The Need for Improved ICU Severity Scoring

How do we know we're doing a good job taking care of critically ill patients? This question is at the heart of the paper recently published in this journal by Raschke and colleagues (1). Currently, one key method we use to assess the quality of patient care is to calculate the ratio of observed to predicted hospital mortality, or the standardized mortality ratio (SMR). Predicted hospital mortality is estimated with prognostic indices that use patient data to approximate their severity of illness (2). Examples of these indices include the Acute Physiology and Chronic Health Evaluation (APACHE) score, the Simplified Acute Physiology Score (SAPS), the Mortality Prediction Model (MPM), the Multiple Organ Dysfunction Score (MODS), and the Sequential Organ Failure Assessment (SOFA) (3).

Raschke *et al.* (1) evaluated the performance of the APACHE IVa score in subgroups of ICU patients. APACHE is a severity-of-illness score initially created in the 1980s and subsequently updated in 2006 (4,5). This index was developed using data from 110,558 patients from 45 hospitals located throughout the United States, and encompassed 104 intensive care units (ICUs) including mixed medical-surgical, coronary, surgical, cardiothoracic, medical, neurologic, and trauma units. The final model used 142 variables including information from the patient's medical history, the admission diagnosis, and physiologic data obtained during the first day of ICU admission (4). Although it subsequently has been validated using other large general ICU patient cohorts, its accuracy in subgroups of ICU patients is less clear (6).

To benchmark whether the APACHE IVa performed sufficiently, Raschke *et al. (1)* employed an interesting and logical strategy. They created a two-variable severity score (2VSS) to define a lower limit of acceptable performance. As opposed to the 142 variables used in APACHE IVa, the 2VSS used only two variables: patient age and need for mechanical ventilation. They included 66,821 patients in their analysis, encompassing patients from a variety of ICUs located in the southwest United States. The APACHE IVa and 2VSS was calculated for all patients. Although the APACHE IVa outperformed the 2VSS in the general cohort of ICU patients, when patients were divided into subgroups based on admission diagnosis the APACHE IVa showed surprising deficiencies. In patients admitted for coronary artery bypass grafting (CABG), the APACHE IVa did no better in predicting mortality than the 2VSS. The ability of APACHE IVa to predict mortality was significantly reduced in patients admitted for gastrointestinal bleed, sepsis, and respiratory failure as compared to its ability to predict mortality in the general cohort (1).

The work by Raschke *et al*. (1) convincingly shows that APACHE IVa underperforms when evaluating outcomes in subgroups of patients. In some instances, it did no better than a metric that used only two input variables. But why does this matter? One might argue that the APACHE system was not created to function in this capacity. It was designed and validated using aggregate data. It was not designed to determine prognosis on individual-level patients, or even on subsets of patients. However, in real-world practice it *is* used to estimate performance in individual ICUs, which have unique cases mixes of patients that may not approximate the populations used to create and validate APACHE IVa. Indeed, other studies have shown that the APACHE IVa yields different performance assessments in different ICUs depending on varying case mixes (2).

So where do we go from here? The work by Raschke *et al.* (1) is helpful because it offers the 2VSS as an objective method of defining a lower limit of acceptable performance. In the future, more sophisticated and personalized tools will need to be developed to more accurately benchmark ICU quality and performance. Interesting work is being done using local data to customize outcome prediction (7,8). Other researchers have employed machine learning techniques to iteratively improve predictive capabilities of outcome measures (9,10). As with many aspects of modern medicine, the complexity of severity scoring will likely increase as computational methods allow for increased personalization. Given the importance of accurately assessing quality of care, improving severity scoring will be critical to providing optimal patient care.

Sarah K. Medrek, MD
University of New Mexico
Albuquerque, NM USA

References

1. Raschke RA GR, Ramos KS, Fallon M, Curry SC. The explained variance and discriminant accuracy of APACHE IVa severity scoring in specific subgroups of ICU patients. Southwest J Pulm Crit Care. 2018;17:153-64. [CrossRef]
2. Kramer AA, Higgins TL, Zimmerman JE. Comparing observed and predicted mortality among ICUs using different prognostic systems: why do performance assessments differ? Crit Care Med. 2015;43:261-9. [CrossRef] [PubMed]
3. Vincent JL, Moreno R. Clinical review: scoring systems in the critically ill. Crit Care. 2010;14:207. [CrossRef] [PubMed]
4. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med. 2006;34:1297-1310. [CrossRef] [PubMed]
5. Zimmerman JE, Kramer AA, McNair DS, Malila FM, Shaffer VL. Intensive care unit length of stay: Benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV. Crit Care Med. 2006;34:2517-29. [CrossRef] [PubMed]
6. Salluh JI, Soares M. ICU severity of illness scores: APACHE, SAPS and MPM. Curr Opin Crit Care. 2014;20:557-65. [CrossRef] [PubMed]
7. Lee J, Maslove DM. Customization of a Severity of Illness Score Using Local Electronic Medical Record Data. J Intensive Care Med. 2017;32:38-47. [CrossRef] [PubMed]
8. Lee J, Maslove DM, Dubin JA. Personalized mortality prediction driven by electronic medical data and a patient similarity metric. PLoS One. 2015;10:e0127428. [CrossRef] [PubMed]

9. Awad A, Bader-El-Den M, McNicholas J, Briggs J. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. Int J Med Inform. 2017;108:185-95. [CrossRef] [PubMed]
10. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. Lancet Respir Med. 2015;3:42-52. [CrossRef] [PubMed]