

## TO THE EDITOR

## The utility of biomarkers in acute GVHD prognostication

Nikolaos Spyrou,<sup>1</sup> Yu Akahoshi,<sup>1</sup> Francis Ayuk,<sup>2</sup> Ernst Holler,<sup>3</sup> Hannah Choe,<sup>4</sup> Aaron Etra,<sup>1</sup> William J. Hogan,<sup>5</sup> Wolf Rösler,<sup>6</sup> Elizabeth Hexner,<sup>7</sup> Zachariah DeFilipp,<sup>8</sup> Ran Reshef,<sup>9</sup> Chantiya Chanswangphuwana,<sup>10</sup> Muna Qayed,<sup>11</sup> Sabrina Kraus,<sup>12</sup> Matthias Eder,<sup>13</sup> Nora Rebeka Javorniczky,<sup>14</sup> Stephan A. Grupp,<sup>15</sup> Carrie L. Kitko,<sup>16</sup> Pietro Merli,<sup>17</sup> Paibel Aguayo-Hiraldo,<sup>18</sup> Matthias Wöfl, <sup>19</sup> Janna Baez,<sup>1</sup> Rahnema Beheshti,<sup>1</sup> Gilbert Eng,<sup>1</sup> Sigrun Gleich,<sup>3</sup> Nikolaos Katsivelos,<sup>1</sup> Alina Khan,<sup>1</sup> Steven Kowalyk,<sup>1</sup> George Morales,<sup>1</sup> Rachel Young,<sup>1</sup> Ryotaro Nakamura,<sup>20</sup> Yi-Bin Chen,<sup>8</sup> John E. Levine,<sup>1,\*</sup> and James L. M. Ferrara<sup>1,\*</sup>

<sup>1</sup>The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY; <sup>2</sup>Department of Stem Cell Transplantation, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; <sup>3</sup>Department of Hematology and Oncology, Internal Medicine III, University of Regensburg, Regensburg, Germany; <sup>4</sup>Division of Hematology, The James Comprehensive Cancer Center, The Ohio State University, Columbus, OH; <sup>5</sup>Division of Hematology, Mayo Clinic, Rochester, MN; <sup>6</sup>Department of Internal Medicine 5, Hematology/Oncology, University Hospital Erlangen, Erlangen, Germany; <sup>7</sup>Blood and Marrow Transplantation Program, Abramson Cancer Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; <sup>8</sup>Hematopoietic Cell Transplant and Cellular Therapy Program, Massachusetts General Hospital, Boston, MA; <sup>9</sup>Blood and Marrow Transplantation Program, Columbia University Medical Center, New York, NY; <sup>10</sup>Department of Medicine, Chulalongkorn University and King Chulalongkorn Memorial Hospital, Bangkok, Thailand; <sup>11</sup>Aflac Cancer and Blood Disorders Center, Children's Healthcare of Atlanta, Atlanta, GA; <sup>12</sup>Department of Internal Medicine II, University Hospital Würzburg, Würzburg, Germany; <sup>13</sup>Department of Hematology, Hemostasis, Oncology, and Stem Cell Transplantation, Hannover Medical School, Hannover, Germany; <sup>14</sup>Department of Medicine I, Medical Center-University of Freiburg, Faculty of Medicine, Albert Ludwigs University, Freiburg, Germany; <sup>15</sup>Division of Oncology, Children's Hospital of Philadelphia and Perelman School of Medicine, Philadelphia, PA; <sup>16</sup>Pediatric Stem Cell Transplant Program, Vanderbilt University Medical Center, Nashville, TN; <sup>17</sup>Department of Hematology/Oncology, Cell and Gene Therapy, Bambino Gesù Children's Hospital, Rome, Italy; <sup>18</sup>Division of Hematology, Oncology and Blood and Marrow Transplantation, Children's Hospital of Los Angeles, Los Angeles, CA; <sup>19</sup>Pediatric Blood and Marrow Transplantation Program, Children's Hospital, University of Würzburg, Würzburg, Germany; and <sup>20</sup>Hematology/Hematopoietic Cell Transplant, City of Hope National Medical Center, Duarte, CA

The clinical severity of acute graft-versus-host disease (GVHD) at its onset is a modest predictor of long-term patient outcomes. Robin et al recently published a single-center study from Hôpital Saint-Louis (HSL) evaluating the utility of blood biomarkers in acute GVHD prognostication.<sup>1</sup> The authors also developed an HSL clinical model to predict outcomes that included 3 clinical variables: liver involvement, age  $\geq 50$  years, and grade 3 or 4 acute GVHD. They then evaluated the value of biomarkers when added to the HSL clinical model using several different techniques, including  $\Delta C$ -index and decision curve analyses (DCAs). The authors concluded that the benefit of the addition of biomarkers to the HSL clinical model was marginal in predicting GVHD outcomes.

We performed analyses, identical to those reported by Robin et al, among 710 patients who received hematopoietic cell transplantation (HCT) between January 2015 and December 2021, with data and samples from the Mount Sinai Acute GVHD International Consortium (MAGIC) database and biorepository (supplemental Table 1; 44 of 710 patients [6%] were included in the original publication).<sup>2</sup> The same eligibility criteria (aged 18-75 years, grade 1-4 acute GVHD systemically treated with at least 1 mg/kg of steroids daily) were used for both analyses (supplemental Table 2). We collected clinical data, measured the serum concentrations of suppression of tumorigenicity 2 (ST2) and regenerating islet-derived 3-alpha (REG3 $\alpha$ ) (referred to as panel 2 by Robin et al<sup>1</sup>) at systemic treatment initiation, and computed the respective MAGIC algorithm probability (MAP) scores, as previously described.<sup>2,3</sup> We followed the same statistical methodology used in the study by Robin et al to compute the C-index,  $\Delta C$ -index with 1000 bootstrap resamples to derive confidence intervals, and DCAs.<sup>1</sup>

Patients who were classified to be at high risk based on the HSL clinical model (liver involvement or age  $\geq 50$  years old, with grade 3-4 acute GVHD [126 of 710 (18%)]) experienced a threefold increase in nonrelapse mortality (NRM) compared with patients at low risk (45% vs 14%;  $P < .001$ ; Figure 1A); these results were in agreement with those reported by Robin et al (Figure 1A as previously

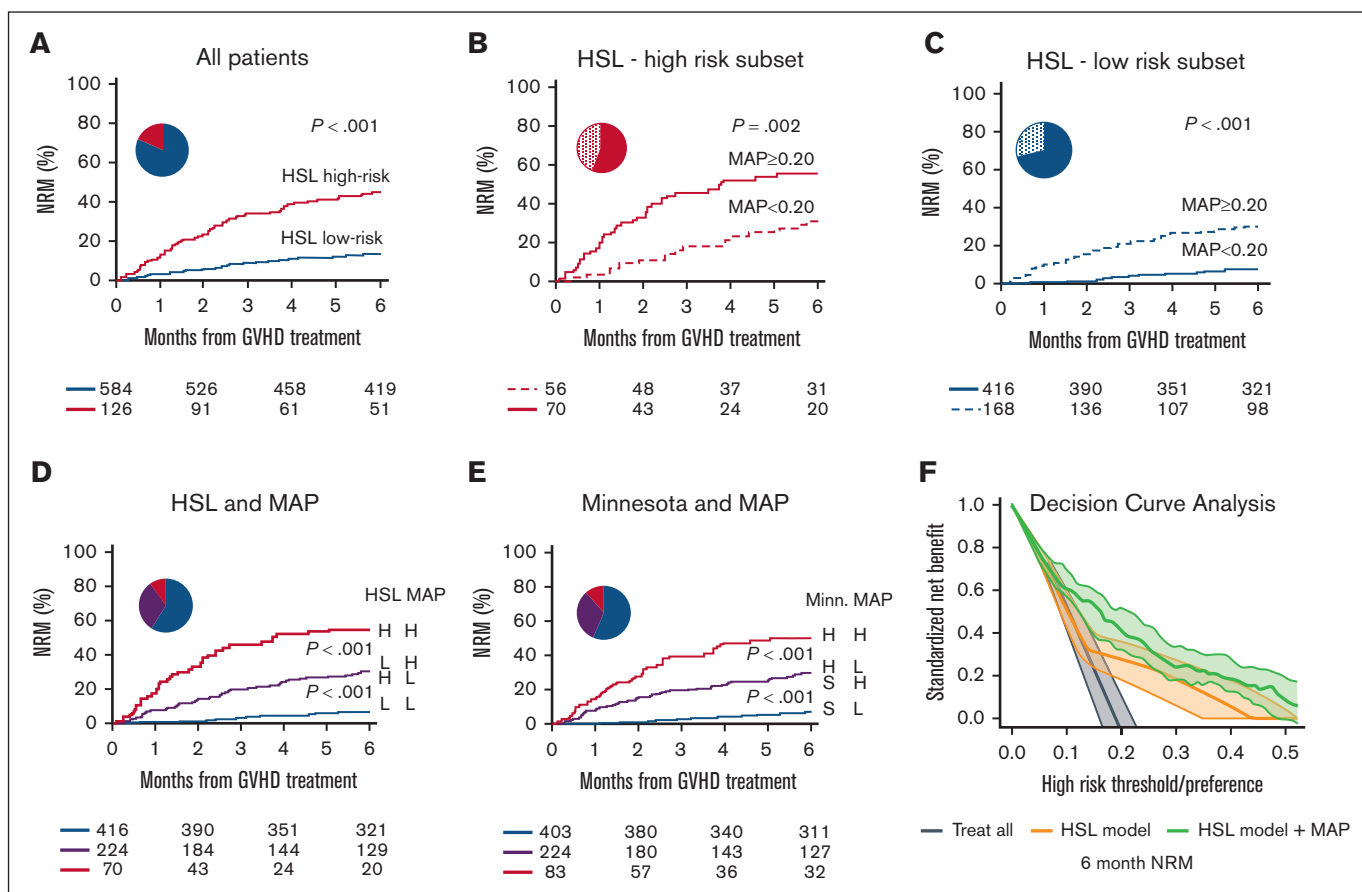
Submitted 6 February 2023; accepted 12 April 2023; prepublished online on *Blood Advances* First Edition 4 May 2023; final version published online 31 August 2023. <https://doi.org/10.1182/bloodadvances.2023009929>.

\*J.E.L. and J.L.M.F. contributed equally to this study.

Data are available on request from the corresponding author, James L.M. Ferrara ([james.ferrara@mssm.edu](mailto:james.ferrara@mssm.edu)).

The full-text version of this article contains a data supplement.

© 2023 by The American Society of Hematology. Licensed under [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International \(CC BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/), permitting only noncommercial, nonderivative use with attribution. All other rights reserved.



**Figure 1. Non-relapse mortality and decision curve analysis of patients stratified by clinical models alone or in combination with biomarkers.** Six-month NRM for (A) MAGIC patients classified to be HSL–high risk (patients with liver involvement or patients  $\geq 50$  years old who also had grade 3–4 acute GVHD) vs HSL–low risk (all other combinations); (B) patients in the HSL–high-risk subgroup classified based on high vs low MAP; (C) patients in the HSL–low-risk subgroup classified based on high vs low MAP; (D) patients classified based on the combination of the HSL model and the MAP (blue, HSL–low-risk and low MAP; purple, HSL–low-risk and high MAP or HSL–high-risk and low MAP; and red, HSL–high-risk and high MAP); (E) patients classified based on the combination of the Minnesota (Minn) risk system and the MAP (blue, Minn standard-risk and low MAP; purple, Minn standard-risk and high MAP or Minn high-risk and low MAP; and red, Minn high-risk and high MAP). (F) DCA for the HSL clinical model (orange) and the combined HSL clinical and MAP (as a continuous variable) model (green) for 6-month NRM.

described<sup>1</sup>). However, the C-index for day 180 NRM of the HSL clinical model was only 0.63 in MAGIC patients, considerably lower than the 0.81 observed in HSL patients (Table 1). The C-index for day 180 NRM of the HSL model was similar to that of the previously validated Minnesota<sup>4</sup> risk system (0.63 vs 0.62;  $\Delta$ C-index, 0.01 [–0.03 to 0.05]; supplemental Table 3).

The proportion of patients with a high MAP and the large differences in day 180 NRM between patients with high MAPs vs low MAPs were similar in both studies (see Figure 1C in the report by Robin et al<sup>1</sup>). When we further stratified MAGIC patients, based on the HSL criteria, into low-risk and high-risk groups using the MAP threshold of 0.20, the MAP further classified both subgroups into populations with significantly different NRM in both high-risk (55% vs 31%;  $P = .002$ ) and low-risk (30% vs 7%;  $P < .001$ ) categories (Figure 1B,C). The combination of the HSL model and the binary MAP score classifies patients into 3 groups of low, intermediate, and high risk with statistically different NRM and responses to primary treatment. Patients categorized as HSL–low risk with a low MAP have only 7% NRM and a high response to steroids (85%); patients categorized as HSL–low risk with a high MAP or patients

categorized as HSL–high risk with a low MAP have intermediate NRM (31%) and treatment response (72%); and patients categorized as HSL high risk with a high MAP experience 55% NRM and only 57% response to treatment (Figure 1D; supplemental Figure 1). We observed a virtually identical pattern producing 3 discrete groups of NRM and treatment response when combining the MAP score with the Minnesota risk system (Figure 1E; supplemental Figures 1 and 2).

In a multivariate regression that incorporated only the HSL clinical model (low- vs high-risk) and the binary MAP (low vs high), a high-risk MAP remained highly significant as a predictor of both NRM (hazard ratio, 4.01 [2.77–5.81];  $P < .001$ ) and overall survival (OS) (hazard ratio, 3.42 [2.48–4.71];  $P < .001$ ) in MAGIC patients (supplemental Figure 3).

We further evaluated the MAGIC patients for any additional value of biomarkers using the HSL clinical criteria. The C-index of the MAP alone (considered as a binary variable to distinguish between patients at high and low risk) was consistent between MAGIC and HSL patients (0.69 vs 0.68). When the HSL clinical criteria and MAP were combined in MAGIC patients, the C-index of the

**Table 1. C-indices of models for the prediction of day 180 NRM and OS**

MAP as a binary variable					
C-index	Model	Day 180 NRM		Day 180 OS	
		MAGIC	Hôpital Saint-Louis	MAGIC	Hôpital Saint-Louis
	Clinical	0.63 (0.58-0.69)	0.81 (0.73-0.86)	0.63 (0.58-0.67)	0.75 (0.68-0.81)
	MAP	0.69 (0.64-0.74)	0.68 (0.59-0.75)	0.67 (0.63-0.72)	0.67 (0.6-0.74)
	Clinical + MAP	0.74 (0.69-0.79)	0.84 (0.73-0.89)	0.71 (0.66-0.76)	0.79 (0.70-0.85)
$\Delta$ C-index	Clinical + MAP*	0.11 (0.05-0.15)	0.03 (−0.002 to 0.07)	0.09 (0.04-0.13)	0.04 (0.006-0.09)
MAP as a continuous variable					
C-index	Model	Day 180 NRM		Day 180 OS	
		MAGIC	Hôpital Saint-Louis	MAGIC	Hôpital Saint-Louis
	Clinical	0.63 (0.58-0.69)	0.81 (0.73-0.86)	0.63 (0.58-0.67)	0.75 (0.68-0.81)
	MAP	0.74 (0.69-0.80)	0.72	0.72 (0.66-0.77)	0.70
	Clinical + MAP	0.76 (0.71-0.82)	0.86†	0.74 (0.68-0.79)	0.82†
$\Delta$ C-index	Clinical + MAP*	0.13 (0.08-0.18)	0.05†	0.11 (0.06-0.15)	0.07†

OS, overall survival.

\*Compared with the clinical model.

†Confidence intervals not reported.

combined model (0.74) was superior to the clinical model alone with a  $\Delta$ C-index of 0.11. In HSL patients, the  $\Delta$ C-index for the combined model was only 0.03 because the HSL clinical criteria alone possessed an unusually high C-index. Use of the MAP as a continuous rather than a binary variable produced a greater  $\Delta$ C-index (0.13) for the combined model (Table 1). We observed an identical pattern when day 180 OS was used as the model end point, with a  $\Delta$ C-index improvement of 0.09 in MAGIC patients compared with a 0.04 improvement in HSL patients.

Recently, in an exploratory study comparing several different biomarker combinations, Etra et al published an algorithm that also combines ST2 and REG3 $\alpha$ <sup>5</sup> and is distinct from the original algorithm of Hartwell et al.<sup>2</sup> The algorithm reported by Etra et al produced a strikingly similar C-index for day 180 NRM compared with the Hartwell algorithm both as a standalone predictor (0.74 vs 0.75) and when combined with the HSL clinical criteria (0.76 for both models; supplemental Table 3). We again observed a large  $\Delta$ C-index for the combined model with Minnesota risk categories using both the Hartwell and Etra algorithms (supplemental Table 3).

Robin et al applied DCAs to the HSL data set to determine whether the addition of biomarkers can aid in clinical management by enhancing the prediction of long-term outcomes.<sup>6</sup> We applied DCAs to the MAGIC data set in the same fashion, comparing the HSL clinical model with the model that combined HSL clinical criteria and the MAP as a continuous variable. The combined model (green) increased the net benefit (ie, correct identification of the patients who will experience nonrelapse death by day 180) over a much wider range of thresholds for the changing immunosuppression (Figure 1F; N.B. states that these thresholds are clinical preferences that are completely unrelated to biomarker thresholds). Indeed, the MAP adds the most net benefit to clinical criteria when the concerns for toxicity from GVHD and treatment are closely balanced. This benefit was also evident when the combined model used a binary MAP classification (supplemental Figure 4). This result is consistent with the large  $\Delta$ C-index of 0.13 between the 2 models (Table 1) and with the creation of 3 rather than 2 distinct

risk groups (Figure 1A,D). Similar results were observed when we compared the Minnesota risk system with the combination of the Minnesota system with the MAP (supplemental Figure 5).

In summary, we have confirmed the observation of Robin et al that the MAP accurately predicts day 180 outcomes of acute GVHD in a large, international, multicenter data set that contained similar proportions of patients categorized as high-risk based on both the clinical criteria and MAP. This independent external validation did not confirm, however, the ability of the HSL clinical criteria that included 3 clinical variables (liver involvement, age  $\geq$ 50 years old, and grade 3-4 acute GVHD) to predict long-term outcomes as accurately as that reported by Robin et al. Rather, the HSL criteria were comparable with the widely adopted Minnesota risk system, in terms of its predictive ability.<sup>4</sup> In our large multicenter cohort, the benefit of adding biomarkers to the HSL clinical model was substantial, validating previous analyses of the utility of the MAP in predicting acute GVHD outcomes.<sup>2,5</sup> The combination of HSL and biomarkers produced 3 distinct risk groups with different NRM and responses to GVHD treatment. Clinicians could potentially use this classification to guide primary treatment: for example, patients at low risk could receive low-dose steroids, patients at intermediate risk could receive high-dose steroids, and patients at high risk could receive a second agent in addition to high-dose steroids. We caution, however, that this approach has not been formally tested in the setting of a clinical trial. But the 3 risk categories are consistent with the results of the DCA, in which the net benefit of the combined model was evident over a wide range of threshold probabilities (eg, as the threshold probability increases, concern for infections from unnecessary immunosuppression for a given patient outweighs concern for incomplete resolution of GVHD).

Several factors may explain the differences between these studies. Firstly, the HSL data set represents the experience of a single institution as opposed to the 21 HCT centers in the MAGIC data set. Secondly, the MAGIC data set is significantly larger than the HSL data set (710 vs 204). Thirdly, the HSL data set is older (2013-2016) and, thus, does not include recent trends in

approaches to GVHD such as prophylaxis using posttransplant cyclophosphamide. We conclude that the MAP does indeed provide significant and useful information regarding acute GVHD outcomes and can, therefore, help guide treatment decisions for patients that develop acute GVHD.

**Acknowledgments:** The authors thank the patients, their families, and the research staff for their participation. This work was supported by the National Institutes of Health, National Cancer Institute grants PO1CA039542 and P30CA196521, the Pediatric Cancer Foundation, and the German Jose Carreras Leukemia Foundation (DJCLS 01 GVHD 2016 and DJCLS 01 GVHD 2020).

**Contribution:** N.S. designed the study, conducted the statistical analysis, and wrote the manuscript; Y.A. collected the clinical data, advised statistical methods, and reviewed and revised the manuscript; F.A., E. Hexner, H.C., A.E., W.J.H., W.R., E. Holler, Z.D., R.R., C.C., M.Q., S.K., M.E., N.R.J., S.G., C.L.K., P.M., P.A.-H., M.W., R.N., and Y.-B.C. collected the clinical data and reviewed and revised the manuscript; J.B., S.G., A.K., N.K., and R.Y. collected and reviewed the clinical data and reviewed and revised the manuscript; G.E. provided computational and programming support and reviewed and revised the manuscript; R.B., S.K., and G.M. performed the laboratory analysis and reviewed and revised the manuscript; J.E.L. and J.L.M.F. organized and designed the study, interpreted the data, advised methods, and wrote and revised the manuscript; and all authors contributed to the writing of the report and approved the final version of the manuscript.

**Conflict-of-interest disclosure:** J.E.L. and J.L.M.F. are listed as coinventors on a GVHD biomarker patent. H.C. is on the advisory board of Incyte and receives research funding from Opna. Y.-B.C. declares consulting for Magenta, Moderna, Equilium, Celularity, Incyte, and Actinium. The remaining authors declare no competing financial interests.

**ORCID profiles:** N.S., 0000-0001-5274-1749; Y.A., 0000-0001-6825-9340; W.J.H., 0000-0002-5841-4105; E.H., 0000-0002-1125-4060; Z.D., 0000-0002-7994-8974; M.Q., 0000-0001-7689-343X; P.M., 0000-0001-6426-4046; P.A.-H., 0000-0002-0196-806X; N.K., 0009-0004-7955-1119; R.N., 0000-0002-9082-0680; Y.-B.C., 0000-0002-9554-1058; J.E.L., 0000-0002-5611-7828.

**Correspondence:** James L.M. Ferrara, The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Pl, New York, NY 10029; email: [james.ferrara@mssm.edu](mailto:james.ferrara@mssm.edu).

## References

1. Robin M, Porcher R, Michonneau D, et al. Prospective external validation of biomarkers to predict acute graft-versus-host disease severity. *Blood Adv*. 2022;6(16):4763-4772.
2. Hartwell MJ, Özbek U, Holler E, et al. An early-biomarker algorithm predicts lethal graft-versus-host disease and survival. *JCI Insight*. 2017; 2(3):e89798.
3. Spyrou N, Levine JE, Ferrara JLM. Acute GVHD: new approaches to clinical trial monitoring. *Best Pract Res Clin Haematol*. 2022;35(4): 101400.
4. MacMillan ML, Robin M, Harris AC, et al. A refined risk score for acute graft-versus-host disease that predicts response to initial therapy, survival, and transplant-related mortality. *Biol Blood Marrow Transplant*. 2015;21(4):761-767.
5. Etra A, Gergoudis S, Morales G, et al. Assessment of systemic and gastrointestinal tissue damage biomarkers for GVHD risk stratification. *Blood Adv*. 2022;6(12):3707-3715.
6. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res*. 2019;3(1):1-18.