J Crit Care 82 (2024) 154814

Contents lists available at ScienceDirect

Journal of Critical Care

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

Effect of mortality prediction models on resource use benchmarking of intensive care units

André Moser [a,*], Rahul Raj [b], Matti Reinikainen [c], Stephan M. Jakob [d], Jukka Takala [d]

[a] CTU Bern, Department of Clinical Research, University of Bern, Mittelstrasse 43, 3012 Bern, Switzerland
[b] Department of Neurosurgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[c] Department of Anaesthesiology and Intensive Care, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland
[d] University of Bern, Hochschulstrasse 4, 3012 Bern, Switzerland

ARTICLE INFO

Keywords:
Cost control
Health care benchmarking
Health resources
Hospital mortality
Intensive care unit
Resource allocation

ABSTRACT

Purpose: Intensive care requires extensive resources. The ICUs' resource use can be compared using standardized resource use ratios (SRURs). We assessed the effect of mortality prediction models on the SRURs.

Materials and methods: We compared SRURs using different mortality prediction models: the recent Finnish Intensive Care Consortium (FICC) model and the SAPS-II model ($n = 68{,}914$ admissions). We allocated the resources to severity of illness strata using deciles of predicted mortality. In each risk and year stratum, we calculated the expected resource use per survivor from our modelling approaches using length of ICU stay and Therapeutic Intervention Scoring System (TISS) points.

Results: Resource use per survivor increased from one length of stay (LOS) day and around 50 TISS points in the first decile to 10 LOS-days and 450 TISS in the tenth decile for both risk scoring systems. The FICC model predicted mortality risk accurately whereas the SAPS-II grossly overestimated the risk of death. Despite this, SRURs were practically identical and consistent.

Conclusions: SRURs provide a robust tool for benchmarking resource use within and between ICUs. SRURs can be used for benchmarking even if recently calibrated risk scores for the specific population are not available.

1. Introduction

Care of patients admitted to intensive care units (ICUs) requires extensive resources [1]. Resource use and outcomes of intensive care are relevant for health care planning and process optimization. The primary goal of care in ICUs is to achieve survival from life-threatening critical illness. The amount of resources used per survivor varies considerably between ICUs and different health care systems [2,3] and increases exponentially with increasing severity of illness [2]. Therefore, comparing and benchmarking ICU costs should take into account the risk of death. Standardized resource use ratios (SRURs) have been used to compare severity adjusted resource use to achieve survivors between ICUs.

The SRURs indicate the observed to expected ratio of resources used per survivor for each ICU, adjusted for the severity of illness of each patient [2,3]. If expected resource use is calculated for each year and risk stratum, then for each year the SRURs of ICUs vary. When the admissions from all ICUs are combined the SRURs equal one, because the overall resource use is distributed over all yearly risk strata. If the overall resource use of all admissions over time is averaged instead of stratification by year, the variation of SRURs over time for individual ICUs and the cohorts of all admissions in each year can be evaluated.

For SRURs, severity of illness scores, such as SAPS (simplified acute physiology score II [4]) are used to allocate resource use and patients into severity of illness strata. It is well known that mortality prediction models require regular updates to provide accurate risk predictions and have limited transportability [5]. Thus, the mortality prediction model used for the stratification of resource allocation and for the mortality prediction of each patient may influence the SRUR. The impact of such interaction is not known. Validated and updated mortality prediction models may not be available for most ICUs, and this may present a limitation for the use of SRURs.

We therefore investigated the impact of different mortality risk scoring systems and expected resource use modelling approaches on the SRUR.

* Corresponding author.
E-mail address: andre.moser@unibe.ch (A. Moser).

https://doi.org/10.1016/j.jcrc.2024.154814

Available online 20 April 2024

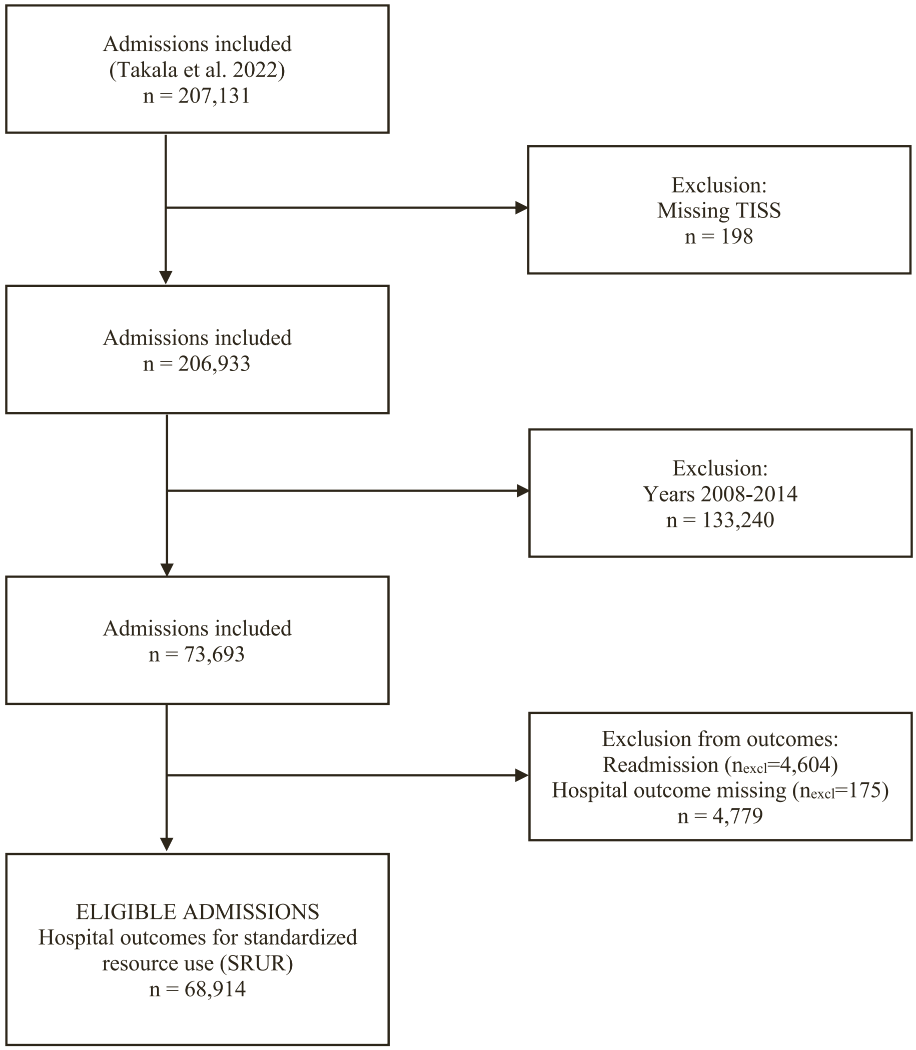

Fig. 1. Flow chart.

2. Methods

2.1. Data source

At the time of data collection, the Finnish Intensive Care Consortium (FICC) consisted of all adult ICUs in Finland except one neurosurgical ICU (21 ICUs, one declined participation and one cardiac ICU was excluded), and the ICUs of one university hospital in Estonia (100% of university ICU admissions and 16% all ICU admissions in Estonia) and one in Switzerland (33% of all university hospital ICU admissions and 11% of all ICU admissions in Switzerland) [3]. Anonymized data were extracted from a benchmarking database. This study is a secondary analysis of data used for a previous publication [3].

2.2. Study population

This secondary analysis used the study population described by Takala et al. [3]. In brief, Takala et al. used data on 207,131 admissions from between 2008 and 2017 from the Finnish Intensive Care Consortium (FICC) database. For the current analysis, we excluded admissions with missing Therapeutic Intervention Scoring System score (TISS score [6]; a measure of resource use) and restricted the analysis period from 2015 to 2017 to be consistent with the FICC mortality prediction model. For hospital survivor counts, readmissions and admissions with missing hospital outcomes were excluded (Fig. 1).

2.3. Study outcomes

Our primary study outcome was SRURs using two different mortality prediction scores and two different expected resource use modelling approaches. SRURs were defined as the observed resource use divided by the expected resource use, stratified by mortality risk scoring, as described in detail below. Resource use was measured by 1) length of stay in days (LOS) or 2) the Therapeutic Intervention Scoring System-76 (TISS) Score [6]. TISS ranks ICU activities according to intensity and resource use.

2.4. Mortality risk

We calculated the individual probability of death from the FICC mortality prediction model described in Moser et al. [7] and from the original SAPS II model [4]. Supplemental Figs. 1 and 2 show calibration plots of the two mortality scoring systems using the underlying study population.

2.5. SRUR calculations

We allocated the resources to severity of illness strata using deciles of predicted mortality from the FICC mortality prediction model and the original SAPS II model. In each risk and year stratum, we calculated the expected resource use from our modelling approaches (as LOS and TISS)

Table 1
Admission characteristics of the study population (N=68,914).

		n (%) / Mean (SD)
Age (years)		60 (18)
Gender		
	Male	42,823 (62%)
	Female	26,082 (38%)
	Missing	9 (<0.1%)
Intensive care unit category		
	University	53,234 (77%)
	Non-university (large)	9630 (14%)
	Non-university (small)	6050 (8.8%)
Treatment type		
	Emergency	48,806 (71%)
	Elective	20,108 (29%)
Diagnosis		
	Non-surgical	36,656 (53%)
	Surgical	32,258 (47%)
SAPS-II score		34 (17)
TISS score sum		120 (197)
Length of stay (days)		2.9 (5.0)
Outcome in hospital		
	Survivor	62,032 (90%)
	Non-survivor	6882 (10%)
Admission year		
	2015	22,521 (33%)
	2016	23,082 (33%)
	2017	23,311 (34%)

Abbreviations: LOS Length of stay; SAPS Simplified acute physiology score; SD Standard deviation; TISS Therapeutic Intervention Scoring System-76 Score.

per survivor. For each ICU, the expected resource use for each mortality risk decile was calculated as (number of survivors)*(expected LOS/TISS per survivor). The expected total resource use for each individual ICU is the sum of expected resource use in all mortality risk deciles The $SRUR_{LOS}$ and $SRUR_{TISS}$ using LOS and TISS were calculated as the ratio of observed to expected total resource use.

2.6. Statistical methods

We describe the study population and resource use by frequencies (n), percentages (%), mean and standard deviation (SD). We modelled the expected absolute resource use (LOS or TISS) in deciles of predicted mortality from FICC and SAPS-II. We used a Gamma log link generalized linear model as primary analysis. We modelled the risk deciles as predictor (without year information) because this approach pools the modelled expected resource use over all three years. We call this approach the "pooled" approach. We report modelled absolute resource use and SRURs with 95% confidence intervals (CIs). We compared the distribution of resource use between different approaches using a Mann-Whitney test. Funnel plots were constructed reporting 95% and 99% control limits using a normal approximation. All analyses are performed in R version 4.2.3 [8].

2.7. Sensitivity analysis

As sensitivity analysis we modelled an interaction effect between risk deciles and years. We call this modelling approach "annual" because it models resource use for all risk strata within each year.

3. Results

3.1. Study population

The admission characteristics are shown in Table 1. Most of the admissions are treated in university ICUs (77%) and are predominantly emergency (71%) admissions. The mean SAPS-II score is 34 (SD 17) with a mean length of stay of 2.9 days (SD 5 days).

Table 2 describes the study population (68,914 admissions with 62,032 survivors from 21 ICUs combined for the years 2015–2017) and resource use per survivor, by mortality risk deciles. Resource use per survivor increased exponentially from 1 LOS-day per survivor in the first decile to 10 LOS-days per survivor in the tenth decile, and from around 50 TISS points per survivor (first decile) to 450 TISS points per survivor (tenth decile), for both risk scoring systems.

3.2. SRUR variation

Funnel plots for $SRUR_{LOS}$ (Panel A: FICC; Panel B: SAPS-II) and

Table 2
Population numbers and resource use per survivor, by mortality risk deciles.

Mortality prediction based on the FICC model and observed mean resource use per survivor					
Risk decile	Predicted mortality in % (SD)	Total population (N = 68,914)	Survivors (N = 62,032)	LOS per survivor	TISS per survivor
1	0.2 (0.1)	7116	7105	1.2	44
2	0.5 (0.1)	7064	7043	1.5	62
3	0.9 (0.1)	7041	6998	1.8	76
4	1.5 (0.2)	6979	6896	2.1	87
5	2.4 (0.3)	6935	6807	2.6	105
6	3.8 (0.5)	6782	6543	3.3	131
7	6.3 (1.0)	6732	6307	4.4	172
8	11.1 (1.9)	6691	5944	5.5	220
9	21.2 (4.4)	6711	5126	7.8	317
10	50.9 (16.6)	6863	3263	10.6	462
Mortality prediction based on the SAPS-II model and observed mean resource use per survivor					
Risk decile	Predicted mortality in % (SD)	Total population (N = 68,914)	Survivors (N = 62,032)	LOS per survivor	TISS per survivor
1	1.3 (0.6)	8601	8581	1.4	51
2	3.2 (0.5)	7393	7352	1.7	68
3	4.8 (0.4)	5280	5223	1.9	79
4	6.9 (0.7)	7397	7270	2.3	92
5	10.1 (1.1)	7749	7515	2.7	110
6	14.6 (1.4)	6116	5844	3.3	132
7	21.2 (2.4)	6480	5994	4.3	171
8	32.4 (4.2)	6533	5741	5.6	222
9	52.1 (7.2)	6684	5116	7.4	305
10	81.0 (9.7)	6681	3396	10.0	443

Abbreviations: LOS Length of stay; SD Standard deviation; TISS Therapeutic Intervention Scoring System-76 Score.
Ties in predicted mortality risk result in variation in patient numbers per decile.

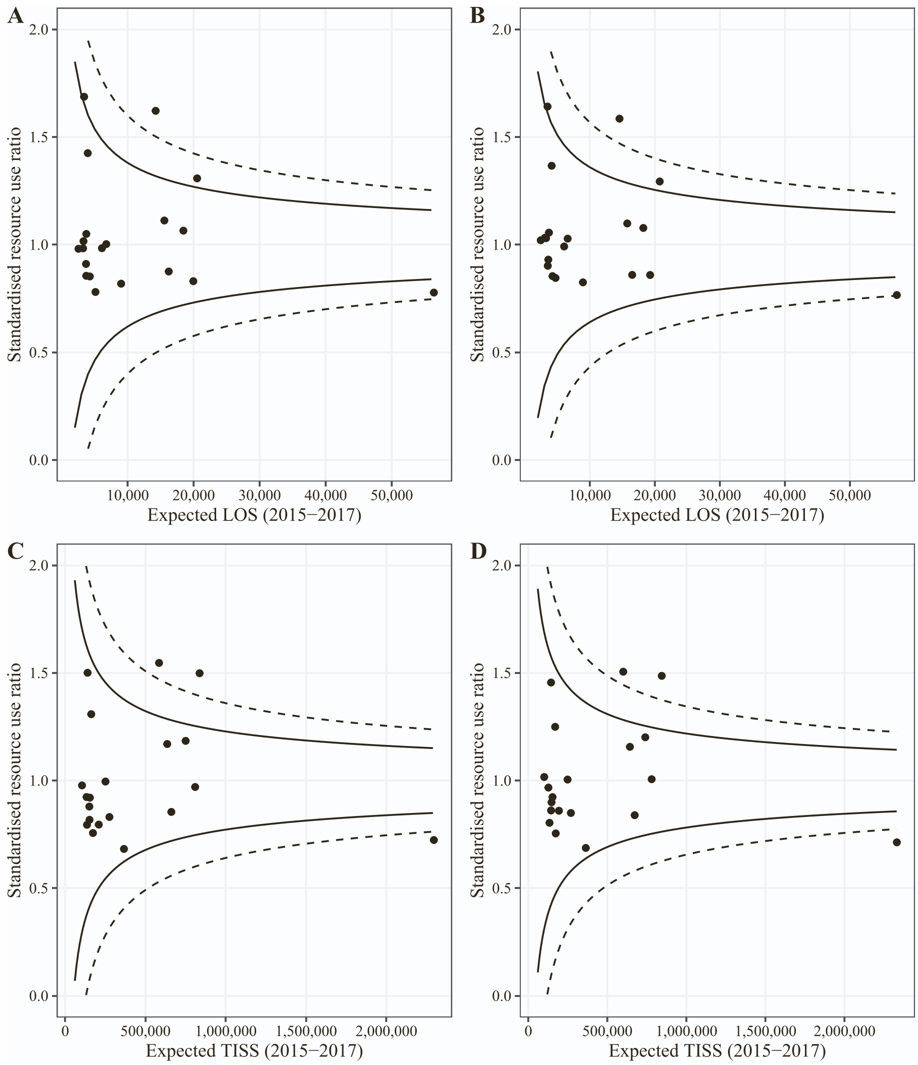

Fig. 2. Funnel plots of standardized resource use ratios (Panels A/C: FICC, Panels B/D: SAPS-II).

$SRUR_{TISS}$ (Panel C: FICC; Panel D: SAPS-II) indicate that most of the ICUs are within the 95% and 99% control limits (Fig. 2). Fig. 3 shows boxplots of SRUR on ICU level. In general, variation remains similar comparing the FICC model and SAPS-II (pairwise *p*-values from Mann-Whitney test >0.70, Supplemental Table 1). Similar variation with the FICC model and SAPS-II was also observed in data of individual ICUs at each risk decile (Fig. 4). Supplemental Fig. 3 shows SRUR estimates with 95% CIs for the whole cohort of eligible admissions. For the pooled approach, SRUR varied from 0.99, 95% CI (0.97–1.01) in 2015 to 0.995, 95% CI (0.98–1.02) in 2017 for $SRUR_{LOS}$ and from 0.99, 95% CI (0.97–1.01) in 2015 to 1.01, 95% CI (1.00–1.04) in 2017 for $SRUR_{TISS}$ using FICC mortality predictions. Results for the SAPS-II model were similar. In the annual approach (sensitivity analysis), the SRURs of all eligible admissions for each year equal one, because SRURs for each year apply the annual resource use and outcome data. In the pooled approach, yearly SRURs for the consortium are not equal to one, because they deviate from the three-year average. Supplemental Fig. 4 shows individual SRUR estimates for ICUs and Supplemental Fig. 5 shows the effect of the used prediction model on individual SRURs estimates for ICUs. The modelled predicted absolute resource use was not different using the pooled approach, with overlapping confidence intervals at each risk decile (Supplemental Figs. 6).

3.3. Sensitivity analysis

The modelled predicted absolute resource use was not different between the pooled and the annual approach, with overlapping confidence intervals at each risk decile (Supplemental Figs. 7–8). Supplemental Fig. 9 shows the distributions of SRUR on ICU level for the annual approach.

4. Discussion

4.1. Summary of main findings

We compared the impact of two different risk scoring systems, the recently developed and validated FICC mortality prediction model and the original SAPS-II model, on SRUR. We used two modelling approaches for expected resource use and two indicators of resource use (LOS and TISS). The FICC prediction model predicted mortality risk accurately whereas the SAPS-II grossly overestimated the risk of death, especially in the high-risk strata. Despite these major differences in mortality risk predictions, the results on SRURs were practically identical and consistent with both LOS and TISS. We conclude that the SRURs provide robust tools for the investigation of resource use variation and reliable SRUR benchmarking results may be achieved even when old risk prediction models with poor calibration are used. This results from grossly similar distribution of patients and resources to the

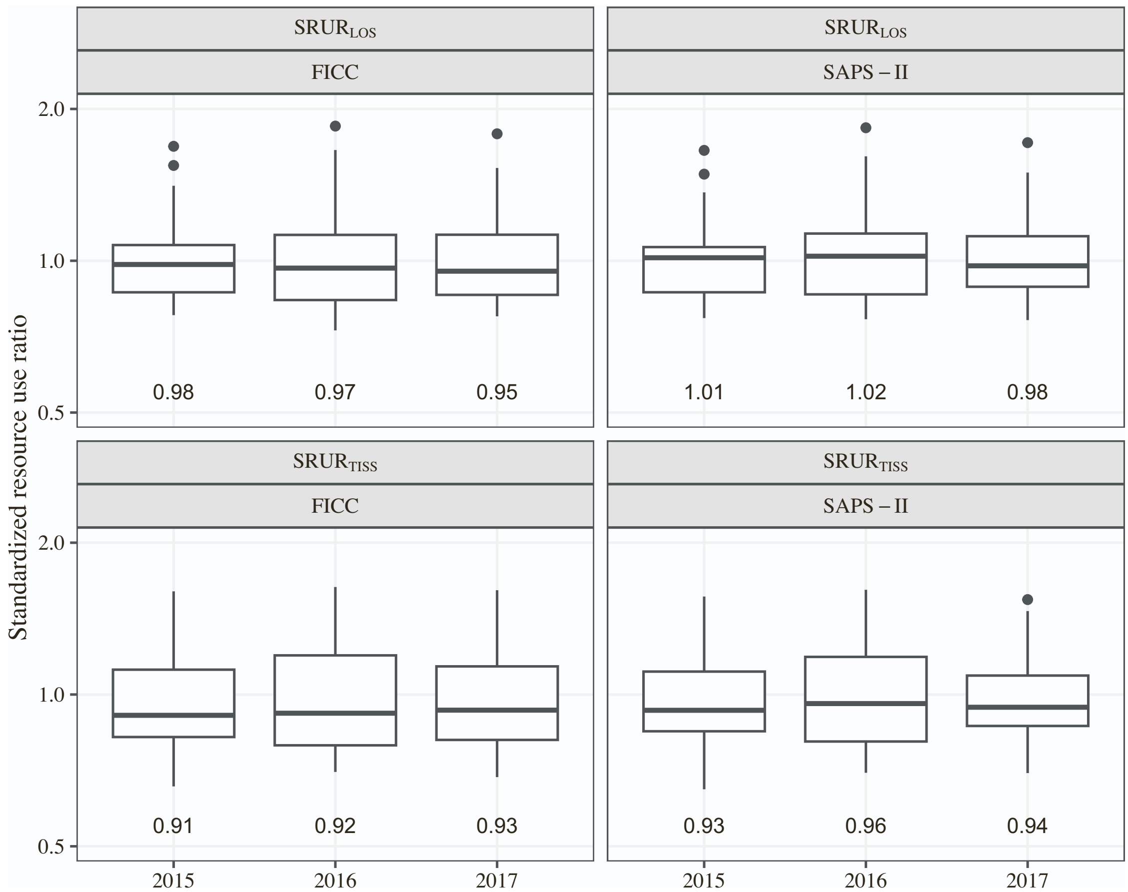

Fig. 3. Boxplots of standardized resource use ratios on ICU level using LOS (top) and TISS (bottom).

risk strata. Still, the use of a well calibrated mortality prediction model is advisable especially when SRUR and standardized mortality ratio (SMR) are used together in benchmarking [3].

4.2. Relation to other studies

Previous studies on SRURs have used either SAPS-3 [2,9] or SAPS-II score strata [3,10,11] to adjust the resource use for the severity of illness. None of the previous studies have addressed the potential relevance of the tool chosen to allocate resources according to the severity of illness. All mortality prediction tools tend to lose their validity over time. This will influence the allocation of resources to severity strata and may thus modify the SRUR. In the present study, we stratified the severity of illness according to risk deciles instead of the severity score strata. The FICC prediction model was well calibrated, in contrast to the overestimated risks predicted by the original SAPS-II. Despite these major differences in the calibration properties, the results on SRURs were practically identical and so was the modelled expected total resource use according to risk of death deciles. Furthermore, the choice of resource use indicator (LOS vs TISS) had no relevant impact on the results. Accordingly, the concept of SRUR is robust enough to be used even if no recently validated mortality prediction model is available, and LOS is the only available indicator of resource use. If the associations of SRUR and SMR are of interest [2,3,9], accurate prediction of mortality with validated and transportable models is necessary. We found no indication for variation of SRURs across risk deciles.

4.3. Strengths and limitations

Our study has several strengths. First, we used the established scoring system SAPS-II besides a risk prediction model which was recently developed using data from the FICC. This allowed us to investigate the impact of a poor calibration of the SAPS-II in the current patient population with a newly developed risk score with good calibration and transportability properties on the SRUR. Second, the multinational database allowed an investigation of SRUR for different health care systems across Finland, Estonia and Switzerland. Since the variation of SRURs highly depend on the underlying health care system processes (for example, health care planning and optimization of care) this multinational characteristics of FICC highlights the robustness of SRUR as a benchmarking tool. Third, the included hospitals covered a wide range of hospital typologies which addresses hospital-related factors (for example, case-mix and volume-driven factors) into the calculation of SRUR.

Our study has some limitations. The study sample 2015–2017 precedes the COVID-19 epidemics. The use and allocation of ICU resources changed dramatically during the epidemics. The long-term effects of such changes are not known and the follow-up is short. Whether the study period represents current ICU resource use can therefore not be confirmed. If the criteria for admission to and discharge from ICUs have not permanently changed, our results on the impact of the mortality prediction models on SRUR should still be valid. The use of LOS and TISS as indicators of resource use can be criticized. Nevertheless, we have previously shown that this approach gives similar results to use of direct costs of care (including salaries, drugs, fluids, and disposables) [3]. This approach does not take into consideration all resources provided by other hospital services The consistency of the results using FICC- or SAPS-II stratification may be influenced by the case-mix. The case-mix in our study is representative of a general ICU population in three countries. Substantially different case-mixes may well lead to poorer correlation between the original SAPS-II and more up to date prediction models. We therefore suggest the use of SAPS-II based approach only in

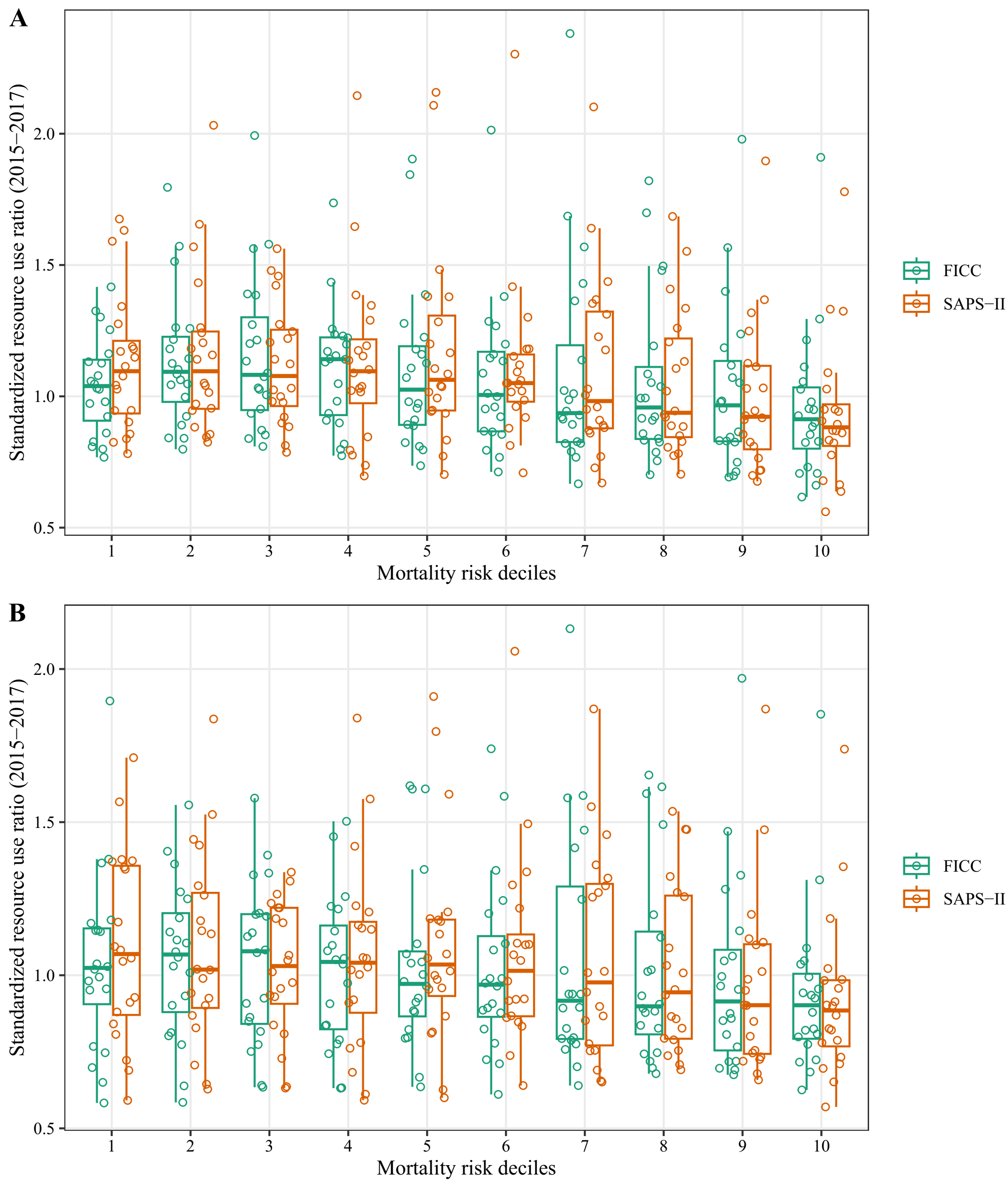

Fig. 4. Standardized resource use ratios and boxplots, by mortality risk deciles (Panel A: LOS, Panel B: TISS).

the absence of an updated prediction model.

5. Conclusions

SRURs provide a robust tool for benchmarking resource use within and between ICUs. SRURs can be used for benchmarking even if recently calibrated risk scores for the specific population are not available.

Ethics approval

The study protocol, database contents and data management process of the original study were approved by the National Institute of Health and Welfare, Finland (Decision THL/1524/5.05.00/2017 and THL/1173/05.00/2018). According to regulations in Finland, Estonia, and Switzerland, no ethics committee approval was needed.

Funding

Open Access funding provided by Universität Bern. The study was supported by departmental funds of the Department of Intensive Care Medicine, Bern University Hospital, University of Bern, Bern, Switzerland. There was no external funding for this project. Dr. Raj has received research grants from Finska Läkaresällskapet, Svenska Kulturfonden and Medicinska Understödsföreningen Liv & Hälsa.

CRediT authorship contribution statement

André Moser: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Rahul Raj:** Data curation, Funding acquisition, Project administration, Resources, Writing – review & editing. **Matti Reinikainen:** Funding acquisition, Project administration, Resources, Writing – review & editing. **Stephan M. Jakob:** Funding acquisition, Project administration, Resources, Writing – review & editing. **Jukka**

Takala: Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal analysis, Funding acquisition.

Declaration of competing interest

None.

Data availability

The authors had permission from FINDATA (Social and Health Data Permit Authority) to analyse the data. A secondary use of the data for other researchers can be obtained through FINDATA (https://findata.fi/en/) according to the Finish Secondary Data Act.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcrc.2024.154814.

References

[1] Hirshberg EL, Wilson EL, Stanfield V, Kuttler KG, Majercik S, Beesley SJ, et al. Impact of critical illness on resource utilization: a comparison of use in the year before and after ICU admission*. Crit Care Med 2019;47:1497–504. https://doi.org/10.1097/CCM.0000000000003970.

[2] 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.

[3] Takala J, Moser A, Raj R, Pettilä V, Irincheeva I, Selander T, et al. Variation in severity-adjusted resource use and outcome in intensive care units. Intensive Care Med 2022;48:67–77. https://doi.org/10.1007/s00134-021-06546-4.

[4] Le Gall J-R, Lemeshow S, Saulnier F. Simplified Acute Physiology Score (SAPS II) Based on a European / North American multicenter study. JAMA 1993;270:2957–63. https://doi.org/10.1001/jama.270.24.2957.

[5] Cox EGM, Wiersema R, Eck RJ, Kaufmann T, Granholm A, Vaara ST, et al. External validation of mortality prediction models for critical illness reveals preserved discrimination but poor calibration. Crit Care Med 2023;51:80–90. https://doi.org/10.1097/CCM.0000000000005712.

[6] Keene AR, Cullen DJ. Therapeutic intervention scoring system: update 1983. Crit Care Med 1983;11:1–3. https://doi.org/10.1097/00003246-198301000-00001.

[7] Moser A, Reinikainen M, Jakob SM, Selander T, Pettilä V, Kiiski O, et al. Mortality prediction in intensive care units including premorbid functional status improved performance and internal validity. J Clin Epidemiol 2022;142:230–41. https://doi.org/10.1016/j.jclinepi.2021.11.028.

[8] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021 [n.d].

[9] 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.

[10] Takala J, Moser A, Reinikainen M, Varpula T, Raj R, Jakob SM. Impact of cardiac surgery and neurosurgery patients on variation in severity-adjusted resource use in intensive care units. J Crit Care 2022;71:154110. https://doi.org/10.1016/j.jcrc.2022.154110.

[11] Raj R, Moser A, Starkopf J, Reinikainen M, Varpula T, Jakob SM, et al. Variation in severity-adjusted resource use and outcome for neurosurgical emergencies in the intensive care unit. Neurocrit Care 2023. https://doi.org/10.1007/s12028-023-01723-3.