UNDERSTANDING THE DISEASE

Understanding intensive care unit benchmarking

CrossMark

Jorge I. F. Salluh[1,2]*, Marcio Soares[1,2] and Mark T. Keegan[3]

Introduction

Originating from the surveyors' practice of placing chiseled horizontal marks in stone structures to form a "bench" for consistent placement of a leveling rod, the term "benchmarking" has evolved to mean the comparison of a business (or healthcare institution) with industry leaders, by evaluating a series of performance metrics. Benchmarking has been divided into the broad categories of process, performance, and strategic benchmarking, and has also been classified as internal (within the same institution) or external benchmarking. In relation to critical care medicine, benchmarking involves the use of quantitative, standardized measurements to allow comparison of performance between intensive care units (ICUs) [1].

For example, predictive models [e.g., the Acute Physiology and Chronic Health Evaluation (APACHE) score, the Simplified Acute Physiology Score (SAPS), and the Mortality Probability Model MPM)], have been developed and allow comparison of expected and actual mortality of critically ill patients through an evaluation of the severity and context of critical illness. Severity-adjusted mortality rates [or standardized mortality ratios (SMRs)] have been used in ICUs around the world for decades, helping to create a culture of performance evaluation [2]. SMRs have been criticized, however, because of the multiple factors that can affect them, including case-mix, cohort size, data collection methodology, bias in lead time, and the performance of the model. It is clear that case-mix is a key factor and should be considered when using SMRs in the comparative analysis of ICUs.

*Correspondence: jorgesalluh@gmail.com
[1] Department of Critical Care and Graduate Program in Translational Medicine, D'Or Institute for Research and Education, Rua Diniz Cordeiro, 30-3° andar, Rio de Janeiro, RJ 22281-100, Brazil
Full author information is available at the end of the article

Although the evaluation of a single ICU over time can produce interesting and insightful results, self-reflection can lead to excessive optimism or criticism. Benchmarking against other ICUs can provide ICU staff and hospital managers with a broader view and clearer perspectives of targets for improvement [1].

Areas of ICU performance suitable for benchmarking include mortality, adherence to processes of care, patient safety, economic outcomes, and patient or family satisfaction (Table 1). The aim of this report is to highlight the strengths and weaknesses of benchmarking and describe how it can be optimally applied in ICUs.

What should we benchmark?

In addition to the evaluation of severity-adjusted mortality rates, the search to identify markers of high-quality care has led to the scrutiny of lengths of ICU stay (LOS) and unplanned (and early) readmission rates. These entangled indicators are surrogates of cost and efficiency and typically reflect several aspects of care, including admission and discharge policies, adherence to best practices, and patient safety. Insightful information can be obtained when such indicators are analyzed in association with data on ICU staffing and resources, bed-availability and capacity strain, case-mix, nosocomial infection rates, and hospital structure. LOS, for example, should be used cautiously as a benchmarking tool as it is influenced by discharge criteria and the availability of step-down units and extra-hospital post-acute care facilities. The European Society of Intensive Care Medicine has recommended the use of specific quality indicators, including SMR, ICU readmission rate within 48 h of ICU discharge, and rates of catheter-related bloodstream infections and unplanned extubations [3].

Business management literature suggests that benchmarks should be "SMART"—specific, measurable, achievable, realistic, and timely. Although not

Springer

Table 1 What should we benchmark in critical care? Main advantages and disadvantages for different measures and indicators

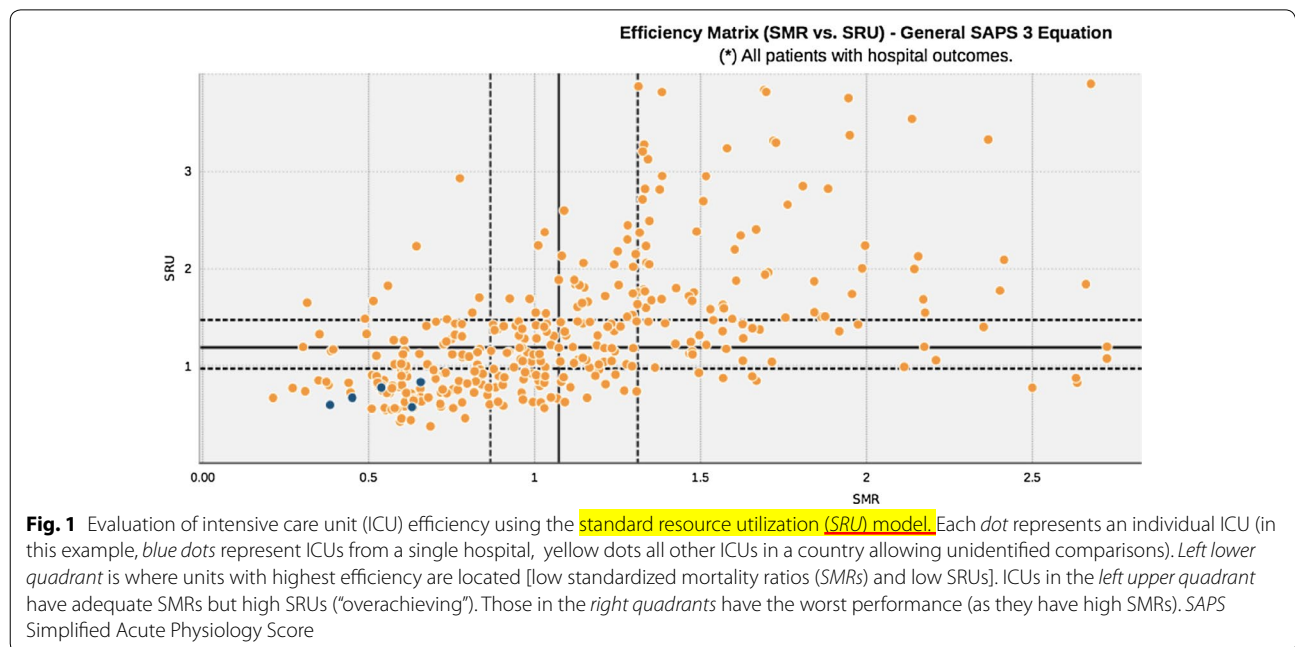| Domain/measure	Advantages	Limitations
Outcomes		
Mortality	Easy to measure, clinically relevant	Has to be risk-adjusted (SMR) with well-calibrated scores, case-mix sensitive, fixed time points are better than ICU/hospital
Length of stay	Easy to measure, clinically relevant, proxy of efficiency	Affected by structure, can be artificially lowered by transfers
ICU readmissions	Easy to measure, clinically relevant, indirect marker of clinical process inside and outside ICU. Evaluations of *unplanned* and *early* ICU readmissions are preferable as they reflect quality and safety	Affected by structure (e.g., step-down units), artificially lowered by transfers and end of life care policies
ICU acquired complications	Indicators of quality of care, there are validated recommended definitions, often modifiable/preventable	Affected by case-mix, frequently under-reported, applied definition may vary
Patient-reported outcomes	Post-ICU vital status and quality of life	Under-reporting, low-adherence, need for specialized platforms
Process of care		
Adherence to best practices and process of care aimed at preventing complications	Reliable surrogate of best practices, extensive EBM literature to support, can be used for audit-feedback purposes	Level of evidence varies according to the measures, effect on outcomes is variable, frequently under-reported, tricky to measure at bedside, frequently requires specialized monitoring system
ICU and hospital organization and structure		
Staffing patterns	Potentially associated with outcomes, easy to measure	Should be adjusted by risk and workload
ICU structure	Can be measured within countries where there are national requirements to provide intensive care; Can allow stratification of levels of care that can be provided by the ICU	Wide variation in national standards as well as in the definition of an ICU bed

ICU Intensive care unit, *SMR* standardized mortality ratio

evidence-based, this is a thoughtful and pragmatic approach. Garland has suggested that ICU performance should be measured in four domains that include medical, economic, psychosocial/ethical, and institutional outcomes [4]. ICU efficiency is also valuable for benchmarking. Rothen et al. evaluated ICU efficiency using the severity-adjusted (SAPS 3) resource, a measure that estimates the average amount of resources used per surviving patient in a specific ICU [the standardized resource use (SRU)] [5]. On the basis of median SMR and median SRU, each ICU is assigned to one of four groups: "most efficient" (all units whose SMR and SRU were below the median SMR and SRU); "least efficient" (units with both SMR and SRU above the median); "overachieving" (low SMR and high SRU); "underachieving" (high SMR and low SRU) (Fig. 1).

Ensuring relevant mortality comparisons

Survival—or not—is irrefutable and a relevant outcome measure. Direct comparisons of mortality among institutions (using funnel plots) and indirect comparisons against a risk-adjustment model (using process control charts) have proven useful [6]. A more nuanced consideration, however, is the selection of the time-point to be used for the assessment of mortality. Early in the history of outcome prediction and performance evaluation it became clear that survival to ICU discharge was an inadequate measure. The three main severity of illness scoring systems use survival to *hospital* discharge as the outcome of interest. For several reasons, however, hospital mortality is also being questioned as the sole point of assessment. The improvement in ICU and hospital survival rates has shifted focus from the evaluation of short-term survival to an assessment of post-ICU medium- and long-term quality of life. Additionally, discharge bias, affected by evolving discharge policies and the increasing availability of long-term post-acute care facilities to which patients may be transferred, may decrease the reliability of hospital mortality as a marker of quality [7, 8]. Therefore, SMR based on case-mix-adjusted mortality at a longer term fixed time-point after ICU admission may be preferable as a quality indicator for benchmarking purposes. For similar discharge-bias associated reasons, ICU LOS and readmission rates should be viewed with caution. Geographic region- and population-specific considerations must be taken into account, potentially requiring customization of predictive models. The heterogeneity of critical illness means that, for some conditions, there is substantial residual mortality in the post-ICU period that is not fully captured by measuring hospital mortality rates [9]. For example, based on epidemiologic data, it would appear that a minimum of 90 days follow-up is necessary to fully capture the mortality effect of sepsis. This contrasts with the apparently sufficient 30-day follow-up in patients who have suffered traumatic injuries not requiring operative intervention. Finally, patient-centered outcomes should be evaluated. Although they are harder to capture and follow, data on quality of life, functional status, and return to work are important measures to benchmark.

Fig. 1 Evaluation of intensive care unit (ICU) efficiency using the standard resource utilization (*SRU*) model. Each *dot* represents an individual ICU (in this example, *blue dots* represent ICUs from a single hospital, yellow dots all other ICUs in a country allowing unidentified comparisons). *Left lower quadrant* is where units with highest efficiency are located [low standardized mortality ratios (*SMRs*) and low SRUs]. ICUs in the *left upper quadrant* have adequate SMRs but high SRUs ("overachieving"). Those in the *right quadrants* have the worst performance (as they have high SMRs). *SAPS* Simplified Acute Physiology Score

Benchmarking processes of care

A complementary approach to benchmarking is to evaluate the adherence to evidence-based practices that are associated with improved outcomes [1, 10]. The rates of adherence to "standards of care" (e.g., low tidal volume ventilation in acute respiratory distress syndrome, prophylaxis against thromboembolism, early recognition and treatment of sepsis) may be ascertained and compared among ICUs. Although it may be argued that the correct benchmark for such measures is 100% adherence, knowledge of the compliance of other units with similar structural characteristics and case-mix may be an incentive to quality improvement, especially if the feasibility of achieving high standards of care in the real world is demonstrated [11, 12].

Comparison of complications

In a perfect world, it would be recommended—and useful—to compare unit-specific rates of hospital-acquired infections (e.g., ventilator associated pneumonia, catheter-related blood stream infections), the occurrence of ICU-acquired multi-resistant organisms or "problem" pathogens (e.g., *Clostridium difficile*, methicillin-resistant *Staphylococcus aureus*), and adverse events (e.g., unanticipated extubation). However, methodologic differences in data acquisition, inter-rater variability, and financial and legal disincentives to report may lead to unreliable incidence and prevalence rates, thus precluding accurate comparisons. Benchmarking these issues is only feasible and accurate in the context of very well-structured and standardized ICU networks, and even in such settings it may remain a complex task. To overcome these limitations ICUs should use the same definitions, potentially through the use of a data-dictionary with specific training and audit.

The future of ICU performance evaluation and benchmarking

The era of "the healthcare data revolution" with its advances in computerization and technologic infrastructure offers the potential for expansion of benchmarking [13]. With the advent of "big data" and "machine learning" updated prognostic models will inevitably become available, likely including a broader range of variables than currently employed [14–16]. If such models are developed on a multinational level, are easy to implement, and use an approach that allows course correction, they may finally make ICU-prediction models useful for individual patients. Widespread implementation of electronic medical records and the availability of real-time information provided by cloud-based structures will provide additional opportunities for comparison within and between institutions. Decreases in the burden of

data abstraction and the development of crowdsourcing will lead to the availability of increasing amounts of standardized, usable patient data. Ultimately, expansion of the domains of benchmarking is likely to occur, allowing evaluation of processes of care and multi-dimensional patient-centered outcomes in addition to the traditional mortality and length of stay comparisons [17].

Conclusions

Benchmarking of ICU performance is here to stay—and its use and complexity will likely expand as the healthcare data revolution proceeds. Although imperfect, severity-adjusted mortality rates and SMRs will continue to be used and refined. Evaluation of processes of care and compliance with commonly accepted practices offer an alternative approach to benchmarking, providing actionable data. It is hoped that widespread implementation of searchable electronic medical records and expansion of databases populated by automated data abstraction will lead to reliable intra- and inter-institutional comparisons, ultimately resulting in improved patient care.

Author details
[1] Department of Critical Care and Graduate Program in Translational Medicine, D'Or Institute for Research and Education, Rua Diniz Cordeiro, 30-3° andar, Rio de Janeiro, RJ 22281-100, Brazil. [2] Programa de Pós-Graduação em Clínica Médica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. [3] Division of Critical Care, Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, MN, USA.

References
1. Woodhouse D, Berg M, van der Putten J, Houtepen J (2009) Will benchmarking ICUs improve outcome? Curr Opin Crit Care 15:450–455
2. Keegan MT, Soares M (2016) What every intensivist should know about prognostic scoring systems and risk-adjusted mortality. Rev Bras Ter Intensiva 28:264–269
3. Rhodes A, Moreno RP, Azoulay E et al (2012) Prospectively defined indicators to improve the safety and quality of care for critically ill patients: a report from the Task Force on Safety and Quality of the European Society of Intensive Care Medicine (ESICM). Intensive Care Med 38:598–605
4. Garland A (2005) Improving the ICU: part 1. Chest 127:2151–2164
5. Rothen HU, Stricker K, Einfalt J et al (2007) Variability in outcome and resource use in intensive care units. Intensive Care Med 33:1329–1336
6. Power GS, Harrison DA (2014) Why try to predict ICU outcomes? Curr Opin Crit Care 20:544–549

7. Kahn JM, Benson NM, Appleby D, Carson SS, Iwashyna TJ (2010) Long-term acute care hospital utilization after critical illness. JAMA 303:2253–2259

8. Reineck LA, Pike F, Le TQ, Cicero BD, Iwashyna TJ, Kahn JM (2014) Hospital factors associated with discharge bias in ICU performance measurement. Crit Care Med 42:1055–1064

9. Taori G, Ho KM, George C et al (2009) Landmark survival as an end-point for trials in critically ill patients—comparison of alternative durations of follow-up: an exploratory analysis. Crit Care 13:R128

10. Watson SR, Scales DC (2013) Improving intensive care unit quality using collaborative networks. Crit Care Clin 29:77–89

11. Noritomi DT, Ranzani OT, Monteiro MB et al (2014) Implementation of a multifaceted sepsis education program in an emerging country setting: clinical outcomes and cost-effectiveness in a long-term follow-up study. Intensive Care Med 40:182–191

12. de Vos Maartje LG, van der Veer SN, Graafmans WC et al (2013) Process evaluation of a tailored multifaceted feedback program to improve the quality of intensive care by using quality indicators. BMJ Qual Saf 22:233–241

13. Ghassemi M, Celi LA, Stone DJ (2015) State of the art review: the data revolution in critical care. Crit Care 19:118

14. Iwashyna TJ, Liu V (2014) What's so different about big data? A primer for clinicians trained to think epidemiologically. Ann Am Thorac Soc 11:1130–1135

15. Desautels T, Calvert J, Hoffman J et al (2016) Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Med Inform 4:e28

16. Pirracchio R, Petersen ML, Carone M et al (2015) Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. Lancet Respir Med 3:42–52

17. Cox CE, Wysham NG, Kamal AH et al (2016) Usability testing of an electronic patient-reported outcome system for survivors of critical illness. Am J Crit Care 25:340–349