Dynamic forecasting of severe acute graft-versus-host disease after transplantation

XUEOU LIU
Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

YIGENG CAO
Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

YE GUO
Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

XIAOWEN GONG
Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

YAHUI FENG
Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

YAO WANG
Yidu Cloud Technology Inc

MINGYANG WANG
Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

MENGXUAN CUI
Yidu Cloud Technology Inc

WENWEN GUO
Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

LUYANG ZHANG
Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

NINGNING ZHAO
Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

XIAOQIANG SONG

Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

XUETONG ZHENG

Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

XIA CHEN

Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

QIUJIN SHEN

Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

SONG ZHANG

Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

ZHEN SONG

Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

LINFENG LI

Yidu Cloud Technology Inc

SIZHOU FENG

Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

MINGZHE HAN

Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

XIAOFAN ZHU

State Key Laboratory of Experimental Hematology, Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Tianjin, 300020, China. https://orcid.org/0000-0002-2572-6495

ERLIE JIANG

Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College

JUNREN CHEN (chenjunren@ihcams.ac.cn)

Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College https://orcid.org/0000-0003-3691-4931

Brief Communication

Keywords: acute graft-versus-host disease, nonparametric approach, dynamic forecasting

Posted Date: November 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1037964/v1

Version of Record: A version of this preprint was published at Nature Computational Science on March 28th, 2022. See the published version at https://doi.org/10.1038/s43588-022-00213-4.

Dynamic forecasting of severe acute graft-versus-host disease after transplantation

Xueou Liu[1,3], Yigeng Cao[1,3], Ye Guo[1,3], Xiaowen Gong[1,3], Yahui Feng[1,3], Yao Wang[2,3], Mingyang Wang[1], Mengxuan Cui[2], Wenwen Guo[1], Luyang Zhang[1], Ningning Zhao[1], Xiaoqiang Song[1], Xuetong Zheng[1], Xia Chen[1], Qiujin Shen[1], Song Zhang[1], Zhen Song[1], Linfeng Li[2], Sizhou Feng[1], Mingzhe Han[1], Xiaofan Zhu[1*], Erlie Jiang[1*], Junren Chen[1*]

[1] State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin, China.

[2] Yidu Cloud Technology Inc., Beijing, China.

[3] These authors contributed equally to this work.

[*] Corresponding authors.

ABSTRACT

To anticipate critical events, clinicians intuitively rely on multidimensional time-series data. It is, however, difficult to model such decision process using machine learning (ML), since real-world medical records often have irregular missing and data sparsity in both feature and longitudinal dimensions. Here we propose a nonparametric approach that updates risk score in real time and can accommodate sampling heterogeneity, using forecasting of severe acute graft-versus-host disease (aGVHD) as the study case. The area under the receiver operator characteristic curve (AUC) rose steadily after transplantation and peaked at >0.7 in both adult and pediatric cohorts. Various numerical experiments provided guidelines for future applications.

MAIN TEXT

Allogeneic hematopoietic stem cell transplantation (allo-HSCT) is a last-resort treatment for many hematological malignancies. Severe aGVHD (grade III–IV) – extensive attack of the skin (**Figure 1a**), gut, and liver of the transplant recipient by the donor's immune cells– remains to this day a leading cause of death after allo-HSCT, with a transplant-related mortality rate as high as ≈30% within 100 days and ≈50% within 3 years.[1] It is desirable to accurately predict severe aGVHD, enabling the medical team to deliver prophylactic immunosuppression specifically to those deemed most likely to benefit from such treatment.[2]

Previously published algorithms for severe aGVHD prognosis were usually based on peri-HSCT 'stationary' (i.e., not time-varying) parameters (including recipient, donor, and transplantation procedural characteristics) or 'landmark' analysis (designating a specific time point post-transplant for biomarker analysis) without modeling multidimensional time-series after HSCT.[3-15] These methods' discrimination capability was limited, especially for patients who develop severe aGVHD later; sizable dynamic clinical data have already accumulated for these patients, and yet much of this new information remains unused. AUCs of models using only stationary parameters were reported to be ≈0.62, even when data from > 20,000 patients were available.[9,12] For landmark analysis, progress has been made on the identification of novel biomarkers.[2-8,14] No biomarker for severe aGVHD, however, is widely adopted

in clinical practice today.[16] Previous studies often relied on biomarkers measured at aGVHD onset for prognosis.[4-6,8] Attempts that tried to use biomarkers measured prior to the appearance of aGVHD signs to forecast severe aGVHD gave conflicting results.[7,10,11,13,14]

ML research on dynamic risk monitoring has been active in intensive care, where blood samples and expert-rated scores are taken frequently and a plethora of devices are connected to the patient.[17,18] HSCT patients, however, have much lower data density and also higher heterogeneity in data collection than patients in intensive care. Limited data capture, non-uniform sampling rates, and data integration issues have all been cited as primary challenges in applying ML in HSCT.[19] One recent study applied penalized logistic regression to vital signs (temperature, heart rate, etc.) that were consistently and frequently recorded within the first 10 days after HSCT[20], and we would like to investigate if we could utilize additional evidence from other dynamic features that were more irregularly measured.

Unlike intensive care[21], currently there is no publicly available multidimensional time-series dataset for HSCT. The Center for International Blood and Marrow Transplant Research and the European Society for Blood and Marrow Transplantation databases collect primarily peri-HSCT stationary parameters, treatment regimes, and treatment outcomes. To close this gap, we compiled and curated post-transplant multidimensional time-series data of HLA- mismatched allo-HSCT patients treated at

the Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College (IHCAMS) (Tianjin, China) between 1 April 2012 to 31 April 2021—hereafter referred to as the 'aGOAT' (aGVHD Onset Anticipation Tianjin) dataset.

aGOAT contained 599 adult and 82 pediatric HLA- mismatched allo-HSCT cases. 12.4% of the adult cohort and 22.0% of the pediatric cohort suffered from severe aGVHD within the first 40 days after transplantation (**Figure 1b and Supplementary Table 1**). aGOAT encompassed a total of 194 dynamic variables for the adult cohort and 159 dynamic variables for the pediatric cohort collected during the first 30 days after transplantation (**Supplementary Table 2**), including vital signs, daily fluid loss (due to diarrhea, vomiting, etc.), complete blood counts (CBC), blood chemistry and electrolytes, blood immune cell profiles (measured by flow cytometry), plasma inflammatory factor levels, etc. The dynamic variables were not measured uniformly across all the patients (**Figure 1b**). Some dynamic variables such as vital signs were available nearly daily, while the others such as blood immune cell profiles and plasma inflammatory factor levels were measured less frequently and not in all patients. In addition, 11 peri-HSCT stationary (i.e., not time-varying) variables were also included in aGOAT (**Supplementary Table 3**); they included information related to primary disease, blood type, stem cell source, conditioning regimen before transplantation, etc.

We then devised a dynamic probabilistic model – 'daGOAT' (dynamic aGVHD Onset

Anticipation Tianjin) – that integrated multidimensional time-series data to calculate risk for severe aGVHD after HLA- mismatched allo-HSCT. Our model updated the risk score $\varphi_i(t)$ daily according to

$$\varphi_i(t) = \sum_{k,t}\big(I_{ikt} \cdot \theta_k(x_{ik}(t), t)\big), \quad (1)$$

where $\theta_k(x_{ik}(t), t)$ was the contribution of 'dynamic' variable $x_{ik}(t)$ to the relative risk of the *i*-th patient developing severe aGVHD. To borrow strength across neighboring time points, the function $\theta_k(\cdot)$ was constrained to be 'smooth' with respect to time *t*.

daGOAT aimed to leverage information from a wide spectrum of clinical variables, even if some of them might be 'spotty'. For instance, even if plasma cytokine data were not available for some patients, the model would still try to infer risk based on other clinical variables. Furthermore, the model updated calculated risk for each patient dynamically, adjusting its assessment whenever new data became available.

To validate our methodology, we compared its performance to the performance of stationary features-only models ('StationaryFeatures'), landmark-specific plasma biomarker levels, and landmark-specific random survival forests models[22] ('LandmarkRSF') using the aGOAT dataset.

daGOAT, StationaryFeatures, and LandmarkRSF were trained and evaluated by 5-fold cross validation (with identical randomization) using all available variables without variable selection. This cross-validation procedure was performed three times for the adult and pediatric cohorts separately and independently to assess the robustness of each modeling approach. In both the adult and pediatric cohorts, daGOAT's discriminative capability reached its peak around the Q1 (25^{th}-percentile) time of severe aGVHD onset, i.e., when ≈75% of the severe aGVHD patients had not yet shown signs of aGVHD. For the adult cohort, AUC increased steadily from 0.65 on day 15 to 0.72 on day 23 (**Fig. 1c**). Performance for the much smaller-sized pediatric cohort was similar, with AUC steadily rising to 0.71 on day 14 (**Fig. 1d**). In both cohorts, daGOAT's peak performance surpassed StationaryFeatures by >16 percentage points. daGOAT outperformed LandmarkRSF by >12 percentage points in the adult cohort (**Fig. 1c**), and its performance was more sustained in time than LandmarkRSF in the pediatric cohort (**Fig. 1d**). We failed to identify statistically significant relationships between severe aGVHD occurrence and plasma levels of suppression of tumorigenicity 2 (ST2), regenerating islet-derived 3-alpha (Reg-3α), and soluble TNF-receptor 1 (sTNFR1) at around day 7 (**Fig. 1e**).

daGOAT was tested for its usage in risk stratification of post-transplant patients: At around the Q1 time of severe aGVHD onset, hazard ratio (HR) between high-risk and low-risk patients identified by the model was 2.045 (95% confidence interval (C.I.): 1.172−3.566) and 2.893 (95% C.I.: 1.057−7.915) for the adult and pediatric cohorts,

respectively, as can be seen by their diverging Kaplan-Meier curves (**Figs. 1f and 1g**).

daGOAT also allowed us to visualize how high-dimensional post-transplant dynamic profile of a patient translated to severe aGVHD risk (**Figs. 2a and 2b**). daGOAT's performance depended on both 'data diversity' (variety) and 'data richness' (quantity). For the adult cohort, it took >100 dynamic variables for the peak AUC (on day 23) to be sustained at >0.7 (**Fig. 2c**), and at least 300 cases' data were needed to train the model properly (**Fig. 2d**). Results of these numerical analysis and experiments shed light on how data should be collected for training and applying the proposed model. In addition, we found that techniques employed in daGOAT such as 'smoothing' and being 'engraftment-aware' (**Methods**) were crucial to the model's stable performance in both the adult and pediatric cohorts (**Fig. 2e**).

Discussion

The comparatively good performance of our modeling approach suggests that it is feasible to reliably and cost-effectively predict severe aGVHD when taking a panoramic and dynamic view of a patient's clinical profile. The average daily cost (charged to the patient) for data collection from day 1 to day 30 post-transplant to support daGOAT was ¥307 per day for one pediatric patient (covering 159 dynamic variables) and ¥425 per day for one adult patient (covering 194 dynamic variables). Despite the large number of dynamic variables included in our model, most of the data were collected in routine clinical care after transplantation and did not incur

additional cost. In contrast, testing a panel of 5-6 plasma biomarkers would cost ¥1000 to ¥1600 extra per sample in China.

Regrettably, due to difficulty in compiling multidimensional post-transplant data from medical records, this study was limited to data from one national hematological center in China, and additional validation at other hospitals will be needed. Deep learning-based risk scoring has been applied to longitudinal data in scenarios where available datasets were orders of magnitude larger than aGOAT (such as predicting respiratory failure in cystic fibrosis[23] and estimating pan-cancer patient survival curves[24]), and it remains to be seen if deep learning can be used to predict severe aGVHD when disparity in data sizes is eventually removed. Regardless of algorithm, the ultimate litmus test of any model would be testing whether we can reduce early mortality after transplantation by applying the model prospectively to administer prophylactic immunosuppression to a targeted subset of HLA-mismatched allo-HSCT patients who are predicted to have high risk for developing severe aGVHD.

As a final note, although this study focused on HSCT, our proposed approach can be generalized to other situations where there is need to integrate non-uniform multidimensional time-series data for dynamic forecasting of adverse events.

METHODS

The aGOAT dataset. We focused on modeling severe aGVHD in HLA-mismatched allo-HSCT, because HLA mismatch is the most important factor associated with aGVHD.[25] Post-transplant multidimensional time-series clinical data of 614 adult patients (age >16) who received HLA-mismatched allo-HSCT between 1 April 2012 and 30 April 2021 and 98 pediatric patients (age $\leq$16) who received HLA-mismatched allo-HSCT between 1 December 2017 and 31 March 2021 at the IHCAMS were able to be electronically retrieved and curated. The adult cohort and the pediatric cohort were treated at different divisions of the IHCAMS.

Outlier values in vital signs (e.g., exorbitant values for body temperature) were made blank. Whenever a dynamic variable was measured more than once on one particular day for one patient, average measurement value of that day was used for that day for that patient. Medical record for each of the cases was reviewed by 2-3 physicians to confirm aGVHD diagnosis and grading (according to the MAGIC criteria[26]). To avoid ambiguity, onset of aGVHD was uniformly defined as the day of initiating aGVHD treatment. After the physicians' review, 20 cases (10 adults and 10 children) were eliminated due to failure of neutrophil engraftment within 30 days of transplantation. Additional 10 cases (4 adults and 6 children) were eliminated, because the recorded date of neutrophil engraftment (defined as 'the date of the first of three consecutive measurements spanning $\geq$3 days of achieving a sustained peripheral blood neutrophil count of $>500\times10^6$/L') did not precede the recorded onset of aGVHD. One additional

adult patient, who had no sign of aGVHD, was also eliminated, because the patient died on day 29 after transplantation, and it was impossible to determine whether the patient would have developed severe aGVHD or not if the patient had survived.

The final dataset, aGOAT, contained 599 adult patients and 82 pediatric patients (**Supplementary Fig. 1 and Supplementary Table 1**). 74 (12.4%) of the adult cohort and 18 (22.0%) of the pediatric cohort suffered from severe aGVHD (grade III–IV) within the first 40 days after transplantation. Eventually 10 (13.5%) of these adult severe aGVHD patients and 2 (11.1%) of these pediatric severe aGVHD patients died within 2 months after aGVHD onset. There was significant difference in 3-year all-cause mortality between the severe aGVHD patients and the other patients in the adult cohort (HR 2.929 (95% C.I.: 1.486–5.774) ($p < 0.0001$, log-rank test)); similar trend appeared to exist in the pediatric cohort also, although it did not pass statistical significance (HR 3.293 (95% C.I.: 0.490–22.140) ($p = 0.096$, log-rank test)) (**Supplementary Fig. 2**). The aGOAT dataset comprised a total of 194 dynamic variables for the adult cohort and 159 dynamic variables for the pediatric cohort (**Supplementary Table 2**). In addition, 11 stationary (i.e., not time-varying) variables were also included in the dataset (**Supplementary Table 3**).

We applied the 'time-limited sample-and-hold' approach commonly used in intensive care unit data analysis[18] to augment the aGOAT dataset (holding time set to 3 days after sampling), based on the hypothesis that most measurements were valid for 3

additional days. This augmented dataset was still very sparse in multiple categories of dynamic variables (**Fig. 1b**). When fitting daGOAT, no other missing-data imputation procedure was conducted to address the problem of non-uniform data measurement. When testing the LandmarkRSF model (using all dynamic variables at one designated time point to fit a random survival forests model[22]), missing data were imputed using the random forest algorithm implemented in the R package 'randomForestSRC'.

The daGOAT model. Our Bayesian-inspired model integrated multidimensional time-series data to calculate risk for severe aGVHD after HLA- mismatched allo-HSCT according to:

$$\varphi_i(t) = \sum_{k,t}\left(I_{ikt} \cdot \theta_k(x_{ik}(t), t)\right),$$

where $x_{ik}(t)$ was the value of the *k*-th "dynamic" variable for the *i*-th patient at time *t*; $I_{ikt} = 0$ when $x_{ik}(t)$ was missing value for the *i*-th patient or (when the model was run in the 'engraftment-aware' mode) before neutrophil engraftment), and $I_{ikt} = 1$ otherwise; $\theta_k(x_{ik}(t), t)$ was the contribution of $x_{ik}(t)$ to the log-odds ratio of the *i*-th patient developing severe aGVHD. $\theta_k(\cdot)$ was constrained to be 'smooth'[27] with respect to time *t*.

We fit daGOAT as follows: First, for every *k* and *t*, we computed the cutoff value c_{kt} that maximized Shannon's mutual information between the *k*-th dynamic variable at

time *t* and severe aGVHD occurrence; then we set $l_k(t)$ and $u_k(t)$ to be the 25th and 75th percentile value among $c_{k,\max\{t_0,\, t-\Delta\tau\}}, \ldots, c_{k,\min\{T,\, t+\Delta\tau\}}$, respectively. (We set $\Delta\tau$ to be ∞.) This step computed the optimal cutoff values $\{l_k(t), u_k(t)\}$ to discretize the *k*-th dynamic variable at time *t*. Second, for every *k* and *t*, we computed

$\rho_{1kt}^{(L)} = P(x_{ik}(t) < l_k(t) \mid i\text{-th patient developed severe aGVHD} \leq 40 \text{ days})$,

$\rho_{1kt}^{(H)} = P(x_{ik}(t) > u_k(t) \mid i\text{-th patient developed severe aGVHD} \leq 40 \text{ days})$,

$\rho_{0kt}^{(L)} = P(x_{ik}(t) < l_k(t) \mid i\text{-th patient did not develop severe aGVHD} \leq 40 \text{ days})$,

and

$\rho_{0kt}^{(H)} = P(x_{ik}(t) > u_k(t) \mid i\text{-th patient did not develop severe aGVHD} \leq 40 \text{ days})$;

then, we computed $\hat{\rho}_{1k}^{(L)}(t)$, $\hat{\rho}_{1k}^{(H)}(t)$, $\hat{\rho}_{0k}^{(L)}(t)$, and $\hat{\rho}_{0k}^{(H)}(t)$ as 'smoothed' versions of $\rho_{1kt}^{(L)}$, $\rho_{1kt}^{(H)}$, $\rho_{0kt}^{(L)}$, and $\rho_{0kt}^{(H)}$, respectively, through smoothing-spline fitting (smooth with respect to *t*). This step computed the discretized probability distribution of the *k*-th dynamic variable that was smooth along the time axis. Finally, we defined

$$\theta_k(x,t) = \begin{cases} \log\left(\frac{\max\left\{0, \hat{\rho}_{1k}^{(L)}(t)\right\}+\gamma}{\max\left\{0,\, \hat{\rho}_{0k}^{(L)}(t)\right\}+\gamma}\right) & \text{if } x < l_k(t) \\ \log\left(\frac{\max\left\{0,\, 1-\hat{\rho}_{1k}^{(L)}(t)-\hat{\rho}_{1k}^{(H)}(t)\right\}+\gamma}{\max\left\{0,\, 1-\hat{\rho}_{0k}^{(L)}(t)-\hat{\rho}_{0k}^{(H)}(t)\right\}+\gamma}\right) & \text{if } l_k(t) \leq x \leq u_k(t), \\ \log\left(\frac{\max\left\{0,\, \hat{\rho}_{1k}^{(H)}(t)\right\}+\gamma}{\max\left\{0,\, \hat{\rho}_{0k}^{(H)}(t)\right\}+\gamma}\right) & \text{if } u_k(t) < x \end{cases}$$

where $\gamma \geq 0$ was a hyperparameter which we set to be 0.1. (When the model was run in the 'no smoothing' mode, $\hat{\rho}_{1k}^{(L)}(t)$, $\hat{\rho}_{1k}^{(H)}(t)$, $\hat{\rho}_{0k}^{(L)}(t)$, and $\hat{\rho}_{0k}^{(H)}(t)$ were not calculated, and $\rho_{1kt}^{(L)}$, $\rho_{1kt}^{(H)}$, $\rho_{0kt}^{(L)}$, and $\rho_{0kt}^{(H)}$ were used instead for calculating $\theta_k(x,t)$.) We set the date t_0 at which log-odds ratio terms started to be cumulated to be day 15 and day 1 post-transplant for adult and pediatric cases, respectively, as the children in the aGOAT dataset tended to have much earlier aGVHD onset than the

adult cases. In addition, when the model was run in the ‘engraftment-aware’ mode, a patient in the validation set did not enter cross-validation until after the patient’s neutrophil engraftment. The pediatric cohort’s data size was too small to support the ‘engraftment-aware’ mode; therefore, daGOAT was set to be not ‘engraftment-aware’ for the children cases.

References

1. Khoury, H.J., *et al.* Improved survival after acute graft-versus-host disease diagnosis in the modern era. *Haematologica* **102**, 958-966 (2017).
2. Paczesny, S. Discovery and validation of graft-versus-host disease biomarkers. *Blood* **121**, 585-594 (2013).
3. Rezvani, A.R., *et al.* Decreased serum albumin as a biomarker for severe acute graft-versus-host disease after reduced-intensity allogeneic hematopoietic cell transplantation. *Biol Blood Marrow Transplant* **17**, 1594-1601 (2011).
4. Vander Lugt, M.T., *et al.* ST2 as a marker for risk of therapy-resistant graft-versus-host disease and death. *N Engl J Med* **369**, 529-539 (2013).
5. McDonald, G.B., *et al.* Plasma biomarkers of acute GVHD and nonrelapse mortality: predictive value of measurements before GVHD onset and treatment. *Blood* **126**, 113-120 (2015).
6. Levine, J.E., *et al.* A prognostic score for acute graft-versus-host disease based on biomarkers: a multicentre study. *Lancet Haematol* **2**, e21-29 (2015).
7. Hartwell, M.J., *et al.* An early-biomarker algorithm predicts lethal graft-versus-host disease and survival. *JCI Insight* **2**, e89798 (2017).
8. Abu Zaid, M., *et al.* Plasma biomarkers of risk for death in a multicenter phase 3 trial with uniform transplant characteristics post-allogeneic HCT. *Blood* **129**, 162-170 (2017).
9. Lee, C., *et al.* Prediction of absolute risk of acute graft-versus-host disease following hematopoietic cell transplantation. *PLoS One* **13**, e0190610 (2018).
10. Zhou, B., *et al.* Prognostic values of increased B7 family proteins in haploidentical hematopoietic stem cell transplantation patients with aGVHD. *Int J Hematol* **109**, 451-462 (2019).
11. Solan, L., *et al.* ST2 and REG3alpha as Predictive Biomarkers After Haploidentical Stem Cell Transplantation Using Post-transplantation High-Dose Cyclophosphamide. *Front Immunol* **10**, 2338 (2019).
12. Arai, Y., *et al.* Using a machine learning algorithm to predict acute graft-versus-host disease following allogeneic transplantation. *Blood Adv* **3**, 3626-3634 (2019).
13. Matsumura, A., *et al.* Predictive Values of Early Suppression of Tumorigenicity 2 for Acute GVHD and Transplant-related Complications after Allogeneic Stem Cell Transplantation: Prospective Observational Study. *Turk J Haematol* **37**, 20-29 (2020).
14. Weissinger, E.M., *et al.* A multicenter prospective, randomized, placebo-controlled phase II/III trial for preemptive acute graft-versus-host disease therapy. *Leukemia* **35**, 1763-1772 (2021).
15. MacMillan, M.L., *et al.* Validation of Minnesota acute graft-versus-host disease Risk Score. *Haematologica* **105**, 519-524 (2020).
16. Zhao, X.S. & Huang, X.J. Seeking biomarkers for acute graft-versus-host disease: where we are and where we are heading? *Biomark Res* **7**, 17 (2019).
17. Henry, K.E., Hager, D.N., Pronovost, P.J. & Saria, S. A targeted real-time early

warning score (TREWScore) for septic shock. *Sci Transl Med* **7**, 299ra122 (2015).
18. Komorowski, M., Celi, L.A., Badawi, O., Gordon, A.C. & Faisal, A.A. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med* **24**, 1716-1720 (2018).
19. Gupta, V., Braun, T.M., Chowdhury, M., Tewari, M. & Choi, S.W. A Systematic Review of Machine Learning Techniques in Hematopoietic Stem Cell Transplantation (HSCT). *Sensors (Basel)* **20**(2020).
20. Tang, S.*, et al.* Predicting Acute Graft-Versus-Host Disease Using Machine Learning and Longitudinal Vital Sign Data From Electronic Health Records. *JCO Clin Cancer Inform* **4**, 128-135 (2020).
21. Johnson, A.E.*, et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* **3**, 160035 (2016).
22. Pickett, K.L., Suresh, K., Campbell, K.R., Davis, S. & Juarez-Colunga, E. Random survival forests for dynamic predictions of a time-to-event outcome using a longitudinal biomarker. *BMC Med Res Methodol* **21**, 216 (2021).
23. Lee, C., Yoon, J. & Schaar, M.V. Dynamic-DeepHit: A Deep Learning Approach for Dynamic Survival Analysis With Competing Risks Based on Longitudinal Data. *IEEE Trans Biomed Eng* **67**, 122-133 (2020).
24. Cheerla, A. & Gevaert, O. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics* **35**, i446-i454 (2019).
25. Kanda, J. Effect of HLA mismatch on acute graft-versus-host disease. *Int J Hematol* **98**, 300-308 (2013).
26. Schoemans, H.M.*, et al.* EBMT-NIH-CIBMTR Task Force position statement on standardized terminology & guidance for graft-versus-host disease assessment. *Bone Marrow Transplant* **53**, 1401-1415 (2018).
27. Friedman, J.H. & Silverman, B.W. Flexible Parsimonious Smoothing and Additive Modeling. *Technometrics* **31**, 3-21 (1989).

Data availability

Upon acceptance of the manuscript for publication, the aGOAT dataset will be deposited at the PRC National Genomics Data Center database (https://ngdc.cncb.ac.cn/) that will be accessible to the research community.

Code availability

R code used in this study is available in Supplementary Information.

Acknowledgements

JR.C. is supported in part by the State Key Laboratory of Experimental Hematology research grant Z20-01. Y.G. is supported in part by the CAMS Innovation Fund for Medical Sciences grant 2020-I2M-C&T-B-089. The authors thank HX. Zhang for assistance in determining neutrophil engraftment dates.

Contributions

JR.C., EL.J., and XF.Z. supervised the study. JR.C., EL.J., and XF.Z. designed the study, with contributions from Y.G. and YG.C.. XO.L. coordinated the study, with contributions from S.Z. and Z.S.. YG.C., Y.G., MY.W., LY.Z., X.C., SZ.F., MZ.H., EL.J., and XF.Z. contributed to data collection. XO.L., YG.C., Y.G., XW.G., YH.F., MY.W., and WW.G. compiled, reviewed, and curated the dataset, with contributions from NN.Z., XQ.S., XT.Z., and X.C.. JR.C. designed the algorithm, with contributions from Y.W., MX.C., and LF.L.. JR.C., Y.W., and MX.C. performed the computation, with contributions from YH.F., XW.G., QJ.S., and XO.L.. J.R.C. wrote

the manuscript, with contributions from Y.W., XO.L., MX.C., XW.G., YG.C., EL.J., and XF.Z..

Corresponding authors

Correspondence to Junren Chen, Erlie Jiang, or Xiaofan Zhu.

Ethics declarations

This retrospective study was initiated in October 2020 and became part of a larger-scope research program, which was approved by the IHCAMS Clinical Research Academic Committee on 11 January 2021 (IIT2021006) and by the IHCAMS Ethics Committee on 7 February 2021 (IIT2021006-EC-1). To avoid biased healthcare or research decisions, patients who received HSCT later than 1 November 2020 (65 adult cases and 15 pediatric cases in the aGOAT dataset) were not included in this study until after 7 February 2021.

388

Fig. 1

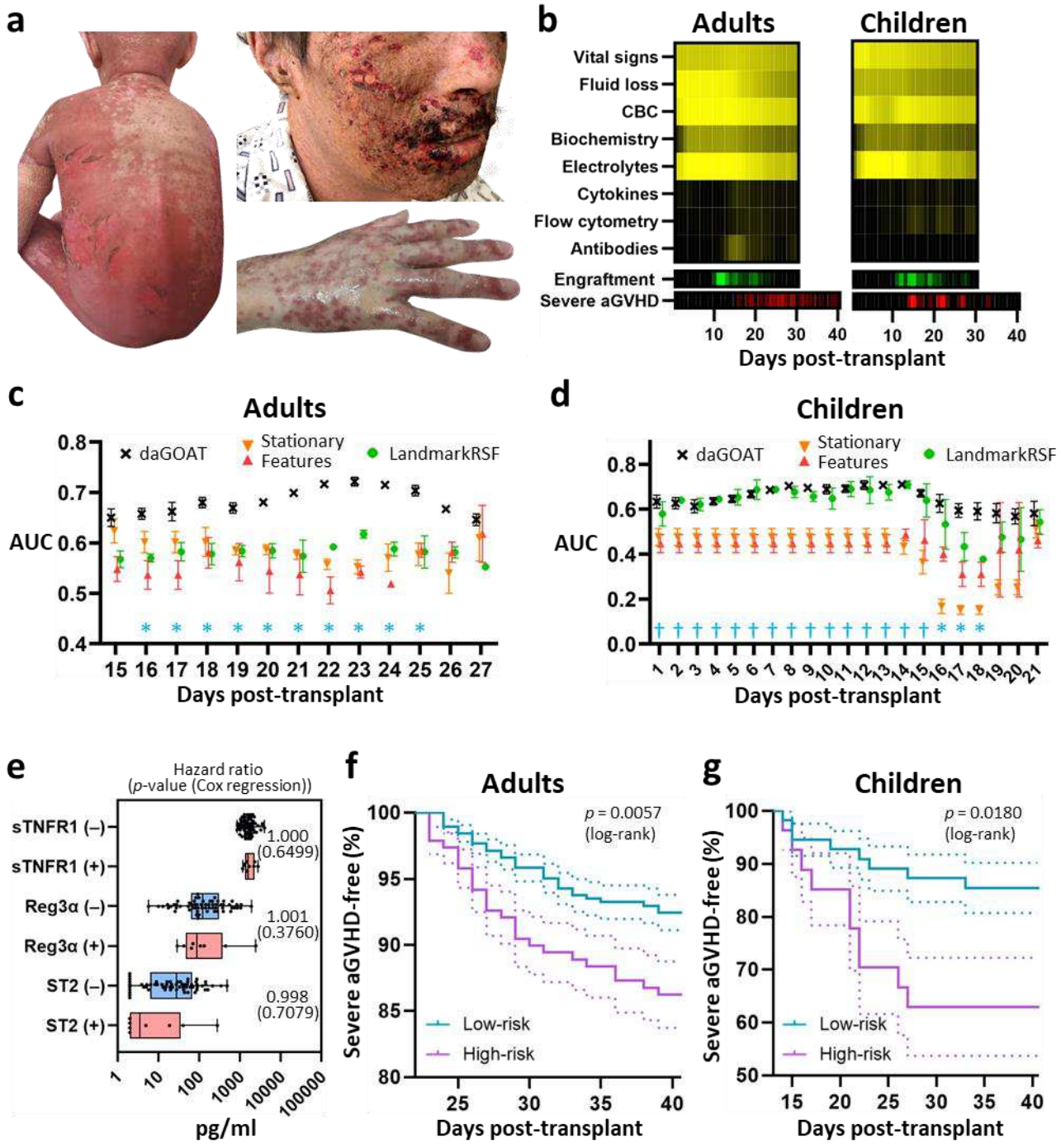
a
b
Adults
Children
Vital signs
Fluid loss
CBC
Biochemistry
Electrolytes
Cytokines
Flow cytometry
Antibodies
Engraftment
Severe aGVHD
10 20 30 40 10 20 30 40
Days post-transplant
c
Adults
daGOAT
Stationary Features
LandmarkRSF
AUC
0.8 0.7 0.6 0.5 0.4
15 16 17 18 19 20 21 22 23 24 25 26 27
Days post-transplant
d
Children
daGOAT
Stationary Features
LandmarkRSF
AUC
0.8 0.6 0.4 0.2 0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Days post-transplant
e
Hazard ratio
(p-value (Cox regression))
sTNFR1 (–)
sTNFR1 (+)
Reg3α (–)
Reg3α (+)
ST2 (–)
ST2 (+)
1.000
(0.6499)
1.001
(0.3760)
0.998
(0.7079)
1 10 100 1000 10000 100000
pg/ml
f
Adults
p = 0.0057
(log-rank)
Severe aGVHD-free (%)
100 95 90 85 80
Low-risk
High-risk
25 30 35 40
Days post-transplant
g
Children
p = 0.0180
(log-rank)
Severe aGVHD-free (%)
100 90 80 70 60 50
Low-risk
High-risk
15 20 25 30 35 40
Days post-transplant

Fig. 1: Dynamic forecasting of severe aGVHD using the proposed algorithm, daGOAT.

a, Skin manifestations of severe aGVHD, a life-threatening complication after transplantation. **b**, Data density and event distributions in the aGOAT dataset. Top: Data density of dynamic variables after transplantation in the adult and pediatric cohorts in the aGOAT dataset (after 'time-limited sample-and-hold' data imputations (**Methods**)). Bottom: Temporal distributions of neutrophil engraftment and severe aGVHD onset. Median onset time of severe aGVHD was day 27 post-transplant (Q1: day 23; Q3: day 31) for the adult cohort and day 21 post-transplant (Q1: day 15; Q3: day 24) for the pediatric cohort, respectively. (Brighter colors in the heat maps indicate higher densities.) **c,d**, Performance of daGOAT in the adult (**c**) and pediatric (**d**) cohorts and comparisons with benchmarks. All models were evaluated by 5-fold cross-validation (with identical randomization for all the models). This procedure was performed three times for the adult and pediatric cohorts separately and independently. In both cohorts, daGOAT was run with 'smoothing'. In the adult cohort, daGOAT was set to be 'engraftment-aware' in addition. The pediatric cohort's data size was too small to run daGOAT in the 'engraftment-aware' mode. Black: daGOAT; Orange and Red: StationaryFeatures, fitted with Naïve Bayes and Random Forest, respectively; Green: LandmarkRSF, fitted using the R package 'randomForestSRC'. (*: daGOAT outperformed all three benchmarks at $p < 0.05$ (paired one-sided t-test). †: daGOAT outperformed both StationaryFeatures models at $p < 0.05$ (paired one-sided t-test).) **e**, Distributions of plasma biomarker levels during

411 days 6–8 post-transplant in the adult cohort. ‘+’ denotes those patients who later
412 developed severe aGVHD, and ‘–’ denotes those who did not. *p*-values were
413 calculated according to Wald test in Cox regression (formula: ‘outcome ~
414 biomarker’). This analysis was not performed in the pediatric cohort due to the very
415 small size of its biomarker data. **f,g**, Risk stratification of adult (**f**) and pediatric (**g**)
416 patients using daGOAT. High-risk: top 1/3 of the patients according to model output
417 at the time of risk stratification (day 23 and day 14 for the adult and pediatric cohorts,
418 respective); Low-risk: bottom 2/3. (Dotted lines: standard errors.)
419

420

Fig. 2

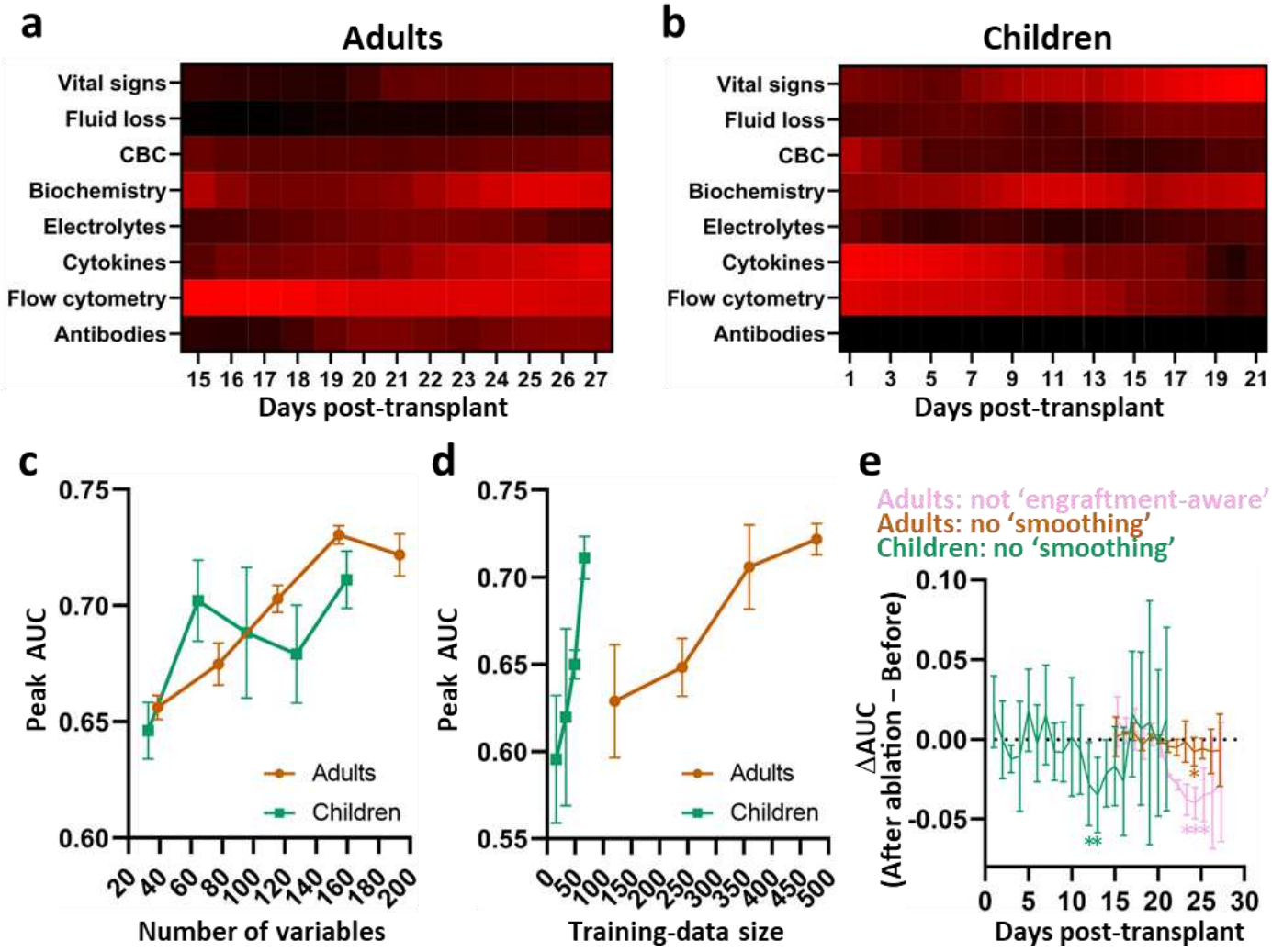
a
Adults
Vital signs
Fluid loss
CBC
Biochemistry
Electrolytes
Cytokines
Flow cytometry
Antibodies
15 16 17 18 19 20 21 22 23 24 25 26 27
Days post-transplant
b
Children
Vital signs
Fluid loss
CBC
Biochemistry
Electrolytes
Cytokines
Flow cytometry
Antibodies
1 3 5 7 9 11 13 15 17 19 21
Days post-transplant
c
Peak AUC
0.75
0.70
0.65
0.60
Adults
Children
20 40 60 80 100 120 140 160 180 200
Number of variables
d
Peak AUC
0.75
0.70
0.65
0.60
0.55
Adults
Children
0 50 100 150 200 250 300 350 400 450 500
Training-data size
e
Adults: not 'engraftment-aware'
Adults: no 'smoothing'
Children: no 'smoothing'
ΔAUC (After ablation – Before)
0.10
0.05
0.00
-0.05
0 5 10 15 20 25 30
Days post-transplant

Fig. 2. Interpretation and characteristics of the daGOAT model.

a,b, Temporal patterns of feature importance identified by daGOAT for the adult (**a**) and pediatric (**b**) cohorts. For each category *C* of dynamic variables (listed in **Supplementary Table 2**), its importance at time *t* was calculated as $\max_{k \in C}\left(\max_{x}\left(\theta_k(x,t)\right) - \min_{x}\left(\theta_k(x,t)\right)\right)$. The brighter a cell is in the heat maps, the more important the corresponding category of dynamic variables at the corresponding time point was for predicting severe aGVHD in the aGOAT dataset. **c**, Relationship between daGOAT's performance and 'data diversity' (variety). Data diversity was measured in the number of variables (randomly selected here) included in model-fitting. **d**, Relationship between daGOAT's performance and 'data richness' (quantity). Data richness was measured in the number of patients (randomly selected here) included in the training set. **e**, Ablation study of the daGOAT model. daGOAT's performance was compared with itself after either 'smoothing' or 'engraftment-aware' was turned off. (*: $p < 0.05$ (paired one-sided t-test).) The pediatric cohort was too small to run daGOAT in the 'engraftment-aware' mode.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- code.txt
- supplementaryinfo.pdf