

Predicting Acute Graft-Versus-Host Disease Using Machine Learning and Longitudinal Vital Sign Data From Electronic Health Records

Shengpu Tang, BSE¹; Grant T. Chappell, BS²; Amanda Mazzoli, BA²; Muneesh Tewari, MD, PhD^{3,6}; Sung Won Choi, MD²; and Jenna Wiens, PhD¹

PURPOSE Acute graft-versus-host disease (aGVHD) remains a significant complication of allogeneic hematopoietic cell transplantation (HCT) and limits its broader application. The ability to predict grade II to IV aGVHD could potentially mitigate morbidity and mortality. To date, researchers have focused on using snapshots of a patient (eg, biomarkers at a single time point) to predict aGVHD onset. We hypothesized that longitudinal data collected and stored in electronic health records (EHRs) could distinguish patients at high risk of developing aGVHD from those at low risk.

PATIENTS AND METHODS The study included a cohort of 324 patients undergoing allogeneic HCT at the University of Michigan C.S. Mott Children's Hospital during 2014 to 2017. Using EHR data, specifically vital sign measurements collected within the first 10 days of transplantation, we built a predictive model using penalized logistic regression for identifying patients at risk for grade II to IV aGVHD. We compared the proposed model with a baseline model trained only on patient and donor characteristics collected at the time of transplantation and performed an analysis of the importance of different input features.

RESULTS The proposed model outperformed the baseline model, with an area under the receiver operating characteristic curve of 0.659 versus 0.512 ($P = .019$). The feature importance analysis showed that the learned model relied most on temperature and systolic blood pressure, and temporal trends (eg, increasing or decreasing) were more important than the average values.

CONCLUSION Leveraging readily available clinical data from EHRs, we developed a machine-learning model for aGVHD prediction in patients undergoing HCT. Continuous monitoring of vital signs, such as temperature, could potentially help clinicians more accurately identify patients at high risk for aGVHD.

JCO Clin Cancer Inform 4:128-135. © 2020 by American Society of Clinical Oncology

Creative Commons Attribution Non-Commercial No Derivatives 4.0 License 

INTRODUCTION

Acute graft-versus-host disease (aGVHD) remains a life-threatening complication of allogeneic hematopoietic cell transplantation (HCT) and is associated with significant morbidity and mortality.^{1,2} To date, our understanding of the pathogenesis of aGVHD has been elucidated by murine studies and important human immunopathogenetic studies,³⁻⁷ which have been particularly useful for investigating new prevention and treatment strategies.⁶ Despite standard prophylaxis, aGVHD still develops in approximately 40% to 60% of transplant recipients.⁸ Numerous research efforts, involving single-center, multicenter, and large registry clinical data, have identified both patient and donor characteristics that increase the risk of aGVHD.^{9,10} Most of these studies have made use of data collected before transplantation. More recently, researchers have identified post-transplantation biologic markers that predict aGVHD before the onset of clinical manifestations

involving the skin, liver, or GI tract.¹¹ Nonetheless, accurately predicting the onset of aGVHD remains a major unmet challenge and barrier to implementing a personalized approach in GVHD management.

To date, most published studies have focused on a snapshot approach based on measurements at a single or few points in time during the peritransplantation period. Meanwhile, researchers have successfully leveraged longitudinal clinical data stored in electronic health records (EHRs) to build patient risk stratification tools for many adverse clinical outcomes.¹²⁻¹⁴ The availability of EHR systems and the power of machine-learning tools present an opportunity for novel GVHD prediction methods.^{15,16} Specifically, we hypothesized that longitudinal data, especially vital sign measurements (eg, body temperature, heart rate, and blood pressure measurements), would show subtle variations resulting from the known effects of immune activation in GVHD on systemic physiology, which could help discriminate

ASSOCIATED CONTENT

Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on January 10, 2020 and published at ascopubs.org/journal/cci on February 21, 2020; DOI <https://doi.org/10.1200/CCI.19.00105>

CONTEXT

Key Objective

To develop and validate a machine-learning model that incorporates longitudinal vital sign data recorded in electronic health records (EHRs) to predict acute graft-versus-host disease (aGVHD) in patients undergoing hematopoietic cell transplantation (HCT).

Knowledge Generated

A model that incorporated vital signs outperformed a similar model that used only pretransplantation information. Feature importance analysis suggested the model relied more on features related to temperature and systolic blood pressure, as well as those features representing longitudinal trends.

Relevance

Leveraging routinely collected longitudinal clinical data recorded in EHRs and machine-learning techniques constitutes a novel strategy for aGVHD prediction. Our approach can be used to inform clinical decisions and disease management of patients undergoing HCT, facilitating targeted pre-emptive strategies in especially high-risk patients and leading to the development of dynamic aGVHD risk scores using longitudinal clinical data.

high-risk from low-risk patients. In this study, using machine-learning techniques, we developed and evaluated an aGVHD risk stratification model using vital sign measurements routinely collected and recorded in EHRs.

PATIENTS AND METHODS

Study Population

A retrospective study was conducted in a cohort of 342 patients undergoing first-time allogeneic HCT admitted to the University of Michigan C.S. Mott Children's Hospital (UM) during 2014 to 2017. This study was approved by the UM Michigan Medicine Institutional Review Board (HUM00123693: Review of the Outcomes of HCT). The time of prediction was set to day 10 post-transplantation to allow for the use of data during the 10-day period; thus, deaths and discharges from the hospital before day 10 post-HCT, in addition to aGVHD diagnoses before day 10, were excluded from the analyses. Patients with any missing demographic or vital sign data were also excluded.

Model Outcome: Clinical Diagnosis of aGVHD

Each patient in the study cohort was evaluated daily by an aGVHD grading team. GVHD adjudication was conducted on a scale of 0 to 4 according to the modified Glucksberg scale.¹⁷ The primary outcome of the study was the incidence of grade II to IV aGVHD by day 100. The inclusion of grade II aGVHD within the range of primary outcomes is consistent with other published studies^{18,19} and results in greater class balance. We considered a binary classification task (ie, labels in {-1, +1}) in which patients who experienced the primary outcome were labeled positive; otherwise, they were labeled negative. Other outcomes potentially competing with aGVHD (eg, deaths resulting from non-GVHD causes) were labeled as negative and not excluded from the cohort, because information about these outcomes would not be available at the time of prediction.

Data Extraction and Feature Engineering

We extracted 2 sets of features for the study population: baseline features and vital sign features.

Baseline features. Baseline features were extracted from the UM Blood and Marrow Transplantation Clinical Research Database. These included patient demographics such as age and sex, as well as conventional risk factors including graft source, HLA match, donor-recipient relation, disease type and status, conditioning regimen, and GVHD prophylaxis regimen. These data were mapped to a 52-dimensional binary feature vector. Categorical variables (eg, sex) were mapped to binary variables (one-hot encoding). Age was discretized using the following bins: (0, 18), (19, 45), and (46, 70), mapping each bin to a binary feature. Final baseline features are provided in the Data Supplement.

Vital sign features. Vital sign features were extracted from the UM EHR system. These data pertained to 6 vital sign measurements: body temperature, heart rate, respiratory rate, diastolic blood pressure, systolic blood pressure, and peripheral capillary oxygen saturation. For each patient, we considered measurements during a 10-day period, beginning at 12:00AM on the day of infusion (day 0) through day 9 post-HCT. We focused on days 0 to 9, the period before the initial discharge after the transplantation, during which we had on average 50 measurements of inpatient vital signs per day (Fig 1). We sought to leverage consistently and frequently recorded vital signs to obtain quantifiable insights into the underlying physiology and its relationship to aGVHD risk. These vital sign measurements were irregularly spaced time series of variable lengths and were mapped to a 600-dimensional binary feature vector representing each patient (Data Supplement). Features captured both daily summary statistics (eg, mean and standard deviation) and trends over time (eg, slope).

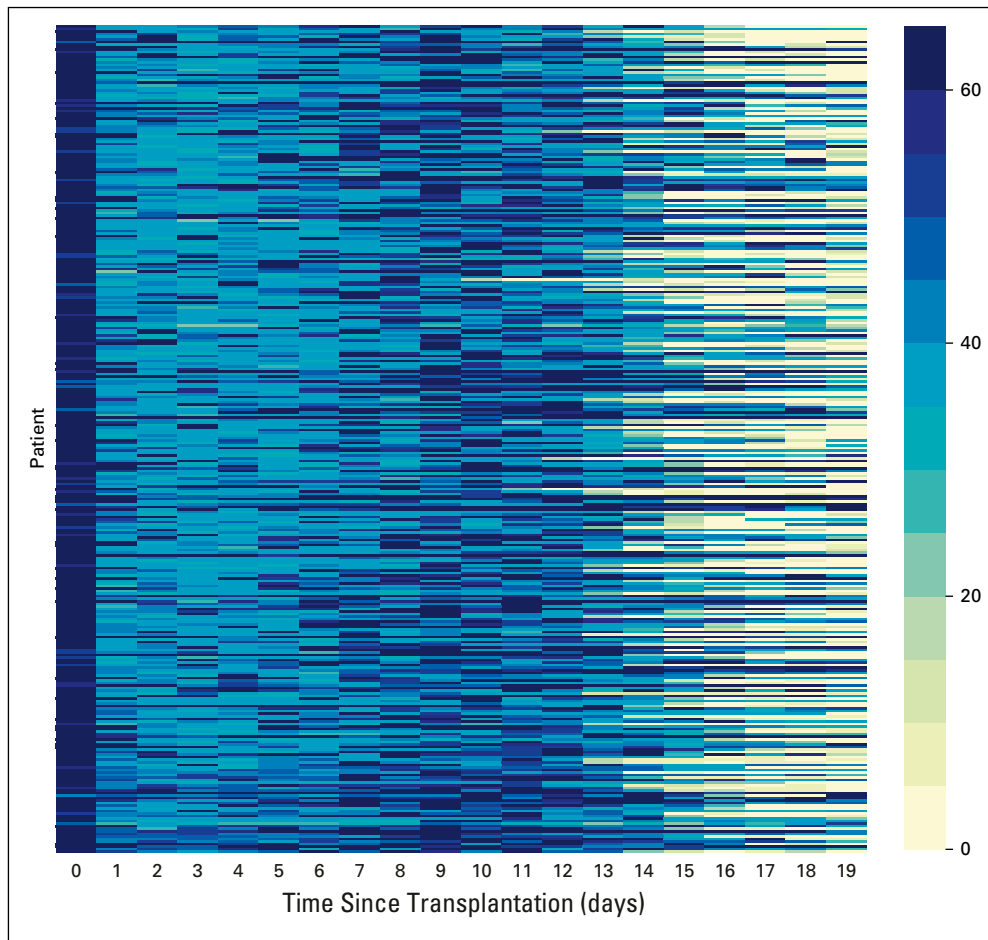


FIG 1. Heatmap of daily measurement frequency of vital signs. Vital sign measurements are routinely collected and recorded in the electronic health record. They include heart rate, respiratory rate, temperature, peripheral capillary oxygen saturation, and systolic and diastolic blood pressures. For a given patient, vital signs were recorded on average 50 times per day during days 0 to 9 and 40 times per day during days 10 to 19. We hypothesize that these data can be used to improve prediction of acute graft-versus-host disease. Each row in the plot represents a patient in the study cohort. For clarity, the darkest color in the heatmap represents a frequency of ≥ 60 .

To incorporate both types of data, we concatenated the baseline features with the vital sign features for each patient, resulting in a 652-dimensional feature vector.

Model Training and Evaluation

To learn and evaluate the risk stratification model, the cohort was split temporally into a training set and a held-out testing set, so that the HCT dates for all patients in the testing set occurred after those of all patients in the training set (Fig 2A). Given the feature vectors and binary labels for each patient in the training data (in years 2014, 2015, and 2016), we learned an L2-regularized logistic regression model to identify patients at greatest risk of developing grade II to IV aGVHD. Model hyperparameters were selected using cross-validation on the training data (Data Supplement). Although deep models specific to time series (eg, LSTM, 1D CNN) may better leverage longitudinal data, they often require large sample sizes. Given our small

sample size, we limited the complexity of our analysis to linear models (Data Supplement provides analyses using random forest, a nonlinear model). In addition, we used L2 regularization over other forms of regularization (eg, L1), because it does not randomly select among correlated features.^{20,21} Applying the model to the held-out testing set from 2017 (the most recent year, in which no relapse occurred before onset), we generated the receiver operating characteristic (ROC) curve and measured discriminative performance using the area under the ROC curve (AUC). Using a decision threshold based on the 55th percentile of the predicted risk scores on the training set (selected based on the incidence rate of aGVHD and to ensure a good sensitivity score), we computed the confusion matrix of the testing set (representing the number of true positives, true negatives, false positives, and false negatives) and reported sensitivity, specificity, and positive predictive value (PPV); 95% CIs were computed using

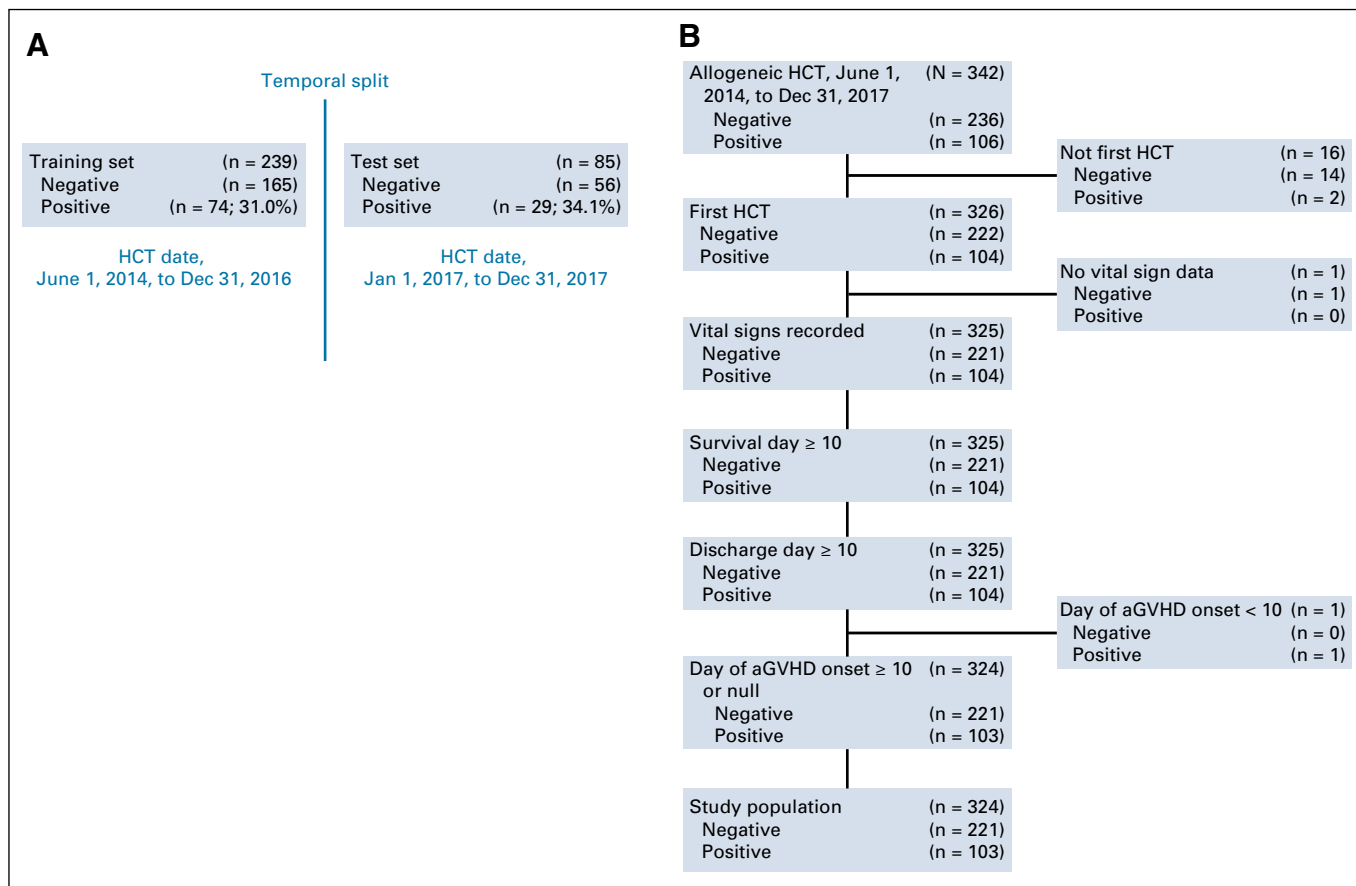


FIG 2. Definition of study cohort and training/test sets. (A) We performed a temporal split to create the held-out test set consisting of patients whose hematopoietic cell transplantation (HCT) date was on or after January 1, 2017. The remainder of the study population (from years 2014, 2015, and 2016) was used for training. (B) Inclusion and exclusion criteria of study population. We excluded patients without vital signs data. Because predictions are made on day 10 post-transplantation, we excluded patients who died, were discharged, or were diagnosed with acute graft-versus-host disease (aGVHD) before day 10.

1,000 bootstrapped samples of the held-out data. We also calculated how far in advance the model could predict high-risk patients, by evaluating the onset time relative to the prediction time (median and interquartile range [IQR]) for true positives.

The discriminative performance of our model that incorporated both baseline features and vital sign features was compared with that of a model using baseline features alone (which is similar to existing approaches that use machine learning^{18,19}). We also compared against a model using only vital sign features. All models were trained and evaluated using the same procedure described in the preceding paragraph; the only difference was the input features. Statistical significance was determined through a Monte-Carlo resampling test²² on the test AUC scores of the 2 models using 1,000 bootstrapped samples, and we reported the one-sided *P* value for whether the proposed model outperformed the baseline model (accordingly, the vitals-only model).

Additionally, we performed several sensitivity analyses to gain insights into how different features were incorporated into the model. To estimate the individual contribution of

each vital sign to model performance, we trained models while excluding each group of vital sign features in turn. To estimate the usefulness of temporal trends versus summary statistics, we trained models while excluding each group of summary features in turn. We estimated the importance of the dropped feature group by observing the decrease in performance (AUC) relative to the original model.²³ Feature groups with a pairwise correlation coefficient of ≥ 0.6 were dropped simultaneously to prevent correlated variables from leaking information. This method effectively measures how much a model relies on the excluded features and thus their importance to the prediction task.

Data Sharing Statement

The data (stripped of any protected health information) and code used in this study will be made available at https://gitlab.eecs.umich.edu/MLD3/JCO_CCI_aGVHD_Prediction. Preprocessing and statistical analyses were performed using Python 3.7²⁴ and Scikit-learn.²⁵ Using this code, other investigators and their institutions can adapt and apply this modeling approach to predict risk of aGVHD in their patients.

TABLE 1. Demographic and Clinical Characteristics (N = 324)

Characteristic	No. (%) of Patients
Recipient age, years	
Median	55
Range	0.3-72
Sex	
Male	190 (59)
Female	134 (41)
Race/ethnicity	
White	296 (91)
African American/black	18 (6)
Asian	2 (1)
Other	6 (2)
Diagnosis	
AML	139 (43)
ALL	52 (16)
Lymphoma (Hodgkin, Non-Hodgkin)	18 (6)
MDS/MPD	64 (20)
Plasma cell (leukemia, myeloma, other)	24 (7)
Other	27 (8)
Disease status	
Low	147 (45)
Intermediate	31 (10)
High	108 (33)
Nonmalignant	38 (12)
Comorbidity index	
Low	58 (18)
Intermediate	98 (30)
High	168 (52)
Conditioning intensity	
Full	284 (88)
Reduced	40 (12)
Donor	
Matched related	118 (36)
Matched unrelated	154 (48)
Mismatched related	13 (4)
Mismatched unrelated	39 (12)
Cell source	
PBSCs	198 (61)
BM	106 (33)
Cord	18 (6)
BM plus cord	2 (1)
CD34 count, 10 ⁶ cells/kg	
Median	5
Range	0.004-480

(Continued in next column)

TABLE 1. Demographic and Clinical Characteristics (N = 324)

Characteristic	No. (%) of Patients
GVHD prophylaxis regimen	
Tacro/MTX	281 (88)
Tacro/MMF	32 (10)
Tacro/MMF/cytoxan	4 (1)
Cyclo/corticosteroid	3 (1)
Cyclo/MTX	3 (1)
Tacro/MTX/ATG	1 (0)

Abbreviations: ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; ATG, antithymocyte globulin; BM, bone marrow; cyclo, cyclosporine; GVHD, graft-versus-host disease; MDS, myelodysplastic syndrome; MMF, mycophenolate; MPD, myeloproliferative disease; MTX, methotrexate; PBSC, peripheral blood stem cell; tacro, tacrolimus.

RESULTS

The study cohort consisted of 324 patients undergoing first-time allogeneic HCT (Fig 2B); 103 (31.8%) had a diagnosis of grade II to IV aGVHD. The onset of positive cases occurred within a median of 36 days (IQR, 24-69 days) from the time of transplantation. Selected demographic and clinical characteristics are listed in Table 1. The proposed model, which incorporated both baseline and vital sign features, modestly but significantly outperformed the baseline model (AUC, 0.659; 95% CI, 0.536 to 0.784 v AUC, 0.512; 95% CI, 0.364 to 0.643; $P = .019$; Fig 3A). The performance of the vitals-only model was similar to that of the proposed model (AUC, 0.634; 95% CI, 0.507 to 0.757; $P > .05$; Fig 3A). At the selected decision threshold, the proposed model achieved a sensitivity of 62.3% (95% CI, 44.8% to 82.8%), specificity of 59.6% (95% CI, 42.9% to 78.6%), and PPV of 39.9% (95% CI, 25.6% to 55.1%; Fig 3B) and identified high-risk patients (true positives) a median of 16 days (IQR, 11-26.2 days) before onset. On the basis of the estimated feature importance, the model relied most on temperature and SBP (Fig 4A), and longitudinal patterns (eg, increasing/decreasing as characterized by positive/negative slopes) were more important than the average values (Fig 4B). A more detailed breakdown of feature importance and follow-up analyses (visualization and clustering) are provided in the Data Supplement.

DISCUSSION

To our knowledge, of the studies investigating predictive modeling in patients undergoing HCT, this is the first approach to use longitudinal data collected over multiple days in the early peritransplantation period. Current research efforts on aGVHD prediction have primarily focused on snapshot approaches involving clinical^{18,19} and biologic¹¹

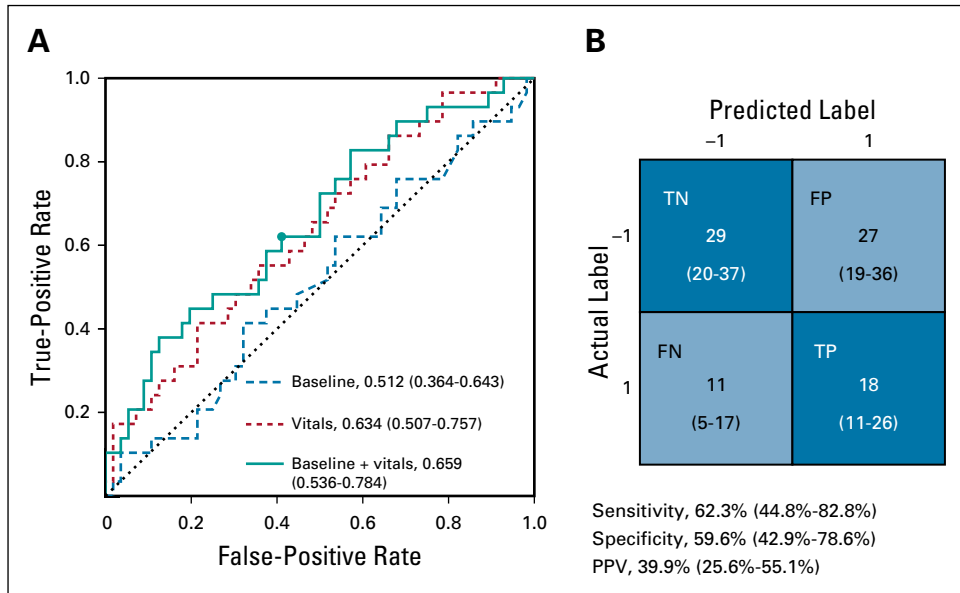


FIG 3. Performance of the model on the 2017 held-out set ($n = 85$), with 95% CIs denoted as ranges in parentheses. (A) Receiver operating characteristic (ROC) curves (with area under the ROC curve [AUC] scores) corresponding to predictive models trained on 239 patients undergoing hematopoietic cell transplantation, using either baseline only, vital sign only, or baseline plus vital sign features. Incorporating vital sign features led to an absolute improvement in AUC score of 0.147 over the baseline model. (B) Confusion matrix with held-out set ($n = 85$) of the proposed model using a risk threshold based on the 55th percentile (marked with filled green circle in panel A). At this risk threshold, the model achieves a sensitivity of 62.3%, specificity of 59.6%, and positive predictive value (PPV) of 39.9%. FN, false negative; FP, false positive; IQR, interquartile range; TN, true negative; TP, true positive.

markers. Rather than considering a single time point, our approach leverages longitudinal data collected over multiple days. By capturing the complex dynamics within vital sign trajectories, we achieved better performance than a model that used only data available at the time of transplantation. Although the baseline model had little predictive advantage, potentially because the baseline

variables only provided weak associations that might not have been detected in the small sample size, the performance of the proposed model was on par with previous studies that also used machine-learning tools, despite our significantly smaller sample size (Arai et al¹⁸ study: $N = 26,695$; AUC, 0.616; Lee et al¹⁹ study: $N = 9,651$; AUC, 0.617). Note that because Lee et al¹⁹ used aGVHD grade III

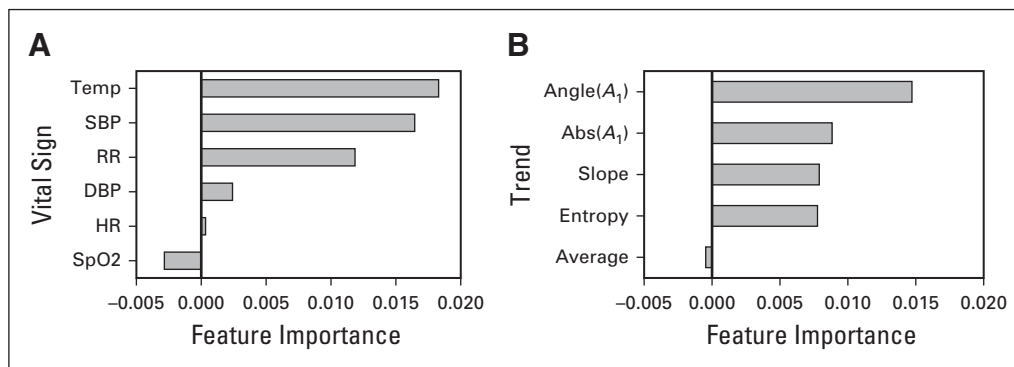


FIG 4. Feature importance of different vital signs and different trend features. The importance of a feature group is defined as the decrease in area under the receiver operating characteristic curve when that group of features is excluded from the model. (A) Among vital signs, temperature (temp) and systolic blood pressure (SBP) are important. (B) Among trends, features pertaining to longitudinal patterns (eg, slope and fast Fourier transform coefficients) are more important than the average values. A_1 denotes the first fast Fourier transform coefficient. DBP, diastolic blood pressure; HR, heart rate; RR, respiratory rate; SpO2, peripheral capillary oxygen saturation.

to IV as the positive label, direct comparisons cannot be made because of differences in the outcome definition (Data Supplement provides an analysis with this alternative outcome). Although the predictions are not yet clinically actionable, and there is clearly room for improvement in the predictive performance, our findings provide a rationale for employing wearable devices in the future to continuously monitor vital signs.

In contrast to prior work, we leveraged data collected as part of routine clinical care. Although there has been immense interest in identifying plasma biomarkers for aGVHD prediction,^{26,27} these measurements are often not routinely obtained in all centers, and efforts to collect and process samples do not scale. EHR data, however, are readily available to most clinicians and researchers. With advances in hospital computing infrastructure through system-wide EHR integration, methods to analyze such data can be extended to more institutions.^{15,28}

Our feature importance analysis showed that certain vital signs, particularly temperature, drove the model predictions of aGVHD. On the basis of the pathophysiology of aGVHD, we hypothesize that cytokines, which are released during the donor immune cell activation characteristic of GVHD, act on the hypothalamus, influencing temperature regulation. Although our approach is based on detecting associations (rather than a causal analysis), these trends and patterns in vital signs could be detected and exploited by our model to identify at-risk patients. These important features differ from prior work that did not consider vital signs collected post-transplantation. Whereas previous studies have found pretransplantation donor and recipient characteristics (ie, baseline features) to be important,¹⁹ we

found a stronger relationship between longitudinal vital signs and outcome, suggesting that going forward, such data should be incorporated when building models of patient risk.

This study is limited in that study data pertained to a single institution, and sample size was small. Moreover, a vast number of other clinical data (eg, medications, laboratory results) were not included, and we did not model interaction effects because of the small sample size and the potentially large feature space. Furthermore, we only considered data collected during the first 10 days post-transplantation. The use of additional data and alternative end points could be considered in future work, especially in reduced-intensity conditioning settings.

In summary, by using readily available clinical data from EHRs, our model provides a novel strategy for aGVHD prediction before the onset of clinical symptoms (eg, rash, vomiting, diarrhea). Although preliminary, our findings highlight the potential use of machine learning and EHR data to predict aGVHD.¹²⁻¹⁴ We focused on a binary classification task with the goal of demonstrating the predictive power of longitudinal vital signs; future work could build upon our study and consider time-to-event analyses. Going forward, additional studies with multicenter collaboration and more registry data may expand our findings to develop a dynamic aGVHD risk score using longitudinal clinical data. Moreover, the proposed approach could be generalized to other major transplantation outcomes (eg, bacterial infections) to enable similar analyses. Such an approach could be used to inform clinical decisions and disease management, facilitating targeted pre-emptive strategies (eg, novel biologics) in especially high-risk patients.

AFFILIATIONS

¹Division of Computer Science and Engineering, Department of Electronic Engineering and Computer Science, University of Michigan, Ann Arbor, MI

²Division of Hematology/Oncology, Department of Pediatrics, University of Michigan, Ann Arbor, MI

³Division of Hematology/Oncology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI

⁴Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI

⁵Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI

⁶Biointerfaces Institute, University of Michigan, Ann Arbor, MI

CORRESPONDING AUTHOR

Jenna Wiens, PhD, 2260 Hayward St, University of Michigan, Ann Arbor, MI 48109; e-mail: wiensj@umich.edu.

EQUAL CONTRIBUTION

S.W.C. and J.W. contributed equally to this study.

SUPPORT

Supported by the Grand Challenge Award from the A. Alfred Taubman Medical Research Institute (M.T., S.W.C., and J.W.).

AUTHOR CONTRIBUTIONS

Conception and design: Muneesh Tewari, Sung Won Choi, Jenna Wiens

Financial support: Muneesh Tewari, Sung Won Choi, Jenna Wiens

Administrative support: Sung Won Choi

Provision of study material or patients: Sung Won Choi

Collection and assembly of data: Grant T. Chappell, Sung Won Choi, Jenna Wiens

Data analysis and interpretation: Shengpu Tang, Amanda Mazzoli, Muneesh Tewari, Sung Won Choi, Jenna Wiens

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Grant T. Chappell**Stock and Other Ownership Interests:** Johnson & Johnson**Muneesh Tewari****Research Funding:** Millennium Pharmaceuticals (Inst), Merck (Inst), Bio-Rad (Inst)**Patents, Royalties, Other Intellectual Property:** Inventor on multiple patent applications related to single-molecule analyte detection technology developed at University of Michigan; the intellectual property has been assigned to University of Michigan, as per university policy for faculty and staff inventors**Travel, Accommodations, Expenses:** Promega**Other Relationship:** Alight Sciences**Jenna Wiens****Research Funding:** Peers Health (Inst)

No other potential conflicts of interest were reported.

ACKNOWLEDGMENT

We thank the patients who participated in the University of Michigan Long-Term Evaluation of the Biology and Outcomes of Hematopoietic Stem Cell Transplantation and Cellular Therapy (HUM43287) protocol.

REFERENCES

1. Zeiser R, Blazar BR: Acute graft-versus-host disease: Biologic process, prevention, and therapy. *N Engl J Med* 377:2167-2179, 2017
2. D'Souza A, Fratham C: Current uses and outcomes of hematopoietic cell transplantation (HCT): CIBMTR summary slides, 2018. <https://www.cibmtr.org/ReferenceCenter/SlidesReports/SummarySlides/pages/index.aspx>.
3. Choi SW, Levine JE, Ferrara JLM: Pathogenesis and management of graft-versus-host disease. *Immunol Allergy Clin North Am* 30:75-101, 2010
4. Shlomchik WD: Graft-versus-host disease. *Nat Rev Immunol* 7:340-352, 2007
5. Ferrara JLM, Levine JE, Reddy P, et al: Graft-versus-host disease. *Lancet* 373:1550-1561, 2009
6. Blazar BR, Murphy WJ, Abedi M: Advances in graft-versus-host disease biology and therapy. *Nat Rev Immunol* 12:443-458, 2012
7. Magenau J, Runaas L, Reddy P: Advances in understanding the pathogenesis of graft-versus-host disease. *Br J Haematol* 173:190-205, 2016
8. Choi SW, Reddy P: Current and emerging strategies for the prevention of graft-versus-host disease. *Nat Rev Clin Oncol* 11:536-547, 2014
9. Harris AC, Ferrara JL, Levine JE: Advances in predicting acute GVHD. *Br J Haematol* 160:288-302, 2013
10. Jagasia M, Arora M, Flowers MED, et al: Risk factors for acute GVHD and survival after hematopoietic cell transplantation. *Blood* 119:296-307, 2012
11. Paczesny S: Discovery and validation of graft-versus-host disease biomarkers. *Blood* 121:585-594, 2013
12. Desautels T, Calvert J, Hoffman J, et al: Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach. *JMIR Med Inform* 4:e28, 2016
13. Oh J, Makar M, Fusco C, et al: A generalizable, data-driven approach to predict daily risk of *Clostridium difficile* infection at two large academic health centers. *Infect Control Hosp Epidemiol* 39:425-433, 2018
14. Zeiberg D, Prahla T, Nallamothu BK, et al: Machine learning for patient risk stratification for acute respiratory distress syndrome. *PLoS One* 14:e0214465, 2019
15. Wiens J, Shenoy ES: Machine learning for healthcare: On the verge of a major shift in healthcare epidemiology. *Clin Infect Dis* 66:149-153, 2018
16. Wiens J, Saria S, Sendak M, et al: Do no harm: A roadmap for responsible machine learning for health care. *Nat Med* 25:1337-1340, 2019 [Erratum: *Nat Med* 25:1627, 2019]
17. Przepiorka D, Weisdorf D, Martin P, et al: 1994 Consensus Conference on Acute GVHD Grading. *Bone Marrow Transplant* 15:825-828, 1995
18. Arai Y, Kondo T, Fuse K, et al: Using a machine learning algorithm to predict acute graft-versus-host disease following allogeneic transplantation. *Blood Adv* 3:3626-3634, 2019
19. Lee C, Haneuse S, Wang HL, et al: Prediction of absolute risk of acute graft-versus-host disease following hematopoietic cell transplantation. *PLoS One* 13:e0190610, 2018
20. Zou H, Hastie T: Regularization and variable selection via the elastic net. *J R Stat Soc B* 67:301-320, 2005
21. Wang J, Oh J, Wang H, et al: Learning credible models. Presented at the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, United Kingdom, August 18-23, 2018
22. LaFleur BJ, Greevy RA: Introduction to permutation and resampling-based hypothesis tests. *J Clin Child Adolesc Psychol* 38:286-294, 2009
23. Breiman L: Random forests. *Mach Learn* 45:5-32, 2001
24. Python: Python Language Reference, version 3.7. <https://docs.python.org/3.7/>
25. Pedregosa F, Varoquaux G, Gramfort A, et al: Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12:2825-2830, 2011
26. Gimondi S, Dugo M, Vendramin A, et al: Circulating miRNA panel for prediction of acute graft-versus-host disease in lymphoma patients undergoing matched unrelated hematopoietic stem cell transplantation. *Exp Hematol* 44:624-634.e1, 2016
27. Budde H, Papert S, Maas JH, et al: Prediction of graft-versus-host disease: A biomarker panel based on lymphocytes and cytokines. *Ann Hematol* 96:1127-1133, 2017
28. Bates DW, Saria S, Ohno-Machado L, et al: Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)* 33:1123-1131, 2014

