0002

Exploring Limiting Factors in the Prediction of Allogeneic HSCT Related Mortality: An In-Silico Machine Learning Analysis of the Acute Leukemia Working Party (ALWP) Registry of the EBMT

R. Shouval[1,], M. Labopin[2], R. Unger[3], S. Giebel[4], F. Ciceri[5], C. Schmid[6], J. Esteve[7], F. Baron[8], N. C. Gorin[9], B. N. Savani[10], M. Mohty[11], A. Nagler[12]*

[1]Division of Hematology and Bone Marrow Transplantation, The Chaim Sheba Medical Center, Tel-Hashomer, Ramat-Gan, Israel, [2]EBMT Paris Office, Hospital Saint Antoine, Paris, France, [3]The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan, Israel, [4]Department of Bone Marrow Transplantation and Oncohematology, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Gliwice Branch, Gliwice, Poland, [5]Hematology and BMT Unit, San Raffaele Scientific Institute, Milan, Italy, [6]Department of Hematology and Oncology, Klinikum Augsburg, Ludwig-Maximilians-University, Munich, Germany, [7]Hematology Department, IDIBAPS, Hospital Clinic, Barcelona, Spain, [8]17-Hematology & GIGA research, University of Liège, Liège, Belgium, [9]4-EBMT Paris Office, Hospital Saint Antoine, Paris, France, [10]Hematology & Stem Cell Transplantation Section, Vanderbilt University Medical Center, Nashville, TN, United States, [11]AP-HP, Hématologie Clinique et Thérapie Cellulaire, Hôpital Saint-Antoine, Paris, France, [12]Division of Hematology and Bone Marrow Transplantation, The Chaim Sheba Medical Center, Tel-Hashomer, Ramat-Gan, Israel

Introduction: The establishment of large transplant registries and introduction of novel statistical techniques have paved the way for large scale data analysis. Nevertheless, contemporary tools for risk prediction of transplant related mortality (TRM) following allogeneic (allo) hematopoietic stem cell transplantation (HCT) are of limited clinical use, owing to a sub-optimal predictive accuracy.

Apart from inherent procedural uncertainty, methodological factors impeding prediction might be attributed to the statistical methodology used, number and quality of features collected, or simply the population size. Using an in-silico approach (i.e. iterative computerized simulations), based on machine learning (ML) algorithms, we aimed to define prediction limiting factors of day 100 TRM and rank variable contribution.

ML is a field of artificial intelligence dealing with the construction and study of systems that can learn from data, rather than follow explicitly programmed instructions. Commonly applied in complex data scenarios, such as financial settings, it may be suitable for outcome prediction of HCT.

Materials (or patients) and methods: Study population was a cohort of 28,236 adult acute leukemia allo-HCT recipients from the EBMT-ALWP. Twenty four variables were analyzed, including recipient, donor and transplant characteristics. Study design involved two phases. The first, focused on development and comparison of several ML based prediction models of day 100 TRM. In the second phase, a repetitive computerized simulation was applied. Factors necessary for optimal prediction were explored: algorithm type, size of data set, number of included variables, and performance in specific subpopulations. Models were assessed and compared on the basis of the area under the receiver operating characteristic curve (AUC).

Results: Six ML based prediction models for day 100 TRM were developed on the entire dataset. Optimal AUCs ranged from 0.65-0.68. Depending on the algorithm used for prediction model development, the *in-silico* experimental system yielded the following results: Predictive performance plateaued on a population size ranging from $n = 5647$-8471; A range of 6-12 ranked variables, selected by a separate feature selection algorithm, were necessary for optimal prediction; Disease status and donor type were consistently top ranking variables. Predictive performance of models developed for specific subpopulations, ranged from 0.59 to 0.67 for patients in second complete remission and patients receiving reduced intensity conditioning respectively.

Conclusion: We present a novel experimental system for assessment of prediction boundaries in HCT. The present approach has clinical implications. We show that predictive performance of day 100 TRM is unlikely to improve with the data routinely gathered on HCT recipients, as different algorithms reach approximately the same performance. In addition, an exhaustive search for variable importance, reveal that few variables "carry the weight" with regard to predictive influence. Predictive performance converged when sampling more than 5000 patients, reflecting the importance of large registry studies. Overall, it seems we have reached a point of predictive saturation. Improving predictive performance will likely require additional types of input like genetic, biologic and procedural factors.

Disclosure of Interest: None declared.