International Journal of Medical Informatics 191 (2024) 105568

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Comparing causal random forest and linear regression to estimate the independent association of organisational factors with ICU efficiency

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

[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, Digital Health and Methodology, Amsterdam, the Netherlands
[e] Department of Intensive Care, Amsterdam UMC, Amsterdam, the Netherlands
[f] D'Or Institute for Research and Education (IDOR), Rio de Janeiro, RJ, Brazil
[g] Brazilian Research in Intensive Care Network (BRICNet), Brazil
[h] PostGraduate, Internal Medicine, Program Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil
[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 unit
Organisational characteristics
Causal random forest
Efficiency
Causal inference

A B S T R A C T

Purpose: Parametric regression models have been the main statistical method for identifying average treatment effects. Causal machine learning models showed promising results in estimating heterogeneous treatment effects in causal inference. Here we aimed to compare the application of causal random forest (CRF) and linear regression modelling (LRM) to estimate the effects of organisational factors on ICU efficiency.

Methods: A retrospective analysis of 277,459 patients admitted to 128 Brazilian and Uruguayan ICUs over three years. ICU efficiency was assessed using the average standardised efficiency ratio (ASER), measured as the average of the standardised mortality ratio (SMR) and the standardised resource use (SRU) according to the SAPS-3 score. Using a causal inference framework, we estimated and compared the conditional average treatment effect (CATE) of seven common structural and organisational factors on ICU efficiency using LRM with interaction terms and CRF.

Results: The hospital mortality was 14 %; median ICU and hospital lengths of stay were 2 and 7 days, respectively. Overall median SMR was 0.97 [IQR: 0.76,1.21], median SRU was 1.06 [IQR: 0.79,1.30] and median ASER was 0.99 [IQR: 0.82,1.21]. Both CRF and LRM showed that the average number of nurses per ten beds was independently associated with ICU efficiency (CATE [95 %CI]: −0.13 [-0.24, −0.01] and −0.09 [-0.17,-0.01], respectively). Finally, CRF identified some specific ICUs with a significant CATE in exposures that did not present a significant average effect.

Conclusion: In general, both methods were comparable to identify organisational factors significantly associated with CATE on ICU efficiency. CRF however identified specific ICUs with significant effects, even when the average effect was nonsignificant. This can assist healthcare managers in further in-dept evaluation of process interventions to improve ICU efficiency.

* 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.ijmedinf.2024.105568
Received 20 October 2023; Received in revised form 4 July 2024; Accepted 26 July 2024
Available online 27 July 2024

1. Introduction

Benchmarking intensive care units (ICUs) provides critical care practitioners and administrators insights for potential improvements in the process of care and outcomes [1–4], which is valuable especially in limited-resource settings [5,6]. To implement setting-adapted quality improvement initiatives, it is crucial to understand which organisational characteristics affect the outcomes and performance in an ICU. Risk-adjusted indicators have been used to measure the efficiency of an ICU: the standardised mortality ratio (SMR) and the standardised resource use (SRU) [7–10]. Furthermore, previous studies have shown the associations between several organisational factors and ICU outcomes in various ways [7–12].

Confounding adjustment is commonly performed with propensity score-based methods or by directly including variables in parametric regression models. Recently tree-based machine learning methods have been proposed to enhance confounder adjustment [13,14]. One promising approach is Causal Random Forest (CRF), which consists of a Random Forest-based method [13,15–17]. As a tree-based model, CRF employs a data partitioning procedure similar to decision trees (like the CART algorithm). However, in CRF, each underlying tree, referred to as a "causal tree," is dedicated to learning the heterogeneity of treatment effects rather than classifying the dependent variable [18]. The concept behind CRF is to identify leaves within these trees that represent combinations of covariates where the treatment effect remains constant but differs significantly from other leaves. These subgroups hence represent the heterogeneous treatment effect [13,15]. Decision tree approaches do not require the pre-specification of the model form and allow for non-parametric estimation of the functional form between response and covariates. These characteristics offer clear advantages over linear regression modelling. Aside from the heterogeneity of treatment effects, a CRF provides unbiased estimates of the average treatment effect (ATE) and confidence intervals compared to previous tree-based models [13,15].

Understanding the measured effect of organisational factors on ICU performance is essential to identify targets for improvement. Little is known about the applications of CRF to estimate treatment effect in ICU benchmarking. Our goal was to demonstrate the application of CRF for estimating the effects of organisational factors in ICU performance and compare it with linear regression modelling in order to understand its merits.

2. Methods

2.1. Study design and data source

This study is a retrospective analysis on a large dataset of Brazilian and Uruguayan ICUs from the "Organisational CHaractEriSTics in cRitical cAre" (ORCHESTRA) network [8,12,19]. This prospectively collected data set contains demographic and clinical information, use of resources, and outcomes of adult patients (18 years and older) 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 [8,12]. A detailed description of ORHCESTRA's data collection and the ICU Brazilian settings was described elsewhere [8,9,12,20]. Additional information on critical care settings in Brazilian and Uruguayan are in the supplementary methods (sMethods). Information was retrieved from the Epimed Monitor System® (Epimed Solutions®, Rio de Janeiro, Brazil) [19]. The Brazilian National Ethics Committee (CAAE: 19687113.8.1001.5249) and the Ethics committee of the *Hospital Maciel, Montevidéo*, Uruguay (protocol n<u>o</u> 20/2017) approved the study and waived the need for informed consent. All data was anonymised.

2.2. Data collection

Patient and ICU level data were included. Patient data consisted in patient's demographics, the severity of illness at admission using the Simplified Acute Physiology Score 3 (SAPS 3), the in-hospital mortality, and ICU length-of-stay in days (defined as 24-hour periods). Patients were excluded based on the Simplified Acute Physiology Score 3 (SAPS-3) exclusion criteria [21]. We excluded ICUs with more than 10 % of admissions with missing admission diagnosis or hospital outcomes. At the organisational level, information on hospital and ICU organization included the number of hospital and ICU beds, the type of ICU (surgical, oncological or neurological ICU), the presence of certified training programs in critical care, and ICU staffing patterns, such as the average number of physicians (physician/bed) and nurses per 10 ICU beds (nurse/bed) ratios in each shift (considering day and night and excluding trainees and residents), the presence of full-time board-certified intensivists and the average annual ICU occupancy rate. More details on the sMethods.

2.3. Outcomes

Our primary outcome was ICU efficiency expressed as the average standardised efficiency ratio (ASER) [22]. The ASER consists of a continuous metric derived from the combination of the Standardised Mortality Ratio (SMR) and the Standardised Resource Use (SRU) calculated as (SMR+SRU) / 2). We chose ASER as a general continuous efficiency metric since it presents more adequate statistical properties for estimating the effect of organisational factors compared to the efficiency matrix approach [7], especially when SMR and SRU are positively correlated [22].

The SMR is a metric of risk-adjusted mortality and corresponds to the ratio of observed deaths to expected deaths. The expected number of deaths was obtained by adding mortality probabilities from a recalibrated SAPS-3 risk model [21] (**Figure S1**). The mortality risks were recalibrated using first-level customisation to reduce the potential over- or underestimation of the predicted risks [23].

The SRU measures the risk-adjusted use of resources as to the observed-to-expected resource use ratio, and we considered the ICU LOS as a proxy measure of resource use [7]. Expected LOS was calculated with the average LOS per survivor for each decile of the recalibrated mortality probability.

2.4. Data analysis

Data was decribed using median and interquartile range (IQR) for continuous variables and absolute frequency and proportions for categorical variables. The Spearman's rho correlation test and a Wilcoxon Rank Sum test were used to evaluate univariable associations between each organisational variable with the efficiency measure.

We evaluated the association of seven organisational factors with ICU efficiency (ASER), following a causal inference framework [18,24]. Based on our clinical expertise, we included the number of hospital beds, the presence of certified training programs in critical care, the full-time coverage of a certified intensivist, the nurse/bed and physician/bed ratios per shift as exposure variables. For each organisational variable of interest, the process for confounder identification followed a combined approach of using knowledge of specialists and statistical methods, such as the Change-In-Estimate [25,26]. The annual ICU occupancy rate (average occupancy rates in the three years of data collection) and the admissions per bed ratio were also considered as potential confounding variables.

The conditional average treatment effects (CATE) were estimated using the Causal Random Forest (CRF) and, as a baseline, the linear regression model (LRM). The CRF is a tree-based machine learning algorithm [13,15] that can estimate heterogeneous treatment effects. Similar to the Random Forest method, CRF is an ensemble of "causal" trees – decision trees built to maximise the heterogeneity of treatment effects [18]. As a tree-based model, causal trees are non-parametric and their branches correspond to a set of decision rules that represent to the interaction among variables during the partitioning process, without pre-specification or enumerating combinations, thus allowing for estimating heterogeneous treatment effect (HTE). In the LRM, confounder adjustment was performed by including the exposure and confounders as independent variables in the model. To account for HTE, interaction terms as the combination of the exposure and each confounder was included in the model. We estimated the CATE for each exposure in the LRM as the average marginal effects [27]. For both methods, the ASER was the outcome variable. Detailed methods and implementation information are described in **sMethods**.

The CATE estimated for each organisational variable were compared between models. In addition, using a trained CRF, it is possible to obtain a prediction of the CATE for an observation [15,28,29]. Therefore, we obtained the CATEs, and their corresponding 95 % confidence intervals, to explore the individual potential effects at the unit level for each ICU when analysing each exposure variable.

In a sensitivity analysis, extremely good- or worst-performing units (with extreme values of ASER) that may influence the results were removed. Those outlying units were identified using the z-score method with two standard deviations. Then, the mortality probabilities, efficiency measures, and average treatment effects for each organisational factor re-estimated using the linear regression model and the CRF. We also used those results to evaluate the pattern of the associations identified between both models, such as non-linearities. All analyses were performed in R 4.3.1.

3. Results

We analysed 277,459 patients admitted to 128 participating ICUs in 77 hospitals in Brazil and Uruguay (Table 1). The median age was 65 years [IQR: 49, 78], with 58.7 % being older than 60 years old, 51 % were women, and most of the admissions were medical (66 %) (Table 1). The overall median SMR was 0.97 [IQR: [0.76, 1.21], median SRU was 1.06 [IQR: 0.79, 1.30], and the median ASER was 0.99 [IQR: 0.82, 1.21] (Table 2). The SMR and SRU for each ICU are shown in Fig. 1 and their distributions are shown in the **supplemental figure - Figure S2**. There was a mild correlation between the SMR and SRU (Spearman's rho correlation coefficient: 0.54, Fig. 1).

Regarding the organisational characteristics of the included ICUs, the median physician/bed ratio per shift was 1.5 [IQR: 1.33, 1.88], and the median nurse/bed ratio was 1.77 [IQR: 1.43, 2.50]. Of the 128 ICUs, 105 (88 %) were medical-surgical ICUs, and 40 (31 %) were covered full-time by a certified intensivist (details on the distribution of organisational variables are in **Figure S3**).

To measure the association of organisational factors (exposures) with efficiency (ASER), we defined the variables for confounder adjustment (**supplemental Table S1**). The exposure variable ICU beds to hospital (ICU/hospital) beds ratio had the least number of confounders (N = 2), while the nurse/bed ratio had the most (N = 7). There was a negative correlation between the nurse/bed ratio (−0.20, p = 0.01) and the number of admissions per bed (−0.52, p < 0.01) with the ASER, and ICUs with training programs in critical care showed increased ASER values compared to those without (**Figure S4**).

CRF and LRM methods showed comparable results when estimating the average treatment effect (Table 3). Although both approaches showed an association between the nurse/bed ratio and the ICU efficiency, the CRF's CATE presented a higher magnitude and narrower confidence intervals compared to the LRM (CATE [95 % CI]: CRF, −0.13 [-0.24, −0.03] vs LRM, −0.10 [-0.18, −0.02]). Among them, only the CRF model showed an association between training programs in critical care with increased ASER (CATE [95 % CI]: 0.19 [0.06, 0.33]). In the sensitivity analysis where extremely good- or worst-performing units were removed, both models' results were comparable and agreed upon regarding identified associations.

In addition, we obtained CATE estimates (predicted individual treatment effects) for each ICU using the CRF model (Fig. 2). There was a high variability in estimates among ICUs for the nurse/bed ratio, with 64 (50 %) units presenting a significant treatment effect (confidence intervals not overlapping the reference line at zero). Although the average effect was not significant for the number of hospital beds, ICU/hospital beds ratio and presence of training program in critical care, we observed that 17 (13 %), 8 (6 %) and 2 (1 %) units, respectively, presented a significant effect for those variables.

Table 1
Characteristics and outcomes of critically ill patients in 128 intensive care units.

Characteristics and Outcomes	Brazil and Uruguay
Total number of patients	277,459
Age (years), median (IQR)	65 (49, 78)
16–30, N (%)	19,487 (7 %)
31–40	25,798 (9.3 %)
41–50	28,144 (10 %)
51–60	41,238 (15 %)
61–70	52,472 (19 %)
71–80	52,960 (19 %)
81–90	44,918 (16.2 %)
> 90	12,442 (4.5 %)
Gender, N (%)	
Female	140,529 (51 %)
Male	136,790 (49 %)
Unknown/Transgender	140 (<0.1 %)
Admission type, N (%)	
Medical	183,561 (66 %)
Elective surgery	74,610 (27 %)
Urgent surgery	19,288 (7 %)
Simplified Acute Physiology Score (SAPS-3)	43 (34, 54)
SAPS-3 Predicted mortality risk	0.09 (0.03, 0.24)
Outcomes	
ICU LOS (days), median (IQR)	2 (1, 5)
ICU mortality, N (%)	24,470 (8.8 %)
Hospital LOS (days), median (IQR)	7 (4, 15)
Hospital mortality, N (%)	37,557 (14 %)

ICU: Intensive Care Unit; LOS: length-of-stay.
IQR: Interquartile Range (1st Quartile – 3rd Quartile).
SD: Standard Deviation.

4. Discussion

This study compared the application of CRF and LRM for estimating conditional average treatment effects under the causal inference framework. Analyses using a sample of 128 Brazilian and Uruguayan ICUs showed that CATE estimates were comparable in both methods, identifying two (nurse/bed ratio and training programs in critical care) out of seven exposure variables with significant independent effects. Furthermore, despite the nonsignificant average effect in some exposure variables, we could identify specific ICUs that could be affected by the organisational variable using the CRF method.

In this study, we introduced the CRF to estimate the CATE of organisational factors on ICU efficiency and obtain insights into the heterogeneity of treatment effects. To the best of our knowledge, only few studies compared CATE (or ATE) derived from CRF versus linear regression, and this is the first application of CRF to the analysis of ICU organisational data. This causal machine learning technique has received some attention recently as it provides confounder adjustment similar to stratification procedures, as the tree-building process assesses and includes potential interactions among confounding variables allowing for the identification of effects in specific subgroups [13]. Our results showed that both CRF and LRM with interactions were comparable regarding the exposures with significant CATE, even in the sensitivity analysis. However, CRF identified some ICUs affected by the exposure variable, even though the average effect was nonsignificant. The structure of CRF as a tree-based model that aims to estimate heterogeneous treatment effects allows the estimation of CATE for an individual observation, in our case, an ICU (a "prediction" of the treatment effect) [15], a feature that is limited in regression due to its modelling

Table 2
Outcome measures and organisational characteristics of all Intensive Care Units.

Characteristics and Measurements	Total	First Quartile (ASER<0.82)	Second Quartile (ASER: 0.82 – 0.99)	Third Quartile (ASER: 0.99 – 1.21)	Fourth Quartile (ASER >= 1.21)
Number of units	128	32	32	32	32
Outcome measures, median [IQR]					
Standardised mortality ratio (SMR)	0.97 (0.76, 1.21)	0.73 (0.66, 0.83)	0.79 (0.71, 1.00)	1.07 (0.97, 1.19)	1.32 (1.21, 1.50)
Standardised resource use (SRU)	1.06 (0.79, 1.30)	0.68 (0.62, 0.75)	0.95 (0.84, 1.10)	1.10 (1.00, 1.24)	1.49 (1.40, 1.69)
Averaged standardised efficiency ratio (ASER) [a]	0.99 (0.82, 1.21)	0.73 (0.66, 0.78)	0.90 (0.85, 0.95)	1.10 (1.03, 1.17)	1.47 (1.35, 1.55)
General characteristics					
Hospital beds, median [IQR]	217 (151, 380)	228 (166, 308)	182 (142, 325)	260 (200, 611)	220 (140, 583)
ICU beds, median [IQR]	14 (10, 20)	10 (8, 19)	16 (10, 20)	18 (10, 26)	14 (10, 20)
ICU/Hospital bed ratio (%), median [IQR]	5.9 (3.3, 12.1)	5.8 (3.5, 11.6)	9.1 (4.6, 13.1)	5.1 (2.5, 12.2)	6.9 (2.7, 12.0)
Medical-surgical ICU, N (%)	112 (88 %)	28 (88 %)	30 (94 %)	28 (88 %)	26 (81 %)
Admissions per bed, median [IQR]	136 (87, 168)	183 (142, 219)	148 (98, 170)	129 (79, 155)	80 (48, 125)
Average occupancy rate (%), median [IQR]	84 (77, 90)	82 (72, 88)	83 (76, 86)	86 (78, 91)	89 (83, 96)
Staffing characteristics					
Presence of training programin critical care, N (%)	27 (21 %)	3 (9.4 %)	2 (6.2 %)	10 (31 %)	12 (38 %)
Presence of a full-time board-certified intensivist, N (%)	40 (31 %)	11 (34 %)	7 (22 %)	12 (38 %)	10 (31 %)
Average Physician/bed ratio [c], median [IQR]	1.50 (1.33, 1.88)	1.61 (1.25, 1.88)	1.53 (1.40, 1.91)	1.50 (1.35, 1.94)	1.53 (1.36, 1.86)
Average nurse/bed ratio [c], median [IQR]	1.77 (1.43, 2.50)	2.18 (1.86, 2.59)	1.63 (1.37, 2.27)	1.63 (1.40, 2.04)	1.71 (1.36, 2.00)

ICU: Intensive care unit.
IQR: Interquartile range (1st quartile – 3rd quartile).
Median and IQR are calculated from the distribution of values at unit level.
[a]Average of SMR and SRU.
[b]Includes neurological, and oncological critical care units.
[c]Average number of physicians and nurses per 10 ICU beds in day and night shifts.

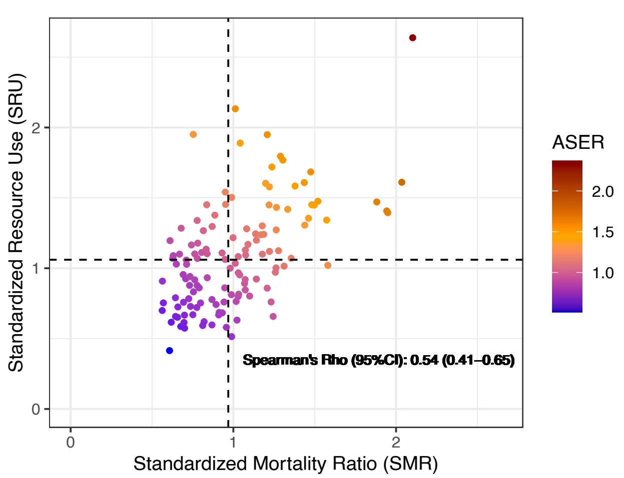

Fig. 1. Distribution of the standardised mortality ratio (SMR), standardized resource use (SRU), and average standardized efficiency ratio (ASER) values in the efficiency matrix in (A) Brazilian/Uruguayan ICUs. Standardised metrics were adjusted by risk using the SAPS-3 expected mortality. Reference lines are the median of SMR and SRU distributions at unit level.

structure. This analysis can assist healthcare managers in identifying subgroups that can be impacted by changes in structure or process and provide insights for new interventions to improve efficiency.

CRF's estimates presented a higher magnitude and slightly wider confidence intervals compared to the LRM with interactions. This may indicate that the interplay among confounders may have been captured in CRF and precision of intervals depend on sample size of trees. Although the CRF does not have a minimum sample size requirement, its estimation process is largely impacted by the number of available observations, as it splits the sample for inducing decision trees and estimating ATE ("honest splitting") [14,16,23]. On the other hand, including interactions in the regression model increases the number of parameters to estimate in a limited sample size, impacting CATE estimates' precision. In our sample, most organisational variables did not show a high univariable correlation nor an explicit nonlinear functional form between ASER and the organisational variable. CRF modelling does not assume linearity but can incorporate the functional form of the evaluated relationship. Our results showed that both models could be applied to the sample. However, although the CRF estimates are likely to be preferred, we cannot validate the results due to the absence of the

Table 3
Estimates of effect for each structural or organisational factor to the efficiency in the linear regression model with marginal effects estimation and the causal random forest.

	Main analysis (128 ICUs)		Sensitivity analysis (123 ICUs)	
	Linear Regression model	Causal Random Forest	Linear Regression model	Causal Random Forest
Exposure	**CATE [95 % CI]**	**CATE [95 % CI]**	**CATE [95 % CI]**	**CATE [95 % CI]**
Number of hospital beds (x100)	−0.01 [-0.04, 0.02]	0.03 [-0.02, 0.07]	0.01 [-0.02, 0.04]	0.04 [-0.01, 0.09]
ICU beds per hospital beds ratio	0.003 [-0.01, 0.01]	0.005 [-0.01, 0.02]	−0.002 [-0.01, 0.01]	0.006 [-0.008, 0.02]
Presence of training program in critical care	0.18 [-0.05, 0.42]	**0.19 [0.06, 0.33]**	**0.21 [0.02, 0.39]**	**0.21 [0.07, 0.36]**
Medical-surgical ICU	−0.05 [-0.22, 0.12]	−0.08 [-0.36, 0.20]	0.009 [-0.13, 0.15]	−0.03 [-0.31, 0.25]
Average physician/bed ratio [a]	0.09 [-0.02, 0.21]	0.06 [-0.15, 0.28]	0.08 [-0.01, 0.18]	0.06 [-0.17, 0.30]
Average nurse/bed ratio [a]	**−0.10 [-0.18, −0.02]**	**−0.13 [-0.24, −0.03]**	**−0.08 [-0.14, −0.01]**	**−0.13 [-0.25, −0.03]**
Presence of a full-time board-certified intensivist	0.03 [-0.08, 0.14]	−0.04 [-0.23, 0.14]	0.01 [-0.08, 0.10]	−0.02 [-0.21, 0.16]

CATE: Conditional average treatment effect.
CI: Confidence interval.
ICU: Intensive care unit.
[a] Average number of physicians and nurses per 10 ICU beds in day and night shifts.

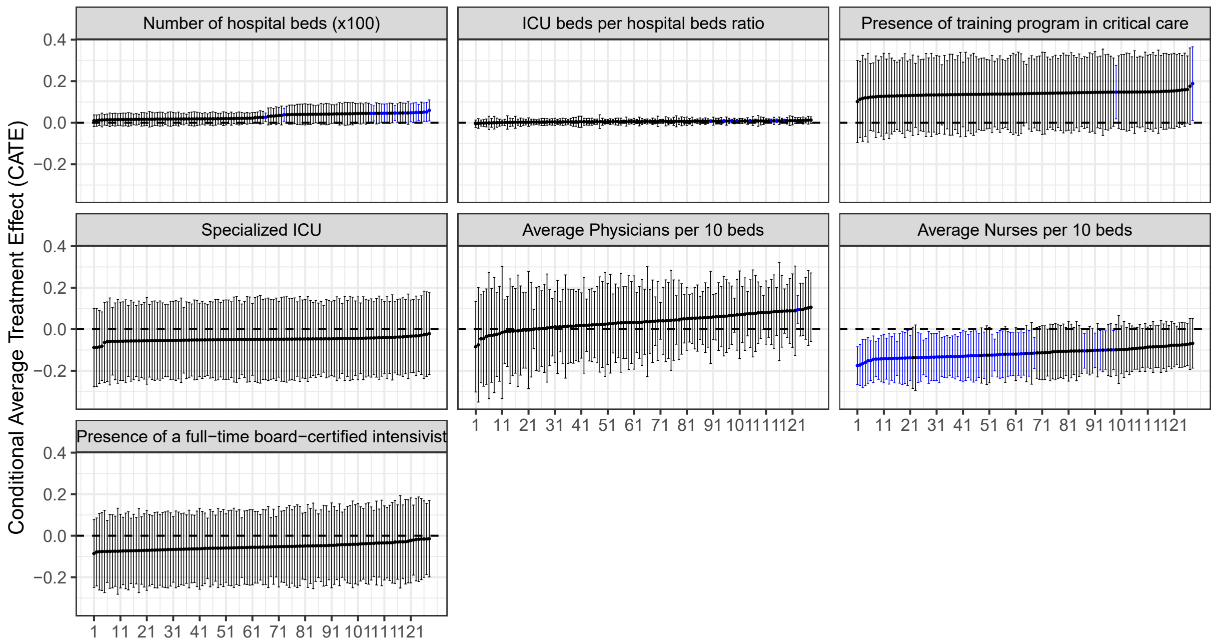

Fig. 2. Conditional Average Treatment Effect (CATE) for each ICU (Average Standardised Mortality Ration – ASER) estimated from Causal Random Forests stratified by exposure variable. Each point corresponds to an ICU, error bars correspond to the 95% confidence interval. Units were ordered from lowest to highest CATE in each exposure. Dashed line corresponds to the reference value of zero ("no effect"). Units with significant conditional treatment effect (confidence intervals not overlapping the reference line at zero) were highlighted in blue.

"ground truth" (the real treatment effect). Previous studies demonstrated that CRF is superior to classical regression modelling, especially due to its estimation of heterogeneous treatment effects, which does not require pre-specified interactions in the modelling stage, allowing for more flexibility in the data [13].

This is not the first study investigating the association between organisational factors and ICU efficiency in the Brazilian setting. At the patient level, Zampieri et al. identified that phenotypes of ICUs with high nursing autonomy, a full-time board-certified intensivist, and a dedicated pharmacist had better patient outcomes [12]. Wortel et al. found that an increased number of intensivists per ICU bed was associated with increased ICU efficiency in Dutch ICUs [10]. In Bastos et al. [9], information on the nurse/bed ratio was not available, and other staffing patterns were not significantly associated with efficiency in the final model. Our findings add to the evidence of these earlier studies that characteristics associated with ICU staffing patterns are associated with ICU efficiency. Also, ICUs from hospitals with a training program in critical care were associated with decreased performance. However, one should be cautious with the interpretation of this results due to potential unmeasured confounders related to management and reimbursement models of hospitals. For instance, in Brazil, training programs are more frequent in public and philanthropic than in private hospitals. In addition, hospitals with training programs admit in general more severe and complex patients, including variables which are not completely accounted for in risk-adjustments in case-mix using severity-of-illness scores such as the SAPS 3.

Our study has strengths and limitations. This study includes one of the largest samples of Brazilian ICUs with available information on organisational variables. We applied a novel machine learning model to estimate average treatment effects. A limitation of our study is that a limited number of organisational characteristics was considered in our analyses. These characteristics were mostly related to the ICU. Unfortunately, other ICU- or hospital-related factors were unavailable in our study data. Hence our analysis was mainly focused on staffing patterns in the ICU. The causal pathways for each variable can be improved by incorporating additional organisational information such as the implementation of protocols, quality-of-care assessment, and bundles of care [8,24]. Second, directly comparing and measuring the goodness-of-fit on linear modelling and causal random forest modelling performance was not possible because the "real" treatment effect is unknown. Therefore, the results of the models should not be interpreted as one approach is necessarily better than the other. However, CRF may be more likely to be closer to the truth since it does not assume a specific functional form between outcome, exposure and covariates in the model, and the model learns from the data. It also incorporates the interactions among variables to estimate heterogeneous treatment effects. Third, this dataset represents a portion of the Brazilian and Uruguayan ICU system, which may limit the generalization of our findings with respect to the effects of organizational factors. However, the ORCHESTRA dataset is the most representative and largest registry of ICU organizational characteristics in the country, composed of units from different regions, and administration/funding status and capacity.

From a methodological perspective, our results show that CATEs could be determined using CRF, especially when interactions (heterogeneous treatment effect) are considered. The clinical results of our study could be useful for manager-clinicians, policymakers and health management in low- and middle-income countries, as our results show that a higher number of nurses per ICU bed will pay off in terms of efficiency. Future studies may investigate potential ceiling effects to determine whether nurse and physician staffing patterns are efficient and whether results differ when benchmarking ICUs from countries with different economic or income status.

5. Conclusion

Using a large sample of Brazilian ICUs, the CRF and LRM with interaction terms identified comparable organisational factors with significant CATE on ICU efficiency. However, compared to the linear model, estimating heterogeneous treatment effects from CRFs can provide more precise insights from CATE without prespecifying interactions. Furthermore, with the CRF, effects on specific ICUs were identified even if the average effect itself was nonsignificant, which was not possible in the regression model. This analysis can assist healthcare managers in identifying specific units or subgroups that can benefit or not from new process interventions to improve efficiency.

Author contribution

Conceptualization and methodology: L. S. L. Bastos, S. A. Wortel, N. F. de Keizer, F. Bakhshi-Raiez, J.I. F. Salluh, D.A. Dongelmans, F. G. Zampieri, G. Burghi, A. Abu-Hanna, S. Hamacher, F. A. Bozza, M. Soares.

Data Curation: L. S. L. Bastos, F. G. Zampieri, G. Burghi, S. A. Wortel, F. Bakhshi-Raiez.

Writing – Original Draft: L. S. L. Bastos, S. A. Wortel.

Writing – Review & Editing: L. S. L. Bastos, S. A. Wortel, N.F. de Keizer, F. Bakhshi-Raiez, J.I. F. Salluh, D.A. Dongelmans, F. G. Zampieri, G. Burghi, A. Abu-Hanna, S. Hamacher, F. A. Bozza, M. Soares.

Visualization: L. S. L. Bastos.

Supervision: N.F. de Keizer, M. Soares.

Standardized Mortality Ratio (SMR).

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.

Declaration of interests

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 commercialises 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. NF 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. Wortel, 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.

Summary table (2–4 bullets each)

What was already known on the topic:

- Confounding adjustment is commonly performed with propensity score-based methods or by directly including variables in parametric regression models
- Novel methods based on machine learning tecniques have been proposed to improve confounder adjustment and provide estimates of interventions at subgroup levels (heterogeneous treatment effect)

What this study added to our knowledge:

- This is the first application of causal random forests to estimate effects of organisational factors in ICU efficiency
- In the analysed settings, both regression modelling with interactions and causal random forests showed comparable average treatment effecets
- With causal random forests, we could predict significant effects of exposures in certain ICUs, even when the average effect was nonsignificant
- The use of machine learning methods such as causal random forests could assist healthcare managers in further in-dept evaluation of process interventions to improve ICU efficiency.

CRediT authorship contribution statement

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

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2024.105568.

References

[1] D.W. De Lange, D.A. Dongelmans, N.F. De Keizer, Small steps beyond benchmarking, Revista Brasileira De Terapia Intensiva 29 (2017) 128–130, https://doi.org/10.5935/0103-507X.20170022.

[2] A. Garland, Improving the ICU: Part 2, Chest 127 (2005) 2165–2179, https://doi.org/10.1378/chest.127.6.2165.

[3] J.I.F. Salluh, M. Soares, M.T. Keegan, Understanding intensive care unit benchmarking, Intensive Care Medicine 43 (2017) 1703–1707, https://doi.org/10.1007/s00134-017-4760-x.

[4] D. Woodhouse, M. Berg, J. Van Der Putten, J. Houtepen, Will benchmarking ICUs improve outcome? Current Opinion in Critical Care 15 (2009) 450–455, https://doi.org/10.1097/MCC.0b013e32833079fb.

[5] F.A. Bozza, J.I.F. Salluh, An urban perspective on sepsis in developing countries, The Lancet Infectious Diseases 10 (2010) 290–291, https://doi.org/10.1016/S1473-3099(10)70074-8.

[6] J.I.F. Salluh, M. Soares, M. Singer, Spreading the knowledge on the epidemiology of sepsis, The Lancet Infectious Diseases 17 (2017) 1104–1106, https://doi.org/10.1016/S1473-3099(17)30480-2.

[7] H.U. Rothen, K. Stricker, J. Einfalt, P. Bauer, P.G.H. Metnitz, R.P. Moreno, J. Takala, Variability in outcome and resource use in intensive care units, Intensive Care Medicine 33 (2007) 1329–1336, https://doi.org/10.1007/s00134-007-0690-3.

[8] M. Soares, F.A. Bozza, D.C. Angus, A.M. Japiassú, W.N. Viana, R. Costa, L. Brauer, B.F. Mazza, T.D. Corrêa, A.L.B. Nunes, T. Lisboa, F. Colombari, A.T. Maciel, L.C. P. Azevedo, M. Damasceno, H.S. Fernandes, A.B. Cavalcanti, P.E.A.A. do Brasil, J. M. Kahn, J.I.F. Salluh, Organizational characteristics, outcomes, and resource use in 78 Brazilian intensive care units: the ORCHESTRA study, Intensive Care Medicine 41 (2015) 2149–2160, https://doi.org/10.1007/s00134-015-4076-7.

[9] L.S.L. Bastos, S. Hamacher, F.G. Zampieri, A.B. Cavalcanti, J.I.F. Salluh, F.A. Bozza, Structure and process associated with the efficiency of intensive care units in low-resource settings: An analysis of the CHECKLIST-ICU trial database, Journal of Critical Care 59 (2020) 118–123, https://doi.org/10.1016/j.jcrc.2020.06.008.

[10] S.A. Wortel, N.F. de Keizer, A. Abu-Hanna, D.A. Dongelmans, F. Bakhshi-Raiez, Number of intensivists per bed is associated with efficiency of Dutch Intensive Care Units, (2020).

[11] W. Checkley, G.S. Martin, S.M. Brown, S.Y. Chang, O. Dabbagh, R.D. Fremont, T. D. Girard, T.W. Rice, M.D. Howell, S.B. Johnson, J. O'Brien, P.K. Park, S. M. Pastores, N.T. Patil, A.P. Pietropaoli, M. Putman, L. Rotello, J. Siner, S. Sajid, D. J. Murphy, J.E. Sevransky, Structure, Process, and Annual ICU Mortality Across 69 Centers: United States Critical Illness and Injury Trials Group Critical Illness Outcomes Study*, Critical Care Medicine 42 (2014) 344–356, https://doi.org/10.1097/CCM.0b013e3182a275d7.

[12] F.G. Zampieri, J.I.F. Salluh, L.C.P. Azevedo, J.M. Kahn, L.P. Damiani, L.P. Borges, W.N. Viana, R. Costa, T.D. Corrêa, D.E.S. Araya, M.O. Maia, M.A. Ferez, A.G. R. Carvalho, M.F. Knibel, U.O. Melo, M.S. Santino, T. Lisboa, E.B. Caser, B.A.M. P. Besen, F.A. Bozza, D.C. Angus, M. Soares, ICU staffing feature phenotypes and their relationship with patients' outcomes: an unsupervised machine learning analysis, Intensive Care Medicine 45 (2019) 1599–1607, https://doi.org/10.1007/s00134-019-05790-z.

[13] S. Wager, S. Athey, Estimation and Inference of Heterogeneous Treatment Effects using Random Forests, Journal of the American Statistical Association 113 (2018) 1228–1242, https://doi.org/10.1080/01621459.2017.1319839.
[14] H.A. Chipman, E.I. George, R.E. McCulloch, BART: Bayesian additive regression trees, Annals of Applied Statistics (2012), https://doi.org/10.1214/09-AOAS285.
[15] S. Athey, J. Tibshirani, S. Wager, Generalized random forests, Annals of Statistics (2019), https://doi.org/10.1214/18-AOS1709.
[16] S. Linde, H. Shimao, An observational study of health care provider collaboration networks and heterogenous hospital cost efficiency and quality outcomes, Medicine 101 (2022) e30662.
[17] L. Li, R.A. Levine, J. Fan, Causal effect random forest of interaction trees for learning individualized treatment regimes with multiple treatments in observational studies, Stat 11 (2022), https://doi.org/10.1002/sta4.457.
[18] S. Athey, G. Imbens, Recursive partitioning for heterogeneous causal effects, Proceedings of the National Academy of Sciences of the United States of America 113 (2016) 7353–7360. DOI: 10.1073/pnas.1510489113.
[19] F.G. Zampieri, M. Soares, L.P. Borges, J.I. Figueira Salluh, O.T. Ranzani, The Epimed Monitor ICU Database®: A cloud-based national registry for adult intensive care unit patients in Brazil, Revista Brasileira De Terapia Intensiva 29 (2017) 418–426, https://doi.org/10.5935/0103-507X.20170062.
[20] J.I.F. Salluh, T. Lisboa, Critical Care in Brazil, 16 (2016). https://healthmanagement.org/c/icu/issuearticle/critical-care-in-brazil-1.
[21] P.G.H. Metnitz, R.P. Moreno, E. Almeida, B. Jordan, P. Bauer, R.A. Campos, G. Iapichino, D. Edbrooke, M. Capuzzo, J.-r., Le Gall, on behalf of the SAPS 3 Investigators, SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description, Intensive Care Med 31 (2005) 1336–1344, https://doi.org/10.1007/s00134-005-2762-6.
[22] L.S.L. Bastos, S.A. Wortel, N.F. de Keizer, F. Bakhshi-Raiez, J.I.F. Salluh, D. A. Dongelmans, F.G. Zampieri, G. Burghi, A. Abu-Hanna, S. Hamacher, F.A. Bozza, M. Soares, Comparing continuous versus categorical measures to assess and benchmark intensive care unit performance, Journal of Critical Care 70 (2022) 154063, https://doi.org/10.1016/j.jcrc.2022.154063.
[23] R. Moreno, G. Apolone, Impact of different customization strategies in the performance of a general severity score, Critical Care Medicine 25 (1997) 2001–2008, https://doi.org/10.1097/00003246-199712000-00017.
[24] D.B. Rubin, Estimating causal effects of treatments in randomized and nonrandomized studies, Journal of Educational Psychology 66 (1974) 688–701, https://doi.org/10.1037/h0037350.
[25] G. Maldonado, S. Greenland, Simulation study of confounder-selection strategies, American Journal of Epidemiology (1993), https://doi.org/10.1093/oxfordjournals.aje.a116813.
[26] R.M. Mickey, S. Greenland, The impact of confounder selection criteria on effect estimation, American Journal of Epidemiology (1989), https://doi.org/10.1093/oxfordjournals.aje.a115101.
[27] E.C. Norton, B.E. Dowd, M.L. Maciejewski, Marginal Effects—Quantifying the Effect of Changes in Risk Factors in Logistic Regression Models, JAMA 321 (2019) 1304, https://doi.org/10.1001/jama.2019.1954.
[28] J. Tibshirani, S. Athey, E. Sverdrup, S. Wager, grf: Generalized Random Forests, 2022. https://CRAN.R-project.org/package=grf.
[29] J. Tibshirani, S. Athey, S. Wager, generalized random forests: package reference, (n.d.). https://grf-labs.github.io/grf/ (accessed December 8, 2020).