Intensive Care Med (2023) 49:223–225
https://doi.org/10.1007/s00134-022-06914-8

LASTING LEGACY IN INTENSIVE CARE MEDICINE

ICU scoring systems

Amanda Quintairos[1,2], David Pilcher[3,4] and Jorge I. F. Salluh[1,5*]

Four decades of prognostic scores in intensive care

Intensive care units (ICUs) are people- and technology-intensive environments where timely and wise use of advanced monitoring and life support is crucial to revert or avoid life-threatening conditions. From their inception, this highly complex environment has been confronted with a need to demonstrate its effectiveness to healthcare stakeholders [1]. In the 1970s and 80s, increasing costs of intensive care, associated with poor outcomes of patients with multi-organ failure, urged intensivists and healthcare managers to look for metrics could concisely express 'severity of illness' and, thus, allow measurement of risk-adjusted outcomes [2].

In the early 1980s, the Acute Physiology and Chronic Health Evaluation (APACHE) system was a milestone in the history of ICU outcome prediction. This scoring system translated domains of pre-morbid conditions (age and co-morbidities), diagnoses and early physiologic derangements (organ failures, laboratory and physiological abnormalities) into a numeric expression of illness severity. In addition to the absolute value of the score, the APACHE system provided an estimate of the risk of death for each individual patient. APACHE was soon followed by the development of the Mortality Prediction Model (MPM) in the United States and the Simplified Acute Physiology Score (SAPS) in Europe. As the technologies improved, new treatments and protocols of care were applied, and the case-mix of the ICU changed (more elderly, co-morbidities and immunocompromised), scores needed to be updated to remain valid predictors of outcomes. The pioneering early versions of APACHE, SAPS and MPM were updated, with SAPS3, APACHE IV, and MPM0-III published, respectively, in 2005, 2006, and 2007 [3]. One of the important differences among these scores relates to the time when they are calculated. SAPS3 and MPM0-III use data from the first hour of ICU admission; whereas, APACHE IV uses the "worst" measurements from the first 24 h. Therefore SAPS3/MPM0-III potentially reflects the early severity of the non-resuscitated patient. APACHE IV provides more time for data collection and less missing data.

Finally, dynamic scores may be applied in the ICU. The most commonly used is the Sequential Organ Failure Assessment (SOFA) which was developed to define the degree of organ failure, to stratify risk particularly in patients with sepsis, and to monitor response to treatment. Scoring systems, although useful for individual risk assessment, are not as applicable for mortality prediction and ICU performance monitoring. They are often used to complement more general mortality scores in ICU.

National and international comparisons: ICU benchmarking

Initially, SAPS3 presented an advantage over APACHE as it was developed in different geographic regions with region-specific equations (supplementary table 1). Geo-economic aspects play a substantial role in case-mix, resource availability, organizational structure, and ultimately outcomes. Many national ICU registries including those in the United Kingdom [4], Italy, Australia, and New Zealand (leveraging previous illness severity scores) developed, updated, and recalibrated their own mortality prediction models to ensure better ICU performance evaluation and benchmarking. While these provided better local benchmarking when compared to currently used prognostic scores [5], international comparisons are still challenging.

Over 40 years, scoring systems have created value for different stakeholders. They are relevant for clinical researchers to assess severity of disease and to interpret the success of interventions; for ICU managers and staff

*Correspondence: jorgesalluh@gmail.com
[1] D'OR Institute for Research and Education, Rua Diniz Cordeiro, 30 – 3º andar, Rio de Janeiro, RJ 22281-100, Brazil
Full author information is available at the end of the article

to benchmark performance, identify outliers and engage in quality improvement; and for healthcare managers and funders, to evaluate performance in intensive care, plan resource allocation, and design performance-based incentives (Fig. 1).

ICU scoring systems have given us a legacy which goes beyond the scores themselves. As sophisticated data science techniques, machine learning, and artificial intelligence bring more predictive capacity and algorithms into healthcare, it is important to remember these lessons learnt over time [6]. Application of a score or prediction to an individual may perpetuate biases inherent in the score's development and can potentially influence perverse behavior. Illness severity scores were initially developed for individual patient application, but clinicians rapidly learnt that they were best used to compare and interpret the risk-adjusted outcomes of patient groups [7].

Applications of illness severity scores have also evolved over time. The SOFA score has been applied to the definition of sepsis [8]. The use of its 'derivative' the Quick-SOFA has been suggested as a screening tool to identify patients at risk of sepsis. Severity of illness scores have been combined with measures of resource utilization, to provide robust and reproducible metrics of ICU efficiency using Rapoport-Teres plots. Techniques used to develop and validate illness severity scoring systems, have led to models to predict length of stay [9], readmission to ICU [10], and the development of complications within the ICU such as pressure injuries [11].

Both SAPS3 and APACHE IV were published more than 15 years ago, it is natural that their calibration would worsen over time. Although scores provide a fixed point of reference, their derived estimates of mortality require iterative updating. Scores developed in general populations may also be inaccurate when applied to specific groups such as transplant or neurocritical patients and more recently coronavirus disease 2019 (COVID-19) [12]. The COVID-19 pandemic has also highlighted the need for global comparisons of disease burden and outcomes. However, there are also potential risks in developing scores for a new condition where outcomes have been heavily influenced by strain placed on the local healthcare system.

Importantly, the development of illness severity scores would not have been possible without collaboration between individuals prepared to collate data from multiple sources, institutions and jurisdictions. Many countries including Brazil, Argentina, Uruguay, and regions of Africa and Asia now have well-established ICU registries which allow for comparison of ICUs both within and between countries, allowing better understanding of

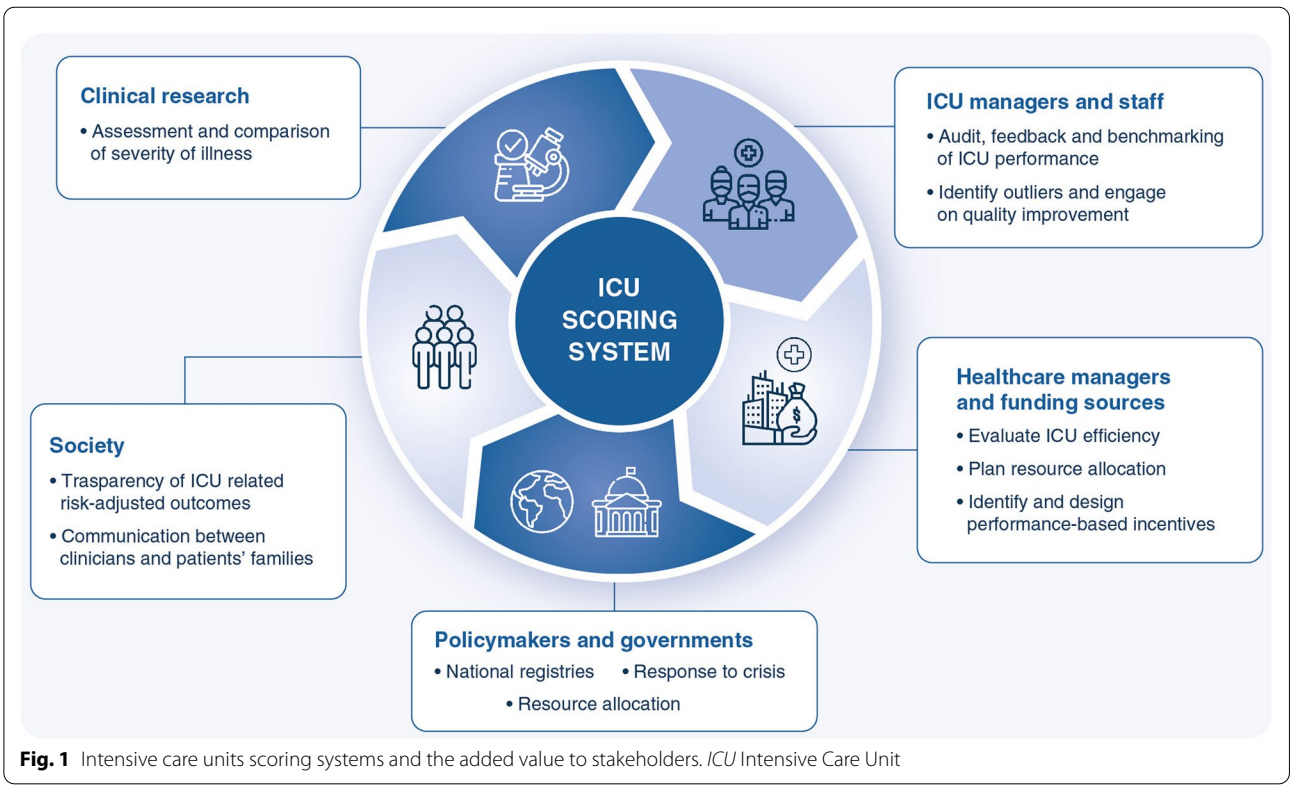

Fig. 1 Intensive care units scoring systems and the added value to stakeholders. *ICU* Intensive Care Unit

differences and common aspects between health systems, access to information and limitations [13, 14].

A global ICU scoring system?

Initiatives such as the Linking of Global Intensive Care (LOGIC) consortium of national ICU Registries and The Global Open Source Severity of Illness Score (GOSSIS) are currently working on effective ways to provide international comparisons of ICUs by using both existing scoring systems and creating new ones [13, 15]. These novel international scoring systems should be more accessible, with heterogeneous case-mix representative of distinct health systems. This type of collaboration should improve benchmarking ICU practice worldwide. Yet, there is still a long way to go until a single ICU scoring system meets all the necessary requisites (supplementary Fig. 1).

Take-home message

After more than 40 years, ICU scoring systems remain the mainstem of ICU performance evaluation. However, their role has become increasingly important beyond the evaluation of an individual or a single ICU, to being more relevant to a wider variety of healthcare stakeholders. Combining our experience of both their limitations and strengths with the rise of technology, big data, and machine learning techniques, we may soon see major changes and improvements that will ultimately allow broader and accurate implementation as well as international comparisons among ICUs.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s00134-022-06914-8.

Author details

[1] D'OR Institute for Research and Education, Rua Diniz Cordeiro, 30 – 3º andar, Rio de Janeiro, RJ 22281-100, Brazil. [2] Department of Critical and Intensive Care Medicine, Academic Hospital Fundación Santa Fe de Bogota, Bogota, Colombia. [3] Department of Intensive Care, Alfred Health, Commercial Road, Prahran, VIC 3004, Australia. [4] The Australian and New Zealand Intensive Care Society (ANZICS) Centre for Outcome and Resource Evaluation, Camberwell, Australia. [5] Post-Graduation Program, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

Author contributions

All authors designed, drafted, and revised the present manuscript.

Funding

JIFS is supported in part by individual research grants from CNPq and FAPERJ. He is a shareholder of Epimed Solutions, a cloud-based electronic system of quality improvement.

Declarations

Conflicts of interest

The other authors declare that they have no conflicts of interest.

Publisher's Note

Received: 17 August 2022 Accepted: 11 October 2022
Published: 31 October 2022

References

1. Power GS, Harrison DA (2014) Why try to predict ICU outcomes? Curr Opin Crit Care 20(5):544–549. https://doi.org/10.1097/MCC.0000000000000136
2. Zimmerman JE, Kramer AA (2014) A history of outcome prediction in the ICU. Curr Opin Crit Care 20(5):550–556. https://doi.org/10.1097/MCC.0000000000000138
3. Pellathy TP, Pinsky MR, Hravnak M (2021) Intensive Care Unit scoring systems. Crit Care Nurse 41(4):54–64. https://doi.org/10.4037/ccn2021613
4. Harrison DA, Parry GJ, Carpenter JR, Short A, Rowan K (2007) A new risk prediction model for critical care: the Intensive Care National Audit & Research Centre (ICNARC) model. Crit Care Med 35(4):1091–1098. https://doi.org/10.1097/01.CCM.0000259468.24532.44
5. Paul E, Bailey M, Kasza J, Pilcher D (2016) The ANZROD model: better benchmarking of ICU outcomes and detection of outliers. Crit Care Resusc 18(1):25–36
6. Yang S, Stansbury LG, Rock P, Scalea T, Hu PF (2019) Linking big data and prediction strategies: tools, pitfalls, and lessons learned. Crit Care Med 47(6):840–848. https://doi.org/10.1097/CCM.0000000000003739
7. Kramer AA, Zimmerman JE, Knaus WA (2021) Severity of illness and predictive models in society of critical care medicine's first 50 years: a tale of concord and conflict. Crit Care Med 49(5):728–740. https://doi.org/10.1097/CCM.0000000000004924
8. Singer M, Deutschman CS, Seymour CW et al (2016) The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA 315(8):801–810. https://doi.org/10.1001/jama.2016.0287
9. Straney LD, Udy AA, Burrell A, Bergmeir C, Huckson S, Cooper DJ, Pilcher DV (2017) Modelling risk-adjusted variation in length of stay among Australian and New Zealand ICUs. PLoS One 12(5):e0176570. https://doi.org/10.1371/journal.pone.0176570
10. Badawi O, Breslow MJ (2012) Readmissions and death after ICU discharge: development and validation of two predictive models. PLoS ONE 7(11):e48758. https://doi.org/10.1371/journal.pone.0048758
11. Ranzani OT, Simpson ES, Japiassú AM, Noritomi DT, Amil Critical Care Group (2016) The challenge of predicting pressure ulcers in critically ill patients. A Multicenter Cohort Study. Ann Am Thorac Soc 13(10):1775–1783. https://doi.org/10.1513/AnnalsATS.201603-154OC
12. Kurtz P, Bastos LSL, Salluh JIF, Bozza FA, Soares M (2021) SAPS-3 performance for hospital mortality prediction in 30,571 patients with COVID-19 admitted to ICUs in Brazil. Intensive Care Med 47(9):1047–1049. https://doi.org/10.1007/s00134-021-06474-3
13. Dongelmans DA, Pilcher D, Beane A, Soares M, Del Pilar Arias Lopez M, Fernandez A, Guidet B, Haniffa R, Salluh JIF (2020) Linking of global intensive care (LOGIC): an international benchmarking in critical care initiative. J Crit Care 60:305–310. https://doi.org/10.1016/j.jcrc.2020.08.031
14. Haniffa R, Isaam I, De Silva AP, Dondorp AM, De Keizer NF (2018) Performance of critical care prognostic scoring systems in low and middle-income countries: a systematic review. Crit Care 22(1):18
15. Raffa JD, Johnson AEW, O'Brien Z, Pollard TJ, Mark RG, Celi LA, Pilcher D, Badawi O (2022) The global open source severity of illness score (GOSSIS). Crit Care Med 50(7):1040–1050. https://doi.org/10.1097/CCM.0000000000005518