Alert Timing in Sepsis Prediction Models—An Opportunity to Tailor Interventions

Emily A. Balczewski, BA; Patrick G. Lyons, MD, MSc; Karandeep Singh, MD, MMSc

Given the important role that early treatment plays in mitigating the adverse outcomes of sepsis, clinical decision support (CDS) tools have become an essential component of modern sepsis surveillance programs. Recent advancements in electronic health records have made more sophisticated CDS tools relying on prediction models feasible to use, but their value relative to traditional scoring systems has not been established.

In their cohort study in this issue of JAMA Network Open, Schertz et al compared one such prediction model, the proprietary Sepsis Prediction Model (SPM; Epic Systems), to 3 traditional tools used to predict sepsis onset and severity: the Sequential Organ Failure Assessment (SOFA), quick Sepsis-Related Organ Failure Assessment (qSOFA), and Systemic Inflammatory Response Syndrome (SIRS) criteria. Using more than 60,000 hospital admissions from 5 hospitals in a health system, they evaluated the performance of these tools prior to the clinical recognition of sepsis. Because prediction model performance varies based on the threshold, comparing prediction models to traditional tools requires the selection of model thresholds that allow for apples-to-apples comparisons. For example, at a threshold of 5, the SPM achieves a similar sensitivity (95%) to the SIRS criteria (95%) and SOFA scores (97%) while achieving a much higher specificity (53%) than both the SIRS criteria (42%) and SOFA scores (43%). On the other hand, at a higher threshold of 9, the SPM achieves a sensitivity similar to the qSOFA score (SPM/H11350 9: 82%; qSOFA: 83%) at a higher specificity (SPM/H11350 9: 77%; qSOFA: 69%).

While SPM’s similar sensitivity and higher specificity compared with other tools (at varying thresholds) suggests that the SPM may be superior, Schertz et al found that the higher specificity achieved by the SPM may come at a cost, which is a delay in identifying sepsis. For example, at a threshold of 9, where the SPM has similar sensitivity to the qSOFA score, the SPM would have generated an alert 93.5 minutes after time zero vs 74 minutes for qSOFA. These differences are unlikely to be clinically meaningful or statistically significant given the wide confidence intervals, but the authors’ findings raise an important issue: whether a patient is recognized by a CDS tool should be considered in the context of when that recognition occurs. The timing of recognition is important because clinically effective actions differ at various time points over the course of a sepsis-related hospitalization. Predicting the onset of sepsis before it occurs may be helpful for assigning a patient’s acuity of care. Identifying sepsis within a few hours after onset may be useful for initiation of antibiotics or other components of the sepsis bundle. And identifying sepsis at the end of a hospitalization may be useful for tracking sepsis epidemiology.

At first glance, both the SPM and other scores appear to provide meaningful information depending on the need for either a high sensitivity or high specificity. But are any of them feasible to implement in a clinical workflow? A major emphasis in sepsis care has been placed on identifying, measuring, and minimizing alert fatigue. Using the results from Table 2 in the study by Schertz et al and the 2.7% sepsis prevalence, we can calculate the positive predictive value and the number needed to evaluate (NNE) for each of the sepsis scores, which represents the number of patients who would need to be screened to capture 1 case of sepsis. The NNE is calculated as 1 divided by the positive predictive value. Assuming that each tool were to maximally alert once per patient—the most conservative strategy with respect to alert fatigue—the NNE is 12.4 for the SPM (at the selected threshold of 8), 14.5 for the qSOFA score, 22.2 for the SOFA score, and 23 for the SIRS criteria. Thus, these tools are only feasible to implement if evaluating more than 10 patients per alert is possible, and the timing of the alert is relevant to help correct a deficiency in current practice (eg, poor sepsis...
bundle compliance). If all alerting patients cannot actually be evaluated, the clinical benefits of these tools diminish greatly. Thus, clinicians overseeing the use of these tools must not only focus on selecting the optimal threshold or the best-performing tool. They must also determine whether any tools are sufficiently high performing to be useful at all.

Recently, there are sepsis prediction models whose implementation has led to improvements in clinical outcomes, including at least 1 study involving the SPM. However, performance of the SPM can vary substantially across hospitals (at least in version 1), and we do not know the minimal performance required within a clinical setting before the use of a sepsis tool will translate to clinical benefits. Better clinical outcomes require both effective models and effective interventions, and both may benefit from tailoring to a local context. Through their analysis, Schertz et al have identified a critical aspect of that tailoring: timing of alerts relative to sepsis onset. By estimating when alerts are expected to fire relative to sepsis onset, we can select interventions most likely to be useful at the time of the alert, hopefully aiding the translation of these tools into better clinical outcomes.

REFERENCES