



DISCOVER THE NEXT GENERATION OF THE *BLOOD* APP

New Innovative Design!



[ASH Home](#) [Blood App](#) [My Folders](#) [Alerts](#) [RSS](#)

[Sign In](#)



Leading the way in experimental and clinical research in hematology

Q

Advanced Search

- |                               |                                    |                             |  |                               |                               |  |  |
|-------------------------------|------------------------------------|-----------------------------|--|-------------------------------|-------------------------------|--|--|
| <a href="#">Home</a>          | <a href="#">About <i>Blood</i></a> | <a href="#">Authors</a>     | <a href="#">Submit to <i>Blood</i></a> | <a href="#">Subscriptions</a> | <a href="#">Classifieds</a>   |  |  |
| <a href="#">Current Issue</a> | <a href="#">First Edition</a>      | <a href="#">Collections</a> | <a href="#">All Issues</a>             | <a href="#">Abstracts</a>     | <a href="#">Video Library</a> |  |  |

Home / December 6, 2014; Blood: 124 (21)

i No markup for post-processing

## Prediction of Allogeneic HSCT Related Mortality in Acute Leukemia: Exploring Boundaries of Prediction through Machine Learning Based Modeling. a Data Mining Study from the Acute Leukemia Working Party (ALWP) of the EBMT

Roni Shouval, MD<sup>\*,1,2,3</sup>, Myriam Labopin, MD<sup>\*,4</sup>, Ron Unger, PhD<sup>\*,2</sup>, Sebastian Giebel<sup>\*,5</sup>, Fabio Ciceri, MD<sup>\*,6</sup>, Christoph Schmid, MD<sup>7</sup>, Jordi Esteve, MD PhD<sup>8</sup>, Frédéric Baron, MD PhD<sup>9</sup>, Bipin N. Savani, MD<sup>10</sup>, Mohamad Mohty, MD PhD<sup>11</sup>, and Arnon Nagler, MD<sup>1,2,4</sup>

[+ Author Affiliations](#)

- |                |                                    |                                    |                           |                     |
|----------------|------------------------------------|------------------------------------|---------------------------|---------------------|
| <b>Article</b> | <a href="#">Figures &amp; Data</a> | <a href="#">Info &amp; Metrics</a> | <a href="#">E-Letters</a> | <a href="#">PDF</a> |
|----------------|------------------------------------|------------------------------------|---------------------------|---------------------|

December 06, 2014 [Table of Contents](#)

[← Previous](#)

### Abstract

Background: Allogeneic hematopoietic stem cell transplantation (allo-HSCT) has been shown to increase survival and induce cure of acute leukemia (AL). Unfortunately, transplant related mortality (TRM) remains high. Risk scores, based on a conventional statistical approach, have been developed for TRM prediction. These have been well validated. Nevertheless, predictive performance is sub-optimal; thus, limiting clinical utility. Factors impeding prediction might be attributed to the statistical methodology, number and quality of features collected, or simply the size of the population analyzed. We set to explore these factors, using a novel computational approach, based on machine learning algorithms (ML).

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

Study design: Using a cohort of 28,236 adult allo-HSCT recipients from the ALWP registry of the EBMT, transplanted between 2000-2011, owing to Acute Myeloid Leukemia or Acute Lymphoblastic Leukemia, and containing 24 variables (i.e., patient, leukemia, donor, and transplant characteristics) we devised a two phase data mining study 1) Development of ML based prediction models for day 100 TRM; 2) *In-silico* analysis (i.e., performed through a computerized simulation) of the developed models. Factors necessary for optimal prediction were explored: type of model, size of data

Volume: 124  
 Issue: 21  
 Pages: 2568 - 2568  
 DOI: <http://dx.doi.org/>

- |                                    |                                       |
|------------------------------------|---------------------------------------|
| <a href="#">✉ Email</a>            | <a href="#">📄 Download PPT</a>        |
| <a href="#">🔔 Citation Alert</a>   | <a href="#">📁 Save to My Folders</a>  |
| <a href="#">🔔 Correction Alert</a> | <a href="#">© Request Permissions</a> |
| <a href="#">🌐 Citation Tools</a>   | <a href="#">↪ Share</a>               |

- [Article](#)
- [Figures & Data](#)
- [Info & Metrics](#)
- [E-Letters](#)

[Related Articles](#) -

set, number of necessary variables, and performance in specific subpopulations; Model development and analysis were performed with "WEKA" a data mining suite. The area under the receiver operating characteristic curve (AUC) is a commonly used evaluation method for binary choice problems, which involve classifying an instance as either positive or negative. A perfect model will score an AUC of 1, while random guessing will score an AUC of around of 0.5. The AUC was used as measure of predictive performance for the developed models.

Results: We developed six machine learning based prediction models for TRM at day 100. Optimal AUCs ranged from 0.65-0.68. Predictive performance plateaued for a population size ranging from n=5647-8471, depending on the algorithm (Figure 1). A feature selection algorithm ranked variables according to importance. Provided with the ranked variable data, we discovered that a range of 6-12 ranked variables were necessary for optimal prediction, depending on the algorithm (Figure 2). Predictive performance of models developed for specific subpopulations, ranged from an average of 0.59 to 0.67 for patient in second complete remission and patients receiving reduced intensity conditioning respectively.

Conclusions: We present a novel computational approach for prediction model development and analysis in the field of HSCT. Using data commonly collected on transplant patients, our simulation elucidate outcome prediction limiting factors. Regardless of the methodology applied, predictive performance converged when sampling more than 5000 patients. Few variables (approximately 6-12), "carry the weight" with regard to predictive influence. In summary, the presented findings describe a phenomenon of predictive saturation, with data traditionally collected. Improving the current performance will likely require additional types of input like genetic, biologic and procedural factors.

No related articles found.

Articles by [Shouval, R.](#)

Articles by [Nagler, A.](#)

Articles by [Shouval, R.](#)

Articles by [Nagler, A.](#)

Advertisement

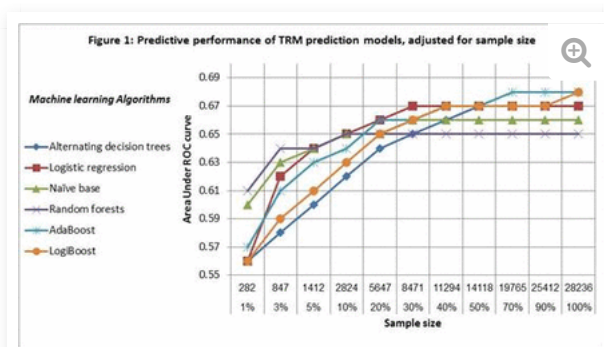


Figure 1 [Download figure](#) | [Open in new tab](#) | [Download powerpoint](#)

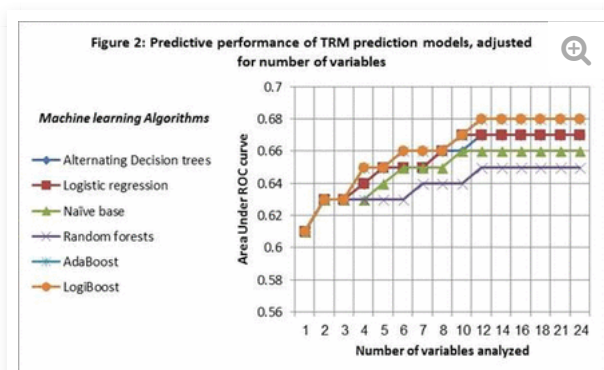


Figure 2

[Download figure](#) | [Open in new tab](#) | [Download powerpoint](#)

**Disclosures** No relevant conflicts of interest to declare.

- \* Asterisk with author names denotes non-ASH members.

© 2014 by The American Society of Hematology

[^ Back to top](#)



How will you treat your patients with the latest available agents?  
LEARN FROM THE EXPERTS.



Advertisement



Leading the way in experimental and clinical research in hematology

**American Society of Hematology**  
2021 L Street NW, Suite 900, Washington, DC 20036  
Phone 202-776-0544 | Fax 202-776-0545

[Current Issue](#)  
[First Edition](#)  
[Topics](#)  
[Collections](#)  
[All Issues](#)  
[Abstracts](#)

[Subscriptions](#)  
[About Blood](#)  
[Newsroom](#)  
[Public Access](#)  
[Permissions](#)  
[Order Reprints](#)

[Submit to Blood](#)  
[Alerts](#)  
[RSS](#)  
[Blood App](#)  
[Contact Us](#)  
[Feedback](#)

**Information for:**

[Authors](#)  
[Subscribers](#)  
[Institutions/Librarians](#)  
[Advertisers](#)  
[ASH Privacy Policy](#)

[ASH Home](#) | [Research](#) | [Education](#) | [Advocacy](#) | [Meetings](#) | [ASH Store](#)



Copyright © 2015 by American Society of Hematology