Predictive Models for In-Hospital Deterioration in Ward Patients

Gabriel J Escobar, MD[1,2]*, Laura C Myers, MD, MPH[1,2], Vincent X Liu, MD, MS[1,2]

[1]The Permanente Medical Group, Oakland, California; [2] Kaiser Permanente Division of Research, Oakland, California.

Adults admitted to general medical-surgical wards who experience in-hospital deterioration have a disproportionate effect on hospital mortality and length of stay.[1] Not long ago, systematic electronic capture of vital signs—arguably the most important predictors of impending deterioration—was restricted to intensive care units (ICUs). Deployment of comprehensive electronic health records (EHRs) and handheld charting tools have made vital signs data more accessible, expanding the possibilities of early detection.

In this issue, Peelen et al[2] report their scoping review of contemporary EHR-based predictive models for identifying ward patients at risk for deterioration. They identified 22 publications suitable for review. Impressively, some studies report extraordinary statistical performance, with positive predictive values (PPVs) exceeding 50% and with 12- to 24-hour lead times to prepare a clinician response. However, only five algorithms were implemented in an EHR and only three were used clinically. Peelen et al also quantified 48 barriers to and 54 facilitators of the implementation and use of these models. Improved statistical performance (higher PPVs) compared to manually assigned scores were the most important facilitators, while implementation in the context of daily practice (alarm fatigue, integration with existing workflows) were the most important barriers.

These reports invite an obvious question: If the models are this good, why have we not seen more reports of improved patient outcomes? Based on our own recent experience successfully deploying and evaluating the Advance Alert Monitor Program for early detection in a 21-hospital system,[3] we suspect that there are several factors at play. Despite the relative computational ease of developing high-performing predictive models, it can be very challenging to create the right dataset (extracting and formatting data, standardizing variable definitions across different EHR builds). Investigators may also underestimate the difficulty of what can be implemented—and sustained—in real-world clinical practice. We encountered substantial difficulty, for example, around alarm fatigue mitigation and the relationship of alerts to end-of-life decisions. Greater attention to implementation is necessary to advance the field.

We suggest that four critical questions be considered when creating in-hospital predictive models. First, what are the statistical characteristics of a model around the likely clinical decision point? Simply having a high C-statistic is insufficient—what matters is the alert's PPV at a clinically actionable threshold.[4] Second, workflow burden—how many alerts per day at *my* hospital—must be measured, including other processes potentially affected by the new system. Third, will the extra work identify a meaningful proportion of the avoidable bad outcomes? Finally, how will model use affect care of patients near the end of life? Alerts for these patients may not make clinical sense and might even interfere with overall care (eg, by triggering an unwanted ICU transfer).

Implementation requires more than data scientists. Consideration must be given to system governance, predictive model maintenance (models can actually decalibrate over time!), and financing (not just the computation side—someone needs to pay for training clinicians and ensuring proper staffing of the clinical response).

Last, rigorous model evaluation must be undertaken. Given the increasing capabilities of comprehensive EHRs, patient-level randomization is becoming more feasible. But even randomized deployments present challenges. Since ward patients are a heterogeneous population, quantifying process-outcome relationships may be difficult. Alternative approaches to quantification of the impact of bundled interventions may need to be considered—not just for initial deployment, but on an ongoing basis. Peelen et al[2] have effectively summarized the state of published predictive models, which hold the tantalizing possibility of meaningful improvement: saved lives, decreased morbidity. Now, we must work together to address the identified gaps so that, one day, implementation of real-time models is routine, and the promise of in-hospital predictive analytics is fulfilled.

References

1. Escobar GJ, Greene JD, Gardner MN, Marelich GP, Quick B, Kipnis P. Intra-hospital transfers to a higher level of care: contribution to total hospital and intensive care unit (ICU) mortality and length of stay (LOS). *J Hosp Med.* 2011;6(2):74-80. https://doi.org/10.1002/jhm.817
2. Peelen REY, Koeneman M, van de Belt T, van Goor H, Bredie S. Predicting algorithms for clinical deterioration on the general ward. *J Hosp Med.* 2021;16(9):612-619. https://doi.org/10.12788/jhm.3675
3. Escobar GJ, Liu VX, Schuler A, Lawson B, Greene JD, Kipnis P. Automated identification of adults at risk for in-hospital clinical deterioration. *N Engl J Med.* 2020;383(20):1951-1960. https://doi.org/10.1056/NEJMsa2001090
4. Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care.* 2015;19(1):285. https://doi.org/10.1186/s13054-015-0999-1

*Corresponding Author: Gabriel J Escobar, MD;
Email: Gabriel.Escobar@kp.org; Telephone: 510-891-5929.