. We have transformed the score into a probability through a logistic transformation. The ADT is appealing for prediction in clinical scenarios, as it is an accurate boosting algorithm in which interpretability is preserved, as opposed to alternative ensemble techniques.

In the study cohort, the majority of patients had Acute Myeloid Leukemia (70%), were in first complete remission (60%) and received myeloablative conditioning (71.5%). Grafts from HLA matched sibling donors were used in 53.9% of patients. OM prevalence at day 100 was 13.9% (n=3,936), underscoring its significance as a valid predictive endpoint. For generation of a prediction model of day 100 OM the ADT algorithm was applied and optimized on the training set using 10 fold cross-validations. After calibrating the score on the validation set, day 100 OM probabilities were calculated and ranged from 3% to 68%. Model’s discrimination on the validation set for the primary objective (day 100 OM) performed better than the EBMT score (AUC=0.706 versus 0.646, p-value<0.00001). Per secondary objectives, cumulative incidence of 2 years NRM was 38.2% (34.7-41.7, 95%-CI) for the patients included in the highest score interval, with corresponding Kaplan Meier estimate of OS and LFS of 19.9% (17.22.9%, 95%-CI) and 17.5% (14.7-20.3%, 95%-CI) respectively. Probabilities of 2 years NRM, OS and LFS, for patients in the lowest score interval, were 9.8% (7.9-12, 95%-CI), 72% (68.8-75.1, 95%-CI) and 64.9% (61.6-68.2, 95%-CI) respectively.

*Corresponding author: Roni Shouval, Division of Hematology and Bone Marrow Transplantation, The Chaim Sheba Medical Center, Tel-Hashomer, Ramat-Gan, Israel, Tel: 972-52-6668162, E-mail: shouval@gmail.com

Received October 30, 2015; Accepted November 20, 2015; Published November 26, 2015


Copyright: © 2015 Shouval R, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Discrimination of the ADT model for 2 years overall survival outperformed the EBMT score (AUC=0.657 versus 0.647, p-value=0.04). An online version of ADT models is available at http://bioinfo.lnx.biu.ac.il/~bondi/web1.html.

The ADT algorithm selected 10 out of 20 variables, while providing information on the variables’ dependence status, importance, and interactions. Key determinants of overall mortality were disease stage and performance status, in line with earlier studies [22,23]. Interactions discovered include the increased risk of aggressive conditioning regimens in the older patients (age>37) and the benefit of experienced centers (>20 transplants per year) in performing transplants from HLA matched unrelated donors, highlighting the need for center accreditation [24-26]. Importantly, age was not an independent risk factor for overall mortality, probably reflecting an improvement in transplantation care and patient selection in recent years [27].

Eligibility of patients with acute leukemia for allo-HSCT is based on a risk benefit-assessment of the relapse risk versus transplant risk [28]. By applying the ADT algorithm, we have developed a novel prediction model based on 10 variables, for day 100 OM. Scores assigned correlated with objectives, enabling an individual continuous probabilistic evaluation of the primary objective (i.e., OM at day 100) and a discretized risk assessment of secondary objectives at 2 years (OS, NRM, and LFS). Apart from the ADT score clinical utility, it marks the introduction of the data mining methodology into HSCT prognostic modeling. Nonetheless, the ADT is a classification algorithm designed for handling binary endpoints, but not censored or continuous endpoints. Therefore, we focused on a short term outcome, in a population where loss of follow-up was below 5%. Prediction of long term survival was achieved indirectly. Future models should be utilizing algorithms designed for survival modeling, accounting for censored data [18,29]. In addition, improving predictive accuracy will likely require incorporation of additional biologic, genetic and clinical features.

## References


