Performance of commonly used risk triage tools in predicting clinical deterioration among hospitalized hematopoietic stem cell transplant recipients

Although hospitalized hematopoietic cell transplantation (HCT) recipients frequently experience critical illness, no prediction tools for clinical deterioration have been validated in this population. We examined the discrimination of five commonly-used risk triage tools among hospitalized HCT recipients at an academic hospital (from January 1, 2019 to December 31, 2022): systemic inflammatory response syndrome (SIRS) criteria, the quick sepsis-related organ failure assessment (qSOFA), the Modified and National Early Warning Scores (MEWS, NEWS), and the Epic Deterioration Index (EDI). We calculated each tool's area under the receiver-operating characteristic curve (AUROC) for a composite of ward death, hospice discharge, or intensive-care unit (ICU) transfer. This outcome occurred in 137 of 1,298 hospitalizations (11%). Hospitalization-level discrimination was lowest for SIRS (AUROC 0.62; 95% confidence interval [CI]: 0.57-0.66) and highest for the EDI (AUROC 0.82; 95% CI: 0.78-0.86). However, the EDI's clinical utility may be limited by long lead times and low positive predictive values.

While outcomes for recipients of hematopoietic cell transplantation (HCT) have improved in recent years, these patients remain at high risk of clinical deterioration (i.e., physiologic worsening resulting in critical illness or death) during hospitalization.[1-3] Because many instances of clinical deterioration represent failure to recognize or act upon early indicators of instability, Early Warning Systems (EWS) have been developed to identify patients at risk of worsening before critical events occur, enabling timely intervention.[4] Critically, these tools have not been well-evaluated in larger series of HCT recipients,[5] who often display different physiology than general inpatients (e.g., fevers, cytopenias, immunocompromise) that could limit performance of risk tools including these variables.[5-7] An inaccurate EWS could drive worse patient outcomes through both false-positive (e.g., alert fatigue, potential for inappropriate testing and/or unnecessary interventions with associated risks) and false-negative (i.e., failure to rescue) results. This study therefore evaluated the performance of several commonly-used risk triage tools for predicting clinical deterioration among hospitalized HCT recipients.

This retrospective cohort study was conducted at the Knight Cancer Institute, Oregon Health & Science University's (OHSU's) National Cancer Institute (NCI)-designated Comprehensive Cancer Center. All ward hospitalizations for adult HCT recipients between January 1, 2019 and December 31, 2022 were included. OHSU's institutional review board approved this study (#25188). Due to the retrospective and minimal risk nature of this study, informed consent was not required by the institutional review board. Brenna Park-Egan, MS, and Patrick G. Lyons, MD, MSc, analyzed the data for this study, and all authors had access to the primary data. Some of these results have previously been presented in the form of an abstract.[8] Electronic health record (EHR) data were obtained from OHSU's Research Data Warehouse. From these, several scores were calculated hourly based on vital sign and laboratory data from the wards: SIRS, qSOFA, MEWS and the NEWS (*Online Supplementary Table S1*).[9] Additionally, the Epic Deterioration Index (EDI, a vendor-supplied model predicting deterioration based on vital signs, demographics, nursing assessments, and lab values) was extracted.[10,11] At OHSU, the EDI is passively displayed within the EHR and used actively to trigger rapid response team evaluation. The other tools in this study are not used by any local clinical protocols. The primary outcome was the composite of ward to ICU transfer, discharge to hospice, or death on the wards. Data were censored after the first of these events.

In the primary analysis, each tool's hospitalization-level area under the AUROC was calculated based on the maximum score from ward admission until deterioration or discharge. Secondarily, discrete time survival analysis was used to predict, every 4 hours, whether the composite outcome would occur in the subsequent 24 hours.[12] For each tool, commonly used thresholds were used to determine the cumulative incidence of "positive" scores (i.e., hypothetical alerts), as well as the time between score-positivity and the primary outcome.[9,10] Subgroup analyses compared tool performance between autologous and allogeneic transplant groups, and index and subsequent admissions. Statistical analyses were performed in R version 4.4.0, using packages caret (v.6.0-94) and pROC (v.1.18.5).[14,15]

In total, 1,298 hospitalizations from 800 patients were included: 732 hospitalizations for allogeneic HCT (allo-HCT) recipients and 566 hospitalizations for autologous HCT (auto-HCT) recipients. Of these, 789 encounters were index admissions, and 509 were subsequent hospitalizations. The composite outcome occurred in 137 hospitalizations (11%), 111 of which involved ICU transfers (Table 1). Among the ICU transfers, 49 (44%) ultimately died and three were discharged to hospice (3%). In overall hospitalizations, allogeneic HCT recipients were significantly more likely to experience the primary outcome than recipients of auto-HCT (N=94, 13% *vs.* N=43, 8%; *P*=0.02). This difference was primarily driven by index transplant hospitalizations (allo-HCT N=41, 12%, *vs.* auto-HCT N=26, 8%; *P*=0.003); no significant difference was observed in outcome rates across the two groups in subsequent hospitalizations. For all scores, median values were significantly higher (*P*<0.01) among patients who deteriorated compared to those who did not (Figure 1A). The EDI had the greatest hospitalization-level discrimination (AUROC 0.82; 95% CI: 0.78-0.86), followed by NEWS (AUROC 0.69; 95% CI: 0.64-0.74), MEWS (AUROC 0.66; 95% CI: 0.61-0.70), qSOFA (AUROC 0.64; 95% CI: 0.59-0.69), and SIRS (AUROC 0.62; 95% CI: 0.57-0.66). Tool performance using discrete time survival analysis was significantly higher for NEWS and MEWS (*P*<0.001) while no significant differences were seen in the other tools (EDI AUROC 0.84; 95% CI: 0.83-0.86; NEWS AUROC 0.80; 95% CI: 0.78-.82; MEWS AUROC 0.77; 95% CI: 0.75-0.79; qSOFA AUROC 0.68; 95% CI: 0.67- 0.70; SIRS AUROC 0.67; 95% CI: 0.65-0.69). MEWS and SIRS had higher AUROC among allog-HCT (*P*=0.01 and *P*<0.01, respectively) than auto-HCT recipients, but no significant differences were seen between transplant groups for the other tools. While performance of every risk prediction tool was observed to be higher in overall subsequent admissions than index admissions, none of these differences were statistically significant (Figure 2). Overall, 1,149 (89%) encounters were SIRS-positive, 971 (75%) were NEWS-positive, 506 (39%) were MEWS-positive, and 225 (17%) were qSOFA-positive at some point during ward hospitalization. A total of 638 (49%) encounters reached the EDI's "moderate" threshold at least once, while 52 (4%) reached its "high" threshold. In five encounters, patients deteriorated without reaching a positive threshold for any tool (these involved facilitating urgent dialysis [N=1], hyponatremia management [N=2], and increased neurological monitoring [N=2]). The time between first reaching a score's positive threshold and deterioration was shortest for qSOFA (3 days, range 1-9 days), followed by MEWS (4 days, 1-10 days), the EDI (5 days, 2-11 days), NEWS (9 days, 3-16 days), and SIRS

Table 1. Patient characteristics and outcomes across index and subsequent admissions.

Index admissions			
	Allogeneic hospitalizations N=347	**Autologous hospitalizations N=442**	***P***
Characteristics			
Age, years, median (IQR)	57 (45-65)	62 (52-68)	<0.001
Female, N (%)	161 (46)	169 (38)	0.023
Min. WBC count x10^9/L, median (IQR)	0.12 (0.10-0.15)	0.12 (0.11- 0.16)	0.2
Max temperature, °C, median (IQR)	38.28 (37.78-39.11)	38.39 (37.89-39)	0.6
Outcomes			
Composite outcome, N (%)	41 (12)	26 (5.9)	0.003
ICU transfer, N (%)	39 (11)	25 (5.7)	0.004
Length of stay, days, median (IQR)	23 (20-29)	16 (14-19)	<0.001
100 day all-cause mortality, N (%)	45 (13)	15 (3.4)	<0.001
Subsequent admissions			
	Allogeneic hospitalizations N=385	**Autologous hospitalizations N=124**	***P***
Characteristics			
Age, years, median (IQR)	56 (41-65)	58 (42-67)	0.4
Female, N (%)	180 (47)	50 (40)	0.2
Min. WBC count x10^9/L, median (IQR)	3.39 (1.59-5.62)	2.28 (0.96-3.94)	<0.001
Max. temperature, °C, median (IQR)	37.28 (37.06-38)	37.72 (37.11-38.67)	0.003
Outcomes			
Composite outcome, N (%)	53 (14)	17 (14)	>0.9
ICU transfer, N (%)	33 (8.6)	14 (11)	0.4
Length of stay, days, median (IQR)	6 (4-15)	5 (3-12)	0.017

ICU: intensive-care unit; WBC: white blood cell; IQR: interquartile range; Min.: minimum; Max.: maximum.

(9 days, 4-15 days).

In this 3-year cohort of hospitalized HCT recipients, commonly used risk prediction tools generally had high discrimination for clinical deterioration. However, values for all scores were frequently at or above typical thresholds for clinical alerting; for instance, only one in ten patients would have avoided meeting SIRS criteria during their hospitalization (likely due to high rates of febrile neutropenia and cytokine release syndrome). Further, hypothetical "lead times" between alerts and clinical deterioration were consistently on the order of days rather than hours. Taken together, these observations suggest limited clinical utility for these scores without substantial modification to how and when they are delivered to, and used by, clinicians. For example, temporarily suppressing subsequent alerts after an initial alert could reduce alarm fatigue. False positives would also be reduced with higher thresholds for positivity; such a strategy could either be applied overall (because the HCT population faces higher baseline risk) or dynamically during periods where risk is expected to be higher or lower based on transplant type and timing. Interestingly, while other scores had similar discrimination to what has been reported among general inpatients, the EDI showed greater discrimination than these comparators.[9,10,13] This finding may reflect the fact that HCT recipients are typically selected based on suitable performance status and fewer comorbidities; for example, supplemental oxygen use (1 parameter in

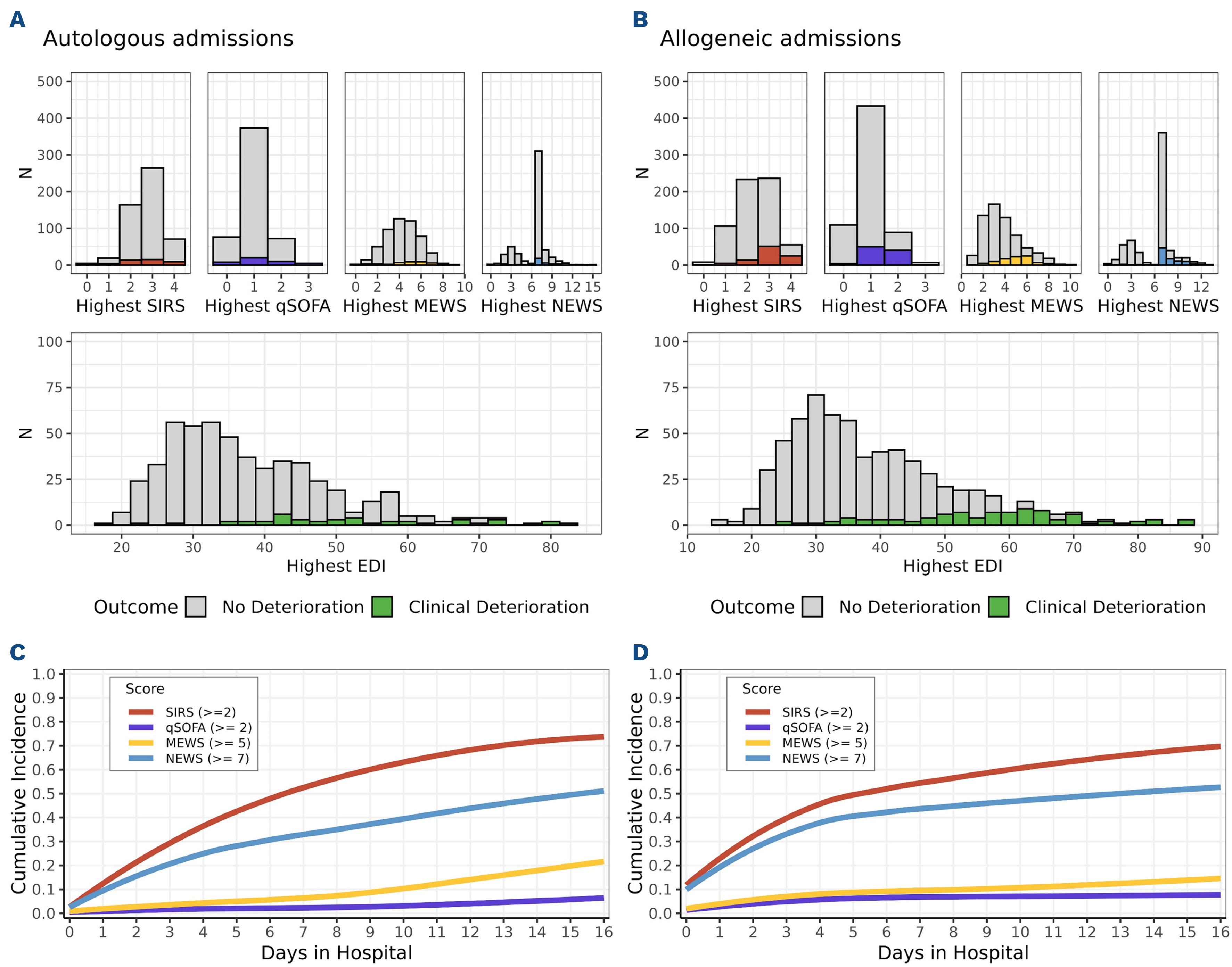

Figure 1. Hospitalization-level score distribution and cumulative incidence of positive scores across autologous and allogeneic admissions. Panels (A) and (B) depict the hospitalization-level highest score distribution of the systemic inflammatory response syndrome (SIRS) criteria, the quick sepsis-related organ failure assessment (qSOFA), the Modified and National Early Warning Scores (MEWS, NEWS), and the Epic Deterioration Index (EDI) across outcome status for autologous and allogeneic hospitalizations, respectively. The x-axis depicts the range of possible scores for each tool, while the y-axis depicts the number of hospital encounters with a maximum of each score. In (C) and (D), SIRS showed the overall highest cumulative incidence of positive scores, followed by NEWS, EDI, MEWS, and qSOFA, for autologous (C) and allogeneic (D) hospitalizations. The x-axis depicts the hospital day and the y axis depicts the proportion of hospitalizations reaching positivity for each score.

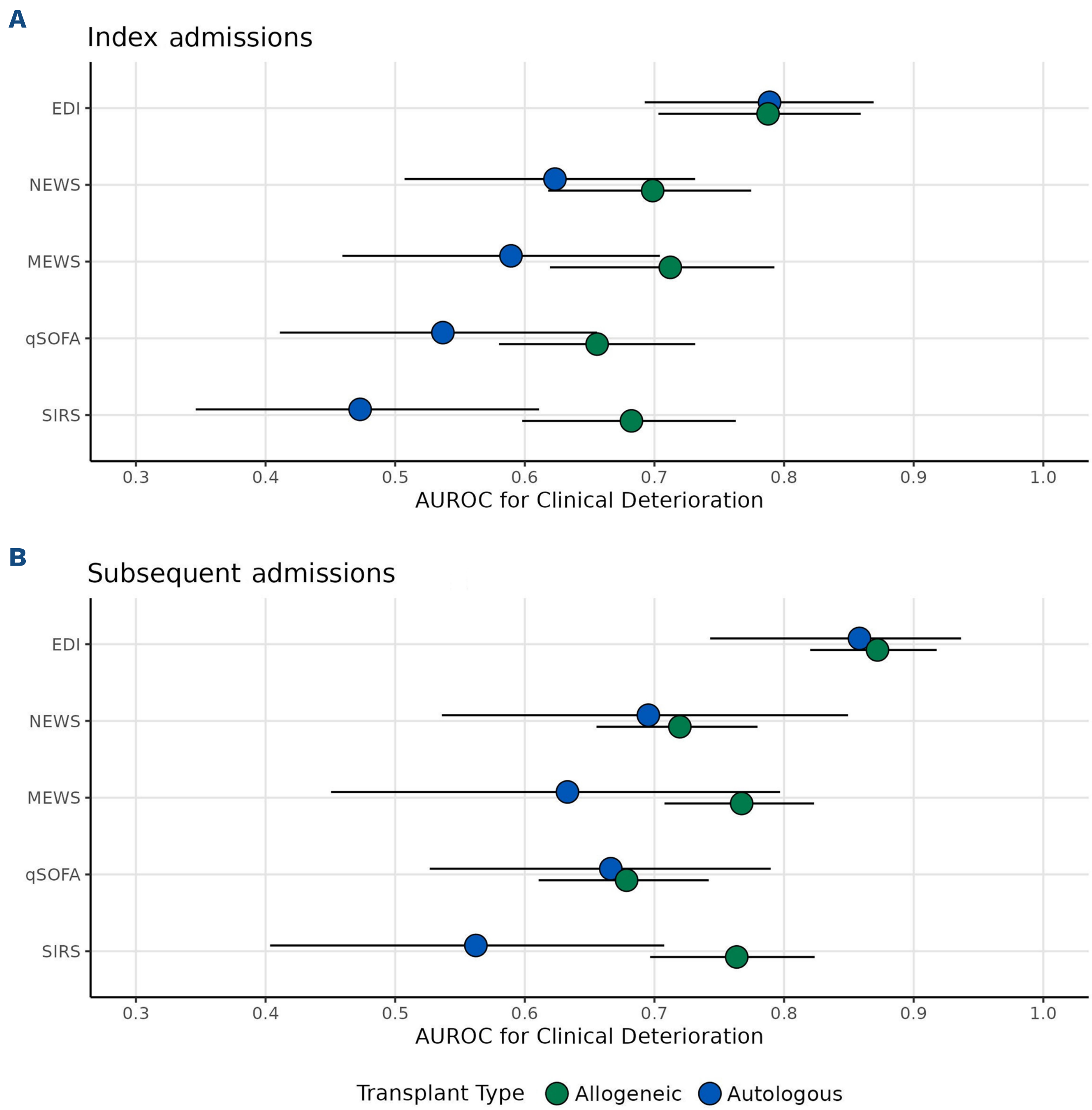

Figure 2. Hospitalization-level Early Warning Scores performance across transplant and admission types. Panels (A) and (B) depict the hospitalization-level performance of the systemic inflammatory response syndrome (SIRS) criteria, the quick sepsis-related organ failure assessment (qSOFA), the Modified and National Early Warning Scores (MEWS, NEWS), and the Epic Deterioration Index (EDI) across outcome status for index and subsequent hospitalizations, respectively, for autologous and allogeneic transplant recipients. The x-axis depicts the area under the receiver operating characteristic for each tool, while the y-axis depicts the selected tool. AUROC: receiver-operating characteristic curve.

the EDI) could be acute or chronic for many patients on the general wards, but is likely representative of acute pathology (and, thus, risk) among HCT recipients. Future work should focus on identifying and validating routinely-available clinical data points where the signal:noise ratio may differ for HCT patients as compared to more general populations; such variables might provide a path to more informative - and therefore more useful - predictive tools.

A limitation of this retrospective study is that some deterioration events may have been prevented or delayed by timely and effective clinical care in response to high EDI scores;[9] our evaluation did not include data on interventions that could have contextualized these results.

Authors

Brenna Park-Egan,[1] Catherine L. Hough,[1] Patrick G. Lyons,[1-3#] and Laura F. Newell[3,4#]

[1]Division of Pulmonary, Allergy, and Critical Care Medicine, OHSU; [2]Department of Medical Informatics and Clinical Epidemiology, OHSU; [3]Knight Cancer Institute, OHSU and [4]Division of Hematology and Medical Oncology, OHSU, Portland, OR, USA

#PGL and LFN contributed equally as senior authors.

Correspondence:
B. PARK-EGAN - parkegan@ohsu.edu

https://doi.org/10.3324/haematol.2025.288892

Received: August 6, 2025.
Accepted: January 8, 2026.
Early view: January 15, 2026.

Disclosures

No conflicts of interest to disclose.

Contributions

PGL and LFN developed the concept and design of the study. BP-E and PGL performed the data cleaning and statistical analysis. All authors contributed to writing the manuscript, with BP-E writing the first draft.

Funding

This study was funded by the KL2TR002370 grant (to PGL).

Data-sharing statement

PGL and LFN developed the concept and design of the study. BP-E and PGL performed the data cleaning and statistical analysis. All authors contributed to writing the manuscript, with BP-E writing the first draft.

References

1. Bayraktar UD, Nates JL. Intensive care outcomes in adult hematopoietic stem cell transplantation patients. World J Clin Oncol. 2016;7(1):98-104.
2. Lyons PG, Klaus J, McEvoy CA, Westervelt P, Gage BF, Kollef MH. Factors associated with clinical deterioration among patients hospitalized on the wards at a tertiary cancer hospital. J Oncol Pract. 2019;15(8):e652-e665.
3. Ji J, Klaus J, Burnham JP, et al. Bloodstream infections and delayed antibiotic coverage are associated with negative hospital outcomes in hematopoietic stem cell transplant recipients. Chest. 2020;158(4):1385-1396.
4. Lyons PG, Edelson DP, Churpek MM. Rapid response systems. Resuscitation. 2018;128:191-197.
5. von Lilienfeld-Toal M, Midgley K, Lieberbach S, et al. Observation-based early warning scores to detect impending critical illness predict in-hospital and overall survival in patients undergoing allogeneic stem cell transplantation. Biol Blood Marrow Transplant. 2007;13(5):568-576.
6. Lyons PG, McEvoy CA, Hayes-Lattin B. Sepsis and acute respiratory failure in patients with cancer: how can we improve care and outcomes even further? Curr Opin Crit Care. 2023;29(5):472-483.
7. O'Mahony M, Wigmore T. Early warning systems and oncological critical care units. In: Nates JL, Price KJ. Oncologic Critical Care: Springer International Publishing; 2019. p. 75-86.
8. Park-Egan B, Hough CT, Newell L, Lyons PG. Discriminatory performance of commonly used risk triage tools in hospitalized recipients of hematopoietic stem cell transplantation. Am J Respir Crit Care Med. 2024;209:A5062.
9. Liu VX, Lu Y, Carey KA, et al. Comparison of early warning scoring systems for hospitalized patients with and without infection at risk for in-hospital mortality and transfer to the intensive care unit. JAMA Netw Open. 2020;3(5):e205191.
10. Singh K, Valley TS, Tang S, et al. Evaluating a widely implemented proprietary deterioration index model among hospitalized patients with COVID-19. Ann Am Thorac Soc. 2021;18(7):1129-1137.
11. Gallo RJ, Shieh L, Smith M, et al. Effectiveness of an artificial intelligence-enabled intervention for detecting clinical deterioration. JAMA Intern Med. 2024;184(5):489-496.
12. Suresh K, Severn C, Ghosh D. Survival prediction models: an introduction to discrete-time modeling. BMC Med Res Methodol. 2022;22(1):207.
13. Churpek MM, Snyder A, Han X, et al. Quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores for detecting clinical deterioration in infected patients outside the intensive care unit. Am J Respir Crit Care Med. 2017;195(7):906-911.
14. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28(5):1-26.
15. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.