Journal of Critical Care 70 (2022) 154063

Contents lists available at ScienceDirect

Journal of Critical Care

journal homepage: www.journals.elsevier.com/journal-of-critical-care

Comparing continuous versus categorical measures to assess and benchmark intensive care unit performance

Leonardo S.L. Bastos [a,1,*], Safira A. Wortel [b,c,d,1], Nicolette F. de Keizer [b,c,d], Ferishta Bakhshi-Raiez [b,c,d], Jorge I.F. Salluh [e,f,g], Dave A. Dongelmans [c,h], Fernando G. Zampieri [e,f], Gastón Burghi [i], Ameen Abu-Hanna [b], Silvio Hamacher [a], Fernando A. Bozza [e,f,j], Marcio Soares [e,f]

[a] Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, RJ, Brazil
[b] Amsterdam UMC location University of Amsterdam, Department of Medical Informatics, Meibergdreef 9, Amsterdam, the Netherlands
[c] National Intensive Care Evaluation (NICE) Foundation, Amsterdam UMC location University of Amsterdam, Department of Medical Informatics, Meibergdreef 9, Amsterdam, the Netherlands
[d] Amsterdam Public Health, Quality of care, Amsterdam, the Netherlands
[e] D'Or Institute for Research and Education (IDOR), Rio de Janeiro, RJ, Brazil
[f] Brazilian Research in Intensive Care Network (BRICNet), Brazil
[g] PostGraduate, Internal Medicine, Program Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil
[h] Department of Intensive Care, Amsterdam UMC, Amsterdam, the Netherlands
[i] Intensive Care Unit, Hospital Maciel, Montevideo, Uruguay
[j] Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, RJ, Brazil

A R T I C L E   I N F O

Keywords:
Intensive care
ICU benchmarking
ICU efficiency
Outcome metrics
Quality indicators

A B S T R A C T

Purpose: To compare categorical and continuous combinations of the standardized mortality ratio (SMR) and the standardized resource use (SRU) to evaluate ICU performance.
Materials and methods: We analysed data from adult patients admitted to 128 ICUs in Brazil and Uruguay (BR/UY) and 83 ICUs in The Netherlands between 2016 and 2018. SMR and SRU were calculated using SAPS-3 (BR/UY) or APACHE-IV (The Netherlands). Performance was defined as a combination of metrics. The categorical combination was the efficiency matrix, whereas the continuous combination was the average SMR and SRU (average standardized ratio, ASER). Association among metrics in each dataset was evaluated using Spearman's rho and R^2.
Results: We included 277,459 BR/UY and 164,399 Dutch admissions. Median [interquartile range] ASER = 0.99 [0.83–1.21] in BR/UY and 0.99[0.92–1.09] in Dutch datasets. The SMR and SRU were more correlated in BR/UY ICUs than in Dutch ICUs (Spearman's Rho: 0.54vs.0.24). The highest and lowest ASER values were concentrated in the least and most efficient groups. An expert focus group listed potential advantages and limitations of both combinations.
Conclusions: The categorical combination of metrics is easy to interpret but limits statistical inference for benchmarking. The continuous combination offers appropriate statistical properties for evaluating performance when metrics are positively correlated.

1. Introduction

Benchmarking processes and outcomes metrics provide healthcare professionals and policymakers opportunities to identify outliers and targets for quality improvement [1]. In intensive care, benchmarking of performance is frequently applied using risk-adjusted mortality and resource use measures [2]. Intensive care unit (ICU) performance should be evaluated in different perspectives [3-5], and standardized outcome measures have been preferred since they are case-mix adjusted and easy to interpret. The two most commonly used metrics to assess ICU performance are the standardized mortality ratio (SMR) and the standardized length-of-stay, also called standardized resource use (SRU), which measure the clinical efficacy and the efficiency of a unit, respectively [5-7].

Quantifying ICU performance based on the combination of these two measures is challenging. A few studies have considered different ways of combining SMR and SRU, for example, categorical and continuous approaches. The first has traditionally been used in ICU benchmarking,

Abbreviations: APACHE, Acute Physiology and Chronic Health Evaluation; ICU, Intensive care unit; IQR, Interquartile Range; SAPS, Simplified Acute Physiology Score; SMR, Standardized Mortality Ratio; SRU, Standardized Resource Use.

* Corresponding author.
E-mail address: lslbastos@puc-rio.br (L.S.L. Bastos).
[1] Authors LSLB and SAW contributed equally to this work

https://doi.org/10.1016/j.jcrc.2022.154063

such as the Rapoport-Teres graph or the "efficiency matrix" [7-11]. This approach uses the median values of SMR and SRU to categorize units into four efficiency groups. The second approach was proposed more recently and averages the SMR and SRU for each unit, thus obtaining a single continuous performance metric [12].

Each approach has its benefits and drawbacks. The categorical approach can identify groups of interest straightforwardly, such as the best and worst-performing units [9,13]. However, categorising a continuous outcome may result in a loss of information for further inference analysis. Otherwise, the continuous approach is an attractive alternative to reduce information loss and improve comparisons when benchmarking. However, the resulting single performance metric has yet to be assessed.

This study aimed to compare the categorical and the continuous approaches combining SMR and SRU to evaluate and benchmark ICU performance. We hypothesize that clinicians have different opinions on the application possibilities of the two approaches and that different relations between SMR and SRU might provide distinct insights and interpretations in practice. An ICU might be considered efficient using one approach but not when using the other, or the data sample could affect ICU efficiency. Hence, we used two datasets of ICUs, one from two South American countries, Brazil, and Uruguay, and the other from The Netherlands. We assessed the categorical and the continuous combinations of the SMR and SRU in the two different settings and provided recommendations on the usage of both approaches.

2. Materials and methods

2.1. Study design and data source

We performed a retrospective observational analysis on data from two large ICU networks in Brazil/Uruguay (BR/UY) and The Netherlands [11,14,15]. BR/UY data was obtained from the "Organizational CHaractEriSTics in cRitical cAre" (ORCHESTRA) network [11,15]. This dataset contains demographic and clinical data, and outcomes for adult patients (≥ 16 years old) admitted to 128 ICUs (123 from Brazil and five from Uruguay) in 77 hospitals (72 from Brazil and five from Uruguay) from 2016 to 2018. In both countries, patient data were retrieved from the *Epimed Monitor System®* (*Epimed Solutions®*, Rio de Janeiro, Brazil) [16]. Dutch patient admission data was obtained from the Dutch National Intensive Care Evaluation (NICE) registry, a non-profit foundation established by intensivists in 1996 [14] to evaluate ICU performance and quality of care. It consists of demographic, physiological, and clinical data and outcomes of ICU patients from all Dutch ICUs in the Netherlands, mainly extracted from electronic patient records and manually validated according to stringent data quality measures. The Brazilian National Ethics Committee (Brazil CAAE: 19687113.8.1001.5249), the Ethics committee of the *Hospital Maciel, Montevidéo*, Uruguay (protocol nº 20/2017), and the Medical Ethics Review Committee of the Amsterdam University Medical Centers (reference number W20_192#20.223) approved the study and waived the need for informed consent.

2.2. Study population

We included all adult ICU patients (age ≥ 18 years old) in both datasets admitted between 2016 and 2018. The countries use different severity of illness scoring systems. Hence, in the BR/UY dataset, patients were excluded based on the Simplified Acute Physiology Score 3 (SAPS-3) exclusion criteria [17], consisting of patients with missing core data such as age, location before ICU admission, main ICU admission diagnosis, and readmissions. In the Dutch dataset, patients were excluded based on the Acute Physiology And Chronic Health Evaluation (APACHE) IV exclusion criteria: patients with an ICU length-of-stay (LOS) ≤ 4 h or longer than one year; readmissions; patients admitted from another Coronary Care Unit (CCU) or ICU; patients with missing admission diagnosis or admission type; patients with burns; transplant patients or CCU or recovery patients [18]. In addition, patients with missing SAPS-3 or APACHE-IV scores were excluded. We excluded ICUs with more than 10% missing data of admission diagnosis or hospital outcomes. Patient-level data furthermore consisted of the patient's demographics (age and gender), type of admission (i.e., medical, elective surgical, or urgent surgical), the severity of illness at ICU admission, i.e., the SAPS-3 score and probability used in BR/UY ICUs, and the APACHE-IV score and probability used in Dutch ICUs, the in-hospital and ICU mortality, and ICU length-of-stay in days (defined as 24-h periods based on admission and discharge dates). Organizational level data consisted of ICU and hospital size expressed as the number of beds.

2.3. Outcomes and ICU performance

Our primary outcome was the ICU performance as the combination of two outcome measures, the SMR and SRU, both adjusted for the severity of illness score as used in each country.

The SMR corresponds to the ratio of the observed number of deaths to the expected number of deaths. The expected number of deaths was obtained by the sum of mortality probabilities obtained from the SAPS-3 standard equation [19] or APACHE-IV risk models [18]. We performed a first-level customization of risk models to reduce the potential over- or underestimation of the predicted risks, thus providing recalibrated mortality risks (Appendix A, Fig. S1).

Similarly, SRU corresponds to the observed resource use to the expected resource use ratio. For this purpose, we considered the ICU LOS as a surrogate measure of ICU resource use [9]. Following the SRU methodology [9,13], the observed use of resources was calculated as the total ICU LOS and the expected use of resources was the average ICU LOS per surviving patient. We obtained ICU LOS estimates for each decile of the recalibrated probabilities obtained from SAPS-3 (BR/UY) or APACHE-IV (The Netherlands) models to reduce potential biases due to miscalibration (Appendix A, Table S1).

We combined SMR and SRU in two ways: a categorical approach and a continuous approach. In the first, we grouped the ICUs using the SMR and SRU efficiency matrix [8,9]. This method uses the respective medians of SMR and SRU distribution to define efficiency groups: the most efficient (both SMR and SRU < median), underachieving (SMR ≥ median and SRU < median), overachieving (SMR < median and SRU ≥ median), and the least efficient (both SMR and SRU ≥ medians). Second, we used the average of SMR and SRU to obtain a single performance metric [12], defined as (SMR + SRU)/2. We refer to this metric as the Average Standardized Efficiency Ratio (ASER). Since this metric is derived from SMR and SRU, the interpretation is similar: the lower the ASER, the better the ICU performance and efficiency than expected.

2.4. Data analysis

We described the study population (patients and ICUs) from both countries. We used median and interquartile range (IQR) or mean and standard deviation (SD) for continuous variables depending on their distribution. For categorical variables, we used absolute frequencies and proportions.

For each country, we analysed the distribution of SMR and SRU values. We evaluated the SAPS-3 and APACHE-IV risk probabilities using the calibration belts [20] (Appendix A, Fig. S1). Using the efficiency matrix, we visualised the association between SMR and SRU and estimated their correlation using Spearman's rho coefficient [21].

We calculated the ASER per ICU per dataset. To assess the ASER in each sample, we evaluated its distribution and association with SMR and SRU both combined and individually. First, we added the ASER in the efficiency matrix, observed the low and high performing units' pattern, and described the ASER distribution per efficiency group. Then, using a linear regression model, we estimated the R^2 coefficient of determination to obtain the level of association using SMR or SRU as the response variable and as the ASER predictor. We also assessed potential unexpected behaviour of units due to under and overestimation in ASER, SMR and SRU using funnel plots [21,22]. As a sensitivity analysis, we evaluated the association between SMR, SRU and ASER stratified by each year of the study period (2016, 2017, and 2018).

Since the severity of illness of patients in ICUs from both datasets was not comparable due to the different risk scores (SAPS-3 vs APACHE-IV) used, we performed all analyses in each of them separately. A *p*-value of 0.05 was considered significant in statistical tests, and 95% confidence intervals were calculated using 5000 bootstrapped samples. We performed all statistical analyses in R version 4.1.2. We followed STROBE guideline recommendations (Appendix B).

Finally, we conducted an expert focus group to obtain a list of recommendations for using both combinations, including limitations, clinical and statistical interpretation, and implications for further use for benchmarking purposes. This group consisted of four intensivists and four statisticians/methodologists, who discussed and identified the main clinical and statistical advantages and disadvantages using both approaches.

3. Results

Of 441,858 patients, 277,459 (63%) were admitted to 128 BR/UY ICUs and 164,399 (37%) were admitted to 83 Dutch ICUs. On average, BR/UY ICUs were larger than Dutch ICUs (Table 1). The median age in BR/UY ICUs (Table 2) was comparable to the median age in Dutch ICUs, but the proportion of patients over 60 years old was higher in the Dutch population (62.9% vs 58.7%). The median SAPS-3 score in BR/UY patients was 43 [IQR: 34–54], and the median APACHE-IV score in Dutch patients was 53 [IQR: 38–75]. Dutch ICUs admitted more male than female patients. This difference was absent in BR/UY (male: 50.6% vs 49.3%). Most ICU admissions were medical in both samples (over 60%). Proportions of crude ICU and in-hospital mortality in BR/UY and Dutch units were comparable (ICU mortality: 8.8% vs 9.8%; hospital mortality: 14% vs 14%). ICU length of stay between BR/UY and Dutch patients was also similar (median ICU LOS = 2 [IQR: 1, 5] days and 2 [IQR: 2, 4] days, respectively).

The distribution of SMR and SRU was quite different between the two countries (Fig. 1, Appendix A, Fig. S2). BR/UY units showed larger SMR and SRU variability than Dutch units (Table 1, Fig. 1). The SMR and SRU were more correlated in BR/UY ICUs than in Dutch ICUs (Spearman's Rho: 0.54 vs 0.24). The proportion of units in overachieving or underachieving groups was lower in BR/UY than in the Dutch dataset (34% vs 41%, respectively, Appendix A, Table S2). When observing the ASER values for both datasets, the highest and lowest values of ASER were concentrated in the least and most efficient groups, respectively. However, the BR/UY dataset has more units with high SMR or SRU, expressed as a larger number of extreme ASER values, mostly concentrated in the "least efficient" group. In contrast, Dutch units are more concentrated toward the median (Fig. 1).

Fig. 2 shows the association between SMR and SRU with the ASER using a linear regression model. In the BR/UY sample, the association between SMR and SRU with ASER was high (R^2: 0.74 and 0.81 for SMR and SRU, respectively), while in the Dutch sample, this association was lower (R^2: 0.67 and 0.60 for SMR and SRU respectively). In addition, the analysis using funnel plots showed that the ICUs were mainly located within the control limits for SMR, SRU and ASER, and, therefore, were no signs of relevant under or overestimation of values (Appendix A, Fig. S3). In our sensitivity analysis, we observed that the results on the association between SMR, SRU and ASER for each year were similar to those in the main analysis in both BR/UY and Dutch datasets (Appendix A, Figs. S4–S6).

Inspired by these quantitative results, the expert focus group composed a list of potential advantages, limitations and statistical and clinical interpretations of considering a categorical and continuous approach for combining SMR and SRU (Table 3). From a statistical perspective, the categorical approach implies that units from the same group have similar performance, whereas the continuous combination provides a span of performance values. The latter feature provides more information for statistical models, thus not requiring a large amount of data to perform benchmarking or providing better estimates in further statistical analysis. Furthermore, when observing their usability and interpretation, categorising performance metrics provides a very straightforward performance indicator (e.g., low or high). However, the classification is often post-hoc and depends on the chosen cut-offs. Additionally, the continuous combination provides a general performance metric, facilitating the benchmarking process, but it is limited when metrics present low or negative correlation.

Table 1
Characteristics and outcome metrics of Brazilian/Uruguayan and Dutch ICUs.

Characteristics and outcomes	Brazil/Uruguay	The Netherlands
Total number of ICUs	128	83
ICU Beds, median (IQR)	14 (10,20)	12 (7, 16)
Hospital Beds, median (IQR)	217 (151, 380)	438 (315, 626)
Proportion of ICU bed/Hospital bed (%), median (IQR)	5.9 (3.3, 12.1)	2.7 (2.1, 3.5)
Overall outcome performance, median (IQR)		
Standardized Mortality Ratio (SMR)	0.97 (0.76, 1.21)	1.00 (0.89, 1.12)
Standardized Resource Use (SRU)	1.06 (0.79, 1.30)	0.98 (0.88, 1.08)
Averaged Standardized Efficiency Ratio (ASER)	0.99 (0.82, 1.21)	0.99 (0.92, 1.09)

ICU: Intensive Care Unit.
IQR: Interquartile Range (1st Quartile, 3rd Quartile).
ASER: Average between SMR and SRU.

Table 2
Characteristics and outcomes of critically ill patients in Brazil/Uruguay and The Netherlands.

Characteristics and Outcomes	Brazil and Uruguay	The Netherlands
Total number of patients	277,459	164,399
Age (years), median (IQR)	65 (49, 78)	66 (54,75)
16–30, N (%)	19,487 (7%)	9599 (5.8%)
31–40	25,798 (9.3%)	8558 (5.2%)
41–50	28,144 (10%)	15,300 (9.3%)
51–60	41,238 (15%)	27,707 (16.9%)
61–70	52,472 (19%)	42,542 (25.9%)
71–80	52,960 (19%)	41,572 (25.3%)
81–90	44,918 (16.2%)	17,716 (10.8%)
>90	12,442 (4.5%)	1405 (0.9%)
Gender, N (%)		
Female	140,529 (50.6%)	69,636 (42.4%)
Male	136,790 (49.3%)	94,748 (57.6%)
Unknown/Transgender	140 (0.1%)	15 (<0.1%)
Admission type, N (%)		
Medical	183,561 (66%)	100,252 (61.0%)
Elective surgery	74,610 (27%)	43,742 (26.6%)
Urgent surgery	19,288 (7.0%)	20,405 (12.4%)
Severity-of-illness score, median (IQR)		
Simplified Acute Physiology Score (SAPS-3)	43 (34, 54)	–
Acute Physiology and Chronic Health Evaluation (APACHE-IV)	–	53 (38, 75)
Predicted mortality risk		
SAPS-3 Predicted mortality risk	0.09 (0.03, 0.24)	–
APACHE-IV Predicted mortality risk	–	0.08 (0.03, 0.26)
Outcomes		
ICU Length-of-Stay (days), median (IQR)	2 (1, 5)	2 (2, 4)
ICU mortality, N (%)	24,470 (8.8%)	16,048 (9.8%)
Hospital mortality, N (%)	37,557 (14%)	23,023 (14%)

ICU: Intensive Care Unit; IQR: Interquartile Range (1st Quartile – 3rd Quartile); SD: Standard Deviation.
–: not available in the dataset.

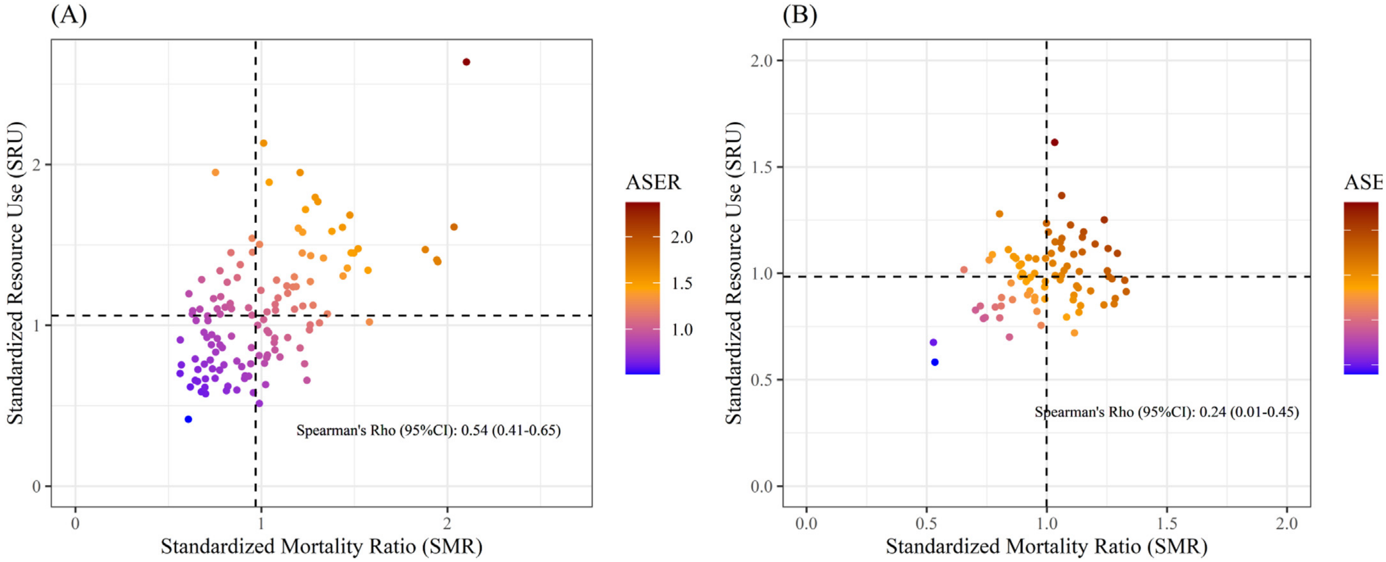

Fig. 1. Distribution of the Standardized Mortality Ratio (SMR), Standardized Resource Use (SRU), and Average Standardized Efficiency Ratio (ASER) values in the efficiency matrix in (A) Brazilian/Uruguayan ICUs and (B) Dutch ICUs. Mortality risks were obtained using the SAPS-3 model for Brazilian/Uruguayan units and the APACHE-IV model for Dutch units.

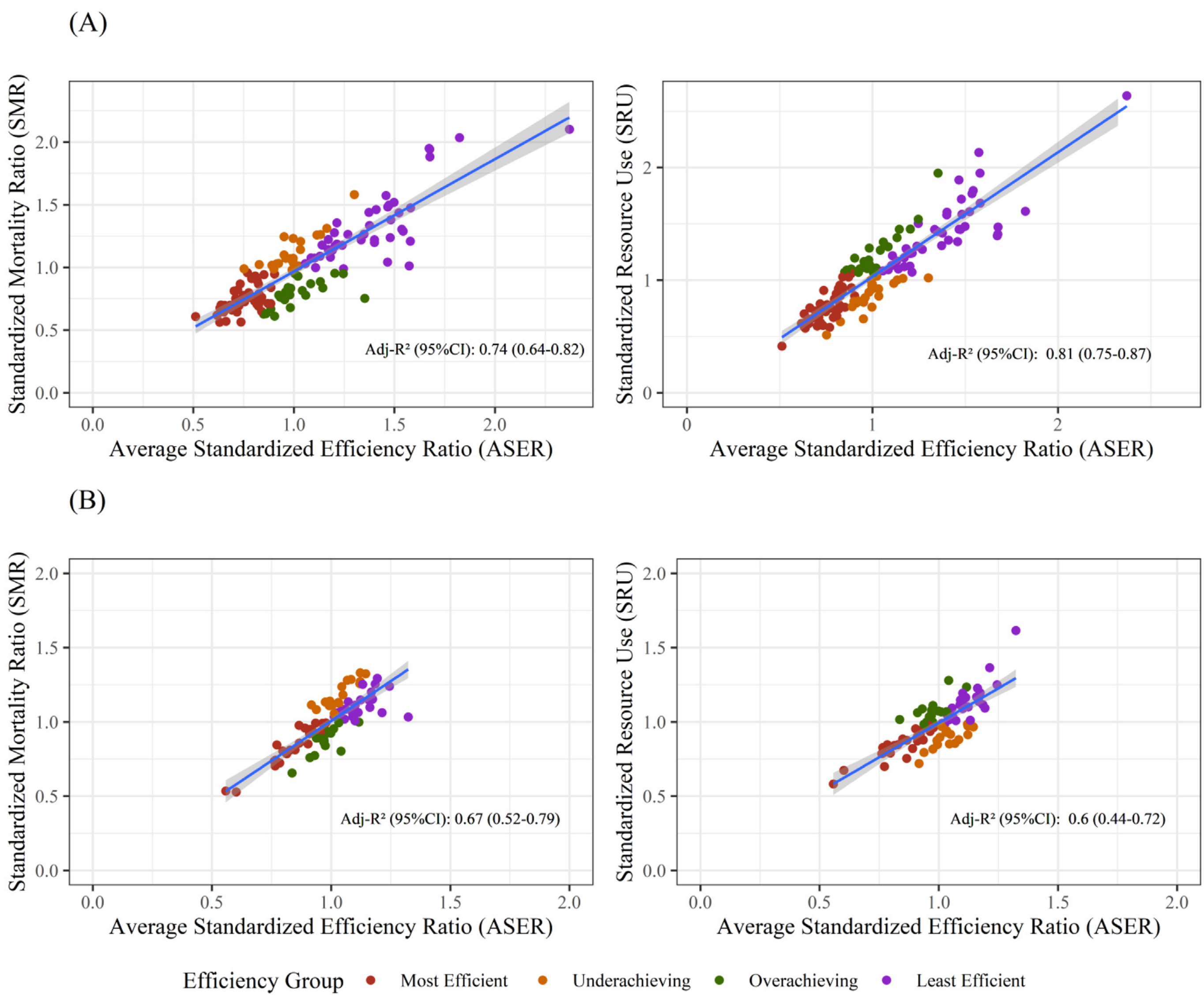

Fig. 2. Association between the Standardized Mortality Ratio (SMR), Standardized Resource Use (SRU), and Average Standardized Efficiency Ratio (ASER) between (A) Brazilian/Uruguayan ICUs and (B) Dutch ICUs. The blue line is the regression with confidence intervals (shaded area). R^2 was obtained from the linear regression model using the ASER as a predictor for SMR or SRU.

4. Discussion

We evaluated two approaches to combine SMR and SRU for ICU benchmarking using data from two large ICU networks in Brazil, Uruguay, and The Netherlands. We observed that the correlation between SMR and SRU influences the properties of their combination. A high positive correlation between SMR and SRU favours using the average as a general efficiency metric. In contrast, a lower correlation provides a balanced distribution of units per quadrant in the efficiency matrix.

As expected, our results showed a high association between SMR and SRU with ASER in the BR/UY sample (i.e., R^2: 0.74 and 0.85 for SMR and SRU, respectively). This association was lower in the Dutch ICUs (i.e., R^2: 0.67 and 0.60, respectively), and the most and least efficient groups become less distinguishable from under- or overachieving groups (Fig. 2). This finding potentially indicates that a low correlation between SMR and SRU affected the distribution of the ASER. In addition, the "cut-offs" for defining the efficiency groups may provide misclassifications and produce unreliable results, thus decreasing the rankability. We noted that, although large, the BR/UY data corresponded to a convenience sample, whereas the Dutch data came from a complete coverage national database; thus, national coverage is different. However, the heterogeneity in the BR/UY healthcare system we describe here has been previously evidenced [10,11,15,23]. We note that although direct comparison of ICU performance between countries was not possible, SMR and SRU estimates were obtained from recalibrated SAPS-3 and APACHE-IV models. This is in line with earlier literature showing that regular assessment of calibration from mortality models is important to provide reliable analysis [24].

Combining metrics is a natural approach when evaluating more than one performance measure [25]. Decisions on whether to combine or not are essential in ICU benchmarking since metrics must represent the unit's performance, and their interpretation can influence clinical and managerial decision-making. Categorising into single continuous metrics or variables have been published before [7,9,10,26]. Our focus group identified similar implications for using the efficiency matrix or the average SMR and SRU.

The continuous approach, such as the ASER, retains the original information and interpretation of performance among units and enables a more straightforward comparison. We noted that this combination is only useful as a general outcome metric if SMR and SRU have a high positive correlation. A negative correlation indicates that the metrics are in the opposite direction, potentially incurring regression to the mean and not favouring its usage as a general metric [21]. On the

Table 3
Advantages and disadvantages of using a categorical/dichotomous versus the continuous representation of ICU.

	Dichotomous/Categorical	Continuous
Definition	Classify ICUs of efficiency groups ("Efficiency matrix", Rapoport et al., 1994/Rothen et al., 2007): Most efficient ICUs are defined as ICUs with an SMR < median SMR and SRU < median SRU, least efficient ICUs are defined as SMR > median SMR and SRU > median SRU	Calculate the arithmetic mean between SMR and SRU. Averaged Standardized Efficiency Ratio (ASER) ASER = (SMR + SRU)/2
Studies where the outcome is used	[Rothen et al. 2007] [9], [Nathanson et al. 2007] [7], [Soares et al. 2015] [11], [Bastos et al. 2020] [10]	[Wortel et al. 2021] [12]
Statistical analyses	Simplifies statistical analyses. Units in the same group are assumed to have similar performance. Can evaluate differences among groups of efficiency using statistical tests or regression analysis (e.g., comparing most efficient vs least efficient units) Subgroups of ICUs can be further explored.	Fewer observations/ICUs are needed to observe effects in statistical modelling. Units are not necessarily similar in terms of performance. A continuous metric provides a span of values of performance metrics that can be analysed. Depending on the distribution of the ASER, multiple parametric or non-parametric statistical analyses can straightforwardly be applied.
Limitations	A larger number of observations/ICUs is needed to obtain consistent estimates in statistical modelling (e.g., regression models). Dichotomisation would lead to a loss of information about the true relationship between variables, resulting in a loss of statistical power and a decreased effect size. Choosing two groups (e.g., most efficient vs least efficient units) may limit the power of the analysis. When regression is being used to adjust for the effect of a confounding variable, dichotomisation will run the risk that a substantial part of the confounding remains.	Averaging SMR and SRU may not be fully representative of a unit's actual performance (e.g., units with high SMR and low SMR may present an average value closer to the reference lines for grouping units) SMR and SRU may have different weights during the decision-making process. The arithmetic mean is a simple approach to combine those metrics. Other averaging metrics and different weights could be incorporated.
Statistical interpretation	Generally, there is no good reason to suppose that there is an underlying dichotomy, and if one exists, there is no reason it should be at the median. Therefore, interpretation of the results is highly dependent on the chosen cut-off point (e.g., median SMR and median SRU). It makes it challenging to model other categories of ICUs, e.g., under- and overachieving ICUs (ICUs with SMR < median SMR and SRU > median SRU and vice versa) Dichotomisation conceals any non-linearity in the relation between the variable and outcome.	For ICUs with very low SMR and very high SRU (or vice versa), the resulting ASER is distorted. Hence, a negative correlation between SMR and SRU may limit the achievement of a general outcome metric (metrics point to opposite directions).
Clinical interpretation of the definition	Interpretation of ICUs performance is straightforward: a unit is either efficient or not. ICUs close to the cut-off point but on opposite sides are characterised as being very different rather than similar.	Interpretation is not always clear: e.g., it is difficult to identify which ICUs are 'good' and 'bad', and there is no cut-off point. However, a single measure score may provide straightforward information on overall efficiency. It might be more important for some ICUs to know how they score on the SMR, while others might find their performance based on SRU more important. It is unclear which of the underlying indicators the ICU could improve with a single average score.

ASER: Averaged Standardized Efficiency Ratio; ICU: Intensive Care Unit; SMR: Standardized Mortality Ratio; SRU: Standardized Resource Use.

other hand, the categorical/dichotomous approach provides a straightforward (clinical) interpretation [27] since an ICU is positioned into a specific performance category. However, units in different efficiency groups but very close to the cut-off points are considered different, while their SMR and SRU values might be very similar. Furthermore, units from the same group are considered equal and, thus, this loss of information may harm further analyses, especially when the sample size is already small [28,29]. In fact, there are rare reasons for discretising a continuous metric, for instance, only if the categories are intrinsic to the variable's nature or known cut-offs in its distribution [30,31].

Choosing one of those combination approaches will depend on the management, research, or clinical objectives. A categorical approach seems adequate to identify performance groups instead of individual comparison. For instance, the ICU manager can quickly identify if the unit must improve mortality or use of resources. However, even within those groups, targets for improvement in organizational processes may be different, and grouping may reduce the comparability among units. On the other hand, if the goal is to compare individual performance, a continuous approach may be preferable. For instance, continuous metrics may provide better statistical properties and more robust results in inference analysis, which assist studies that evaluate the association between organizational factors and ICU efficiency and identify targets for process improvement [10-13]. This behaviour might explain possible non-significant results due to dichotomising performance in addition to reduced sample size when comparing two groups (e.g., "least" and "most" efficient groups).

We evaluated the association between SMR and SRU with their combination approaches. SMR and SRU presented a high positive correlation for BR/UY ICUs, resulting in units concentrated in quadrants of "least" and "most" efficient units in the efficiency matrix. For Dutch units, the correlation was lower, and all efficiency quadrants were more equally distributed (Fig. 1). However, ASER was correlated with SMR and SRU individually in both countries, especially for BR/UY units (Fig. 2). Hence, the average SMR and SRU could be used as a general efficiency measure in those settings of moderate/high correlation between those metrics.

The strengths of our work include the analysis of outcome metrics from large national networks of ICUs in three distinct countries, Brazil, Uruguay, and The Netherlands. We had information on the severity of illness and outcomes for all studied patients. The present study also has some limitations. First, comparisons among countries were not possible due to differences in the locally adopted severity of illness score. However, we performed our analysis per dataset, considering their distribution of SMR and SRU. Second, our analysis was limited to the combination of two metrics. Using the average may give reasonable results if more than two metrics present considerable collinearity. Third, we used the arithmetic mean to compute the ASER, thus implying equal weights for SMR and SRU, which may differ in other circumstances with different priorities. However, we noted that the continuous combination could easily incorporate different weights and keep the properties discussed previously, especially when the correlation between SMR and SRU is high. The categorisation becomes challenging in the case of multiple dimensions. Our study analysed the two main metrics used in ICU benchmarking, SMR and SRU, measuring different performance perspectives. Hence, different combination methods, such as clustering or data envelopment analysis, should be applied.

5. Conclusion

Combining measures of performance indicators will always conceal some degree of information. The continuous combination offers appropriate statistical properties for evaluating benchmarking when outcome metrics are highly positively correlated, assisting the identification of process improvement targets. The categorical combination facilitates the interpretation of outcomes results, such as identifying best and worst-performing ICUs but should be used with caution when the number of units is limited.

Funding

This study was supported by the National Council for Scientific and Technological Development (CNPq), the Coordination for the Improvement of Higher Education Personnel (CAPES) - Finance Code 001, Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ), the Pontifical Catholic University of Rio de Janeiro, and by departmental funds from the D'Or Institute for Research and Education.

Credit authorship contribution statement

Leonardo S.L. Bastos: Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Safira A. Wortel:** Conceptualization, Methodology, Data curation, Writing – original draft, Writing – review & editing. **Nicolette F. de Keizer:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Ferishta Bakhshi-Raiez:** Conceptualization, Methodology, Data curation, Writing – review & editing. **Jorge I.F. Salluh:** Conceptualization, Methodology, Writing – review & editing. **Dave A. Dongelmans:** Conceptualization, Methodology, Writing – review & editing. **Fernando G. Zampieri:** Conceptualization, Methodology, Data curation, Writing – review & editing. **Gastón Burghi:** Conceptualization, Methodology, Data curation, Writing – review & editing. **Ameen Abu-Hanna:** Conceptualization, Methodology, Writing – review & editing. **Silvio Hamacher:** Conceptualization, Methodology, Writing – review & editing. **Fernando A. Bozza:** Conceptualization, Methodology, Writing – review & editing. **Marcio Soares:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Declaration of Competing Interest

The funders had no role in study design, data collection and analysis, decision to publish, or the preparation of the manuscript. Dr. Soares and Dr. Salluh are founders and equity shareholders of Epimed Solutions®, which commercializes the Epimed Monitor System®, a cloud-based software for ICU management and benchmarking. Dr. Zampieri has received grants for investigator-initiated studies from Ionis Pharmaceuticals (USA), Bactiguard (Sweden) and the Brazilian Ministry of Health, none related to the scope of this study. N.F. de Keizer and D.A. Dongelmans are members of the board of the Dutch National Intensive Care Evaluation (NICE) foundation. The NICE foundation pays the department of Medical Informatics, Amsterdam UMC, for processing data of all Dutch ICUs into audit and feedback information. S.A., N.F. de Keizer and F. Bakhshi-Raiez are employees of the department of medical informatics and work on the NICE project. The other authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcrc.2022.154063.

References

[1] Salluh JIF, Chiche JD, Reis CE, Soares M. New perspectives to improve critical care benchmarking. Ann Intensive Care. 2018;8:17. https://doi.org/10.1186/s13613-018-0363-0.

[2] Woodhouse D, Berg M, van der Putten J, Houtepen J. Will benchmarking ICUs improve outcome? Curr Opin Crit Care. 2009;15:450–5. https://doi.org/10.1097/MCC.0b013e32833079fb.

[3] Donabedian A. The quality of care. How can it be assessed? JAMA. 1988;260:1743–8. https://doi.org/10.1001/jama.260.12.1743.

[4] Garland A. Improving the ICU: part 2. Chest. 2005;127:2165–79. https://doi.org/10.1378/chest.127.6.2165.

[5] Salluh JIF, Soares M, Keegan MT. Understanding intensive care unit benchmarking. Intensive Care Med. 2017;43:1703–7. https://doi.org/10.1007/s00134-017-4760-x.

[6] Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. Crit Care Med. 2011;39:163–9. https://doi.org/10.1097/CCM.0b013e3181f96f81.

[7] Nathanson BH, Higgins TL, Teres D, Copes WS, Kramer A, Stark M. A revised method to assess intensive care unit clinical performance and resource utilization*. Crit Care Med. 2007;35:1853–62. https://doi.org/10.1097/01.CCM.0000275272.57237.53.

[8] Rapoport J, Teres D, Lemeshow S, Gehlbach S. A method for assessing the clinical performance and cost-effectiveness of intensive care units: a multicenter inception cohort study. Crit Care Med. 1994;22:1385–91.

[9] Rothen HU, Stricker K, Einfalt J, Bauer P, Metnitz PGH, Moreno RP, et al. Variability in outcome and resource use in intensive care units. Intensive Care Med. 2007;33:1329–36. https://doi.org/10.1007/s00134-007-0690-3.

[10] Bastos LSL, Hamacher S, Zampieri FG, Cavalcanti AB, Salluh JIF, Bozza FA. Structure and process associated with the efficiency of intensive care units in low-resource settings: an analysis of the CHECKLIST-ICU trial database. J Crit Care. 2020;59:118–23. https://doi.org/10.1016/j.jcrc.2020.06.008.

[11] Soares M, Bozza FA, Angus DC, Japiassú AM, Viana WN, Costa R, et al. Organizational characteristics, outcomes, and resource use in 78 Brazilian intensive care units: the ORCHESTRA study. Intensive Care Med. 2015;41:2149–60. https://doi.org/10.1007/s00134-015-4076-7.

[12] Wortel SA, de Keizer NF, Abu-Hanna A, Dongelmans DA, Bakhshi-Raiez F. Number of intensivists per bed is associated with efficiency of Dutch intensive care units. J Crit Care. 2021;62:223–9. https://doi.org/10.1016/j.jcrc.2020.12.008.

[13] Rothen HU, Takala J. Can outcome prediction data change patient outcomes and organizational outcomes? Curr Opin Crit Care. 2008;14:513–9. https://doi.org/10.1097/MCC.0b013e32830864e9.

[14] van de Klundert N, Holman R, Dongelmans DA, de Keizer NF. Data resource profile: the Dutch National Intensive Care Evaluation (NICE) registry of admissions to adult intensive care units. Int J Epidemiol. 2015;44:1850. https://doi.org/10.1093/ije/dyv291.

[15] Zampieri FG, Salluh JIF, Azevedo LCP, Kahn JM, Damiani LP, Borges LP, et al. ICU staffing feature phenotypes and their relationship with patients' outcomes: an unsupervised machine learning analysis. Intensive Care Med. 2019;45:1599–607. https://doi.org/10.1007/s00134-019-05790-z.

[16] Zampieri FG, Soares M, Borges LP, Salluh JIF, Ranzani OT. The Epimed Monitor ICU Database®: a cloud-based national registry for adult intensive care unit patients in Brazil. Revista Brasileira de Terapia Intensiva. 2017;29. https://doi.org/10.5935/0103-507X.20170062.

[17] Metnitz PGH, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3—from evaluation of the patient to evaluation of the intensive care unit. Part 1: objectives, methods and cohort description. Intensive Care Med. 2005;31:1336–44. https://doi.org/10.1007/s00134-005-2762-6.

[18] 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;34:1297–310. https://doi.org/10.1097/01.CCM.0000215112.84523.F0.

[19] Moreno RP, Metnitz PGH, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3—from evaluation of the patient to evaluation of the intensive care unit. Part 2: development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med. 2005;31:1345–55. https://doi.org/10.1007/s00134-005-2763-5.

[20] Finazzi S, Poole D, Luciani D, Cogo PE, Bertolini G. Calibration Belt for quality-of-care assessment based on dichotomous outcomes. PLoS One. 2011;6:e16110. https://doi.org/10.1371/journal.pone.0016110.

[21] Verburg IWM, de Jonge E, Peek N, de Keizer NF. The association between outcome-based quality indicators for intensive care units. PLoS One. 2018;13:e0198522. https://doi.org/10.1371/journal.pone.0198522.

[22] Spiegelhalter DJ. Funnel plots for comparing institutional performance. Stat Med. 2005;24:1185–202. https://doi.org/10.1002/sim.1970.

[23] Machado FR, Cavalcanti AB, Bozza FA, Ferreira EM, Angotti Carrara FS, Sousa JL, et al. The epidemiology of sepsis in Brazilian intensive care units (the Sepsis PREvalence Assessment Database, SPREAD): an observational study. Lancet Infect Dis. 2017;17:1180–9. https://doi.org/10.1016/S1473-3099(17)30322-5.

[24] Minne L, Eslami S, de Keizer NF, de Jonge E, de Rooij SE, Abu-Hanna A. Effect of changes over time in the performance of a customized SAPS-II model on the quality of care assessment. Intensive Care Med. 2012;38:40–6. https://doi.org/10.1007/s00134-011-2390-2.

[25] Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. Med Care. 2007;45:489–96. https://doi.org/10.1097/MLR.0b013e31803bb479.

[26] Soares M, Bozza FA, Angus DC, Japiassú AM, Viana WN, Costa R, et al. Organizational characteristics, outcomes, and resource use in 78 Brazilian intensive care units: the ORCHESTRA study. Intensive Care Med. 2015;41:2149–60. https://doi.org/10.1007/s00134-015-4076-7.

[27] Farrington DP, Loeber R. Some benefits of dichotomization in psychiatric and criminological research. Crim Behav Ment Health. 2000;10:100–22. https://doi.org/10.1002/cbm.349.

[28] Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ. 2006;332:1080.1. https://doi.org/10.1136/bmj.332.7549.1080.

[29] MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. Psychol Methods. 2002;7:19–40. https://doi.org/10.1037/1082-989X.7.1.19.

[30] Nuzzo RL. Making continuous measurements into dichotomous variables. 2019;11:1132–4. https://doi.org/10.1002/pmrj.12228.

[31] Dawson NV, Weiss R. Dichotomizing continuous variables in statistical analysis: a practice to avoid. Med Decis Mak. 2012;32:225–6. https://doi.org/10.1177/0272989X12437605.