

STATISTICAL SYMPOSIUM ORAL PRESENTATIONS

SC03.1

Outcome prediction after allogeneic stem cell transplantation: the doctor's perspective

J. Schetelig^{1,2,*} on behalf of Chronic Malignancies Working Party
¹Clinical Trials Unit, DKMS, ²Medical Department I, University Hospital Carl Gustav Carus, Dresden, Germany

Abstract: This presentation will focus on the doctor's perspective on outcome prediction. Starting points are questions arising in daily practice of allogeneic stem cell transplantation (alloSCT). The goal of the session is to link published data and statistical know-how to these everyday questions, such as giving a prognosis for a specific patient or determining the best moment for an intervention to prevent relapse. For this purpose the difference between risk factor analyses and true outcome prediction for new patients will be discussed and challenges in the use and interpretation of scores and predictive models will be addressed. Outcome prediction after alloSCT based on risk factors known prior to transplantation will be reviewed. The Pretransplant Assessment of Mortality (PAM) - Score which was developed at the Fred Hutchinson Cancer Research Center in Seattle (<http://pamscore.org/>; published by Au *et al*, BBMT 2015 & Parimon *et al*, Annals of Internal Medicine 2006) and an EBMT-registry based multivariate Cox regression analysis of a large cohort of CLL patients will be used as examples. Results from large multivariate regression analyses and validated scoring systems will be compared in terms of utility, challenges and limitations to answer everyday questions. More specifically the following topics will be addressed in this presentation: the assumption of causality, confounding factors, interactions between risk factors and specific model assumptions. Next the need for visualization of results from risk factor analyses and limitations will be addressed. In this context the question will be addressed why it is difficult to interpret hazard ratios. A specific challenge is the visualization of the impact of multiple risk factors on competing risks outcomes. Balanced and biased approaches for visualization will be shown. Finally, the need for cross-validation of scoring systems will be highlighted and examples will be discussed.

Disclosure of Interest: None declared.

SC03.2

Prediction models for competing risks outcomes – general concepts

H. Putter^{1,*}, L. de Wreede¹, N. Kroeger²

¹Dept of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, Netherlands, ²University Hospital Eppendorf, Hamburg, Germany

Abstract: The purpose of this presentation is to give a general overview of statistical concepts relevant for prediction models for competing risks outcomes in stem cell transplantation (SCT). The presentation will cover general aspects, like the purpose of prediction models - the distinction between etiological research and prediction will be addressed by the first presentation in the session. Aspects of model building will be considered next, in particular the choice between Cox models based on cause-specific hazards versus Fine & Gray models, and practical choices to be made, such as covariate selection, how to deal with missing data and with

time-dependent information (e.g., onset of aGvHD). After having constructed a candidate model, checking model assumptions and model validation become important. Checking the proportional hazards assumption and checking functional form of covariates is not straightforward when the prediction model is based on Fine & Gray regression. Ways of presenting the final results of the model will be examined. A useful graphical representation is obtained, for instance, by grouping patients into equally sized risk groups, based on the linear predictor of the risk model. With respect to model validation, predictive measures like the concordance measure of discrimination and prediction error will be reviewed. Finally, the issue of overfitting and the need for internal and external validation will be discussed. All steps in the construction and internal validation of models will be illustrated using an ongoing project on 1638 MDS patients who received an allogeneic stem cell transplantation from an HLA-identical sibling or a matched unrelated donor between 1995 and 2012 and who were reported to the EBMT registry. The aim of the project is to construct a series of prediction models for relapse in the first two years after transplantation, with non-relapse mortality as a competing risk.

Disclosure of Interest: None declared.

SC03.3

Risk factors for predicting non-relapse mortality and event-free survival: a case study in CLL

L. de Wreede^{1,2,*}, H. Putter², J. Schetelig^{1,3}

¹DKMS CTU, German Bone Marrow Donor Center, Dresden, Germany, ²Medical Statistics and Bioinformatics, LUMC, Leiden, Netherlands, ³Medical Department I, University Hospital of the Technical University, Dresden, Germany

Abstract: We will show how we analysed a dataset containing baseline risk factors and outcome variables collected in the CLL Data Quality Initiative of the EBMT as a case study illustrating some issues related to creating risk factor models based on a registry data set. Data on 694 patients were available. The goal of the analysis was to identify patients who could benefit most from alloHCT in the era of targeted therapy by identifying risk factors associated with low non-relapse mortality in the first 2 years after alloHCT and high 5-year event-free survival (using EFS as a surrogate for long-term disease-control).

Cox regression models for (cause-specific) hazards were fitted for the respective endpoints to evaluate a potential clinical impact of risk factors on transplant outcomes. We aimed at getting an overview of the most relevant risk factors and finding their optimal estimate in order to better understand their impact on transplant outcome, and therefore we did not aim for the most parsimonious models or for models optimized for prediction for new patients. To account for missing information for some risk factors we used multiple imputation to avoid bias and make use of all available data. Next, stepwise backward variable selection was performed. With the goal of showing the relevance of the fitted models for separating outcomes of patients with favorable and unfavorable characteristics, we plotted the model-based predicted probabilities for a set of relevant reference patients. For this purpose we selected good and poor risk male and female reference patients in such a way that the risk scores for the EFS model for the male patients were close to the 10th and the 90th

percentile in one MI dataset. For the cumulative incidence of non-relapse mortality, also the impact of the risk factors on the competing event relapse was taken into account.

This case study shows how to create regression models in a risk factor analysis, how missing data can be handled and how the impact of risk factors can be visualized by showing model-based curves for reference patients.

Disclosure of Interest: None declared.

SC03.4

How to Use Risk Factors for Outcome Prediction in Patients: Experience in Umbilical Cord Blood Transplantation with a Data Mining Approach

R. Shouval^{1,*}

¹Hematology and Bone Marrow Transplantation, Sheba Medical Center, Ramat-Gan, Israel

Abstract: Machine learning (ML) is a branch of artificial intelligence, in which the underlying paradigm does not start with a pre-defined model; rather it lets the data create the model by detecting underlying patterns. Thus, this approach

avoids pre-assumptions regarding model types and variable interactions, and may offer additional knowledge, which has eluded detection by standard statistical methods. ML algorithms, have been applied in "big data" scenarios such as computerized visions, and are tools in the data mining approach for knowledge discovery in large datasets.

Using a data mining approach, we have investigated how umbilical cord blood transplantation parameters can be combined to predict OS at 2 years following transplantation in a large cohort acute leukemia patients. Predictors were ranked according to their predictive contribution and interactions were analyzed. Random Survival Forest (RSF), a ML technique for predictive modeling accounting for right censored data, was used for the statistical analysis. RSF reduce variance and bias by using all variables collected and by automatically assessing for nonlinear effects and complex interactions.

The presentation will provide an overview of the data mining process, and explain how machine learning can be utilized for survival analysis, and risk factor identification in hematopoietic stem cell transplantation.

Disclosure of Interest: None declared.