






Missing data methods for intensive care unit SOFA scores in electronic health records studies: results from a Monte Carlo simulation

Journal of **Comparative Effectiveness Research**

Daniel L Brinton^{*1} , Dee W Ford², Renee H Martin², Kit N Simpson¹ , Andrew J Goodwin² & Annie N Simpson¹ 

¹ College of Health Professions, Medical University of South Carolina, SC 29425, USA

² College of Medicine, Medical University of South Carolina, SC 29425, USA

*Author for correspondence: Tel.: +1 843 792 8383; brintond@musc.edu

Aim: Missing data cause problems through decreasing sample size and the potential for introducing bias. We tested four missing data methods on the Sequential Organ Failure Assessment (SOFA) score, an intensive care research severity adjuster. **Methods:** Simulation study using 2015–2017 electronic health record data, where the complete dataset was sampled, missing SOFA score elements imposed and performance examined of four missing data methods – complete case analysis, median imputation, zero imputation (recommended by SOFA score creators) and multiple imputation (MI) – on the outcome of in-hospital mortality. **Results:** MI performed well, whereas other methods introduced varying amounts of bias or decreased sample size. **Conclusion:** We recommend using MI in analyses where SOFA score component values are missing in administrative data research.

First draft submitted: 22 March 2021; Accepted for publication: 14 October 2021; Published online: 2 November 2021

Keywords: administrative data • critical care • health services research • Monte Carlo method • risk adjustment • severity of illness

In both prospective and retrospective outcomes studies, use of a patient severity score – such as the Sequential Organ Failure Assessment (SOFA) score [1] – is vital for risk adjustment in multivariable models. While caution has been advised for using these patient severity scores on the individual level for prognosis, they are effective for severity and case-mix adjustments across larger populations [2]. Therefore, the use of severity score systems is common in critical care research, but not without imperfections in execution. Severity score systems are typically composites of multiple component measures; in the case of the SOFA, there are six items that are physiological indicators of organ failure. Thus, when SOFA scores are derived from electronic health record (EHR) data, it is common that at least one component measure is missing preventing the accurate calculation of the score [1,3–5].

As part of our research program in critical illness, institutional EHR data were used to evaluate process and outcome measures among patients treated for ventilator dependent respiratory failure (VDRF) in intensive care units (ICUs) at a large academic medical center in southeastern USA. SOFA scores were used as a severity adjustment tool, however significant issues with missing EHR data were identified. The Sepsis-3 consensus paper [6] recommends assuming no organ derangement, “unless the patient is known to have preexisting (acute or chronic) organ dysfunction” (p. 805). While the amount of derangement for other organ systems may be known, when one (or more) SOFA component data points are missing, the Sepsis-3 consensus paper is directing one to impute a SOFA component score of zero. Although this approach would allow SOFA scores to be calculated for all patients – increasing statistical power when using SOFA as a severity adjuster – it was hypothesized that this approach might introduce bias. Therefore, an evaluation of four common statistical methods for addressing missing data (namely, complete case, median imputation, zero imputation and multiple imputation [MI]) was conducted in order to inform recommendations for those who may confront similar issues.

Through application via a simulation study approach, the effects of missing SOFA score components and the degree to which results were biased was evaluated. The simulation considered two types of missing data mechanisms, missing at random (MAR) and missing not at random (MNAR), as well as, various percentages of missing data (10, 20, 30 and 40%). Finally, each simulated dataset was compared with the complete dataset with no missing data, to identify which method best minimizes bias.

Materials & methods

Study design

This is a retrospective, observational cohort study utilizing de-identified patient data extracted from the clinical data warehouse of an academic medical center located in the southeastern USA. This study was reviewed by the Institutional Review Board and designated as non-human subjects research.

VDRF cohort development

Data extracted for this study included all adult patients aged at least 18 years who were admitted to one of five adult ICUs at an academic medical center in the southeastern USA between 1 January 2015 through 31 October 2017 and were mechanically ventilated for at least 96 h – as indicated by an ICD-9 procedure code of 96.72 or an ICD-10 procedure code of 5A1955Z. For patients with multiple ICU admissions, only the first admission was included to create a cohort of unique patients. Patients with missing start or end ventilation dates/times were excluded.

Measures extracted for VDRF cohort

Data for this study included: demographic measures (age, sex, race and primary payer); components of SOFA score within 24 h of admission to the ICU (PaO₂, FiO₂, platelets, total bilirubin, mean arterial pressure [MAP], vasopressors, Glasgow Coma Score [GCS] [7], creatinine and urine output); measures relevant to mechanical ventilation (height, ventilator settings, performance of spontaneous breathing trials); and other clinical measures of interest (Richmond Agitation-Sedation Scale (RASS) scores [8,9], Confusion Assessment Method for the ICU (CAM-ICU) scores [10]). In addition, all diagnosis and procedure codes, discharge disposition, hospital length of stay and total charges were extracted. The Charlson comorbidity score [11,12] was calculated for each patient using diagnosis codes. In this study results and findings related to SOFA score calculations and patient in-hospital mortality are reported.

Outcome measure

The outcome measure utilized in this study is in-hospital mortality which is defined by a discharge disposition of death.

Simulation method overview

An algorithm depicting the simulation process used in this study is represented in [Figure 2](#) and described herein. The published reference on conducting simulation studies in medical statistics by Burton *et al.* served as a guide for the simulation methods [13]. At the start of the simulation, a dataset with no missing measures was provided that included all component values of the SOFA score, composite SOFA score, potential covariates and the outcome of in-hospital mortality. The simulation parameter γ which denotes the percentage of observations within the dataset that will contain one or more missing SOFA component values varied across four levels, $\gamma = 10, 20, 30$ and 40% , as these represent moderate amounts of missing data. From the dataset without missing measures, a simple random sample of 1000 observations was chosen. Missing data were then generated under the MAR and MNAR mechanisms by choosing γ percent of the observations to have a missing SOFA score value resulting in four new datasets with variable proportions of missing data. For the process of selecting which records would be chosen to have elements of SOFA score deleted, under MAR those who died were more likely to be selected; under MNAR, records with higher SOFA scores were more likely to be selected. Once records were selected to have SOFA score elements deleted, missing data patterns were assigned to records based on their observed frequency in the original dataset, and elements deleted accordingly.

Next, each of the four different statistical methods for addressing missing data were applied to the four derivative datasets with missing data. Finally, a logistic regression model was developed to predict the outcome of in-hospital mortality for each of the simulated datasets. A total of 1000 simulation loops were ran.

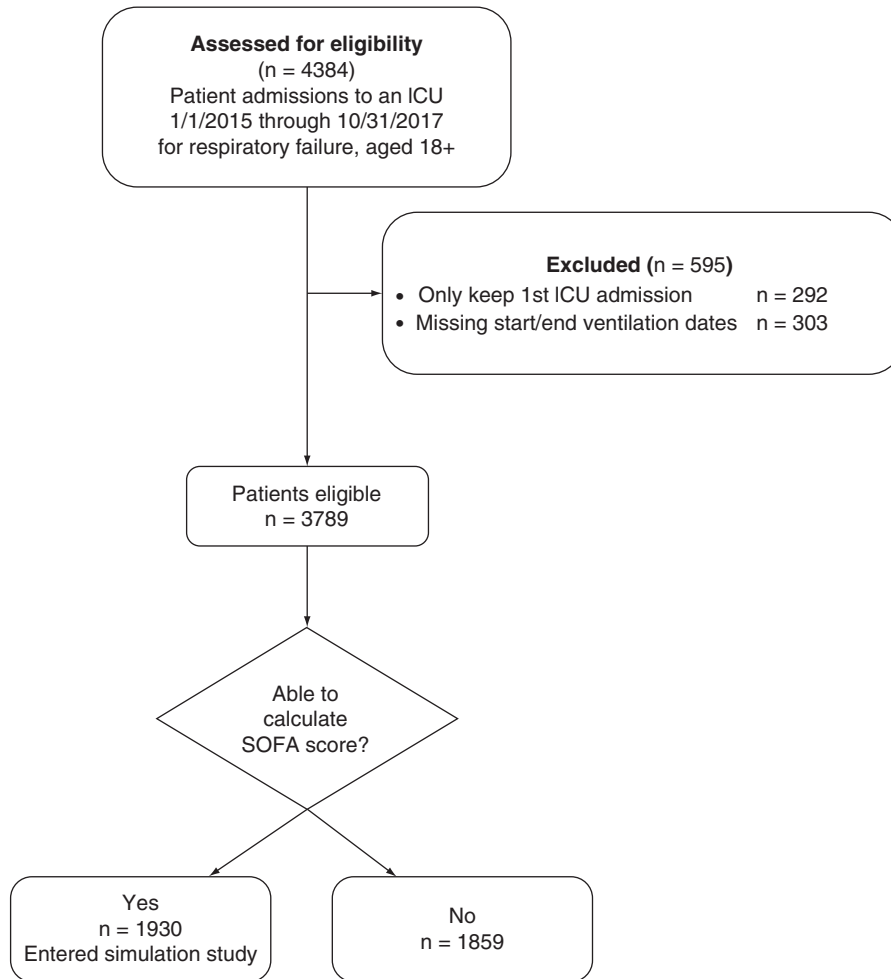


Figure 1. CONSORT flow diagram illustrating the development of the ventilator-dependent respiratory failure cohort for this study.

ICU: Intensive care unit; SOFA: Sequential Organ Failure Assessment.

Missing data methods

Four methods for handling missing data were implemented in this simulation study. The first method, complete case analysis, is the default method of all commercially available statistical software. Using complete case analysis cases are removed from the regression modeling analysis if one or more SOFA score variables are missing. The second missing data method, median imputation, imputes the median component value for the SOFA score across the entire dataset for missing values. The third method, zero imputation, imputes a 0 for missing SOFA score variables because they are assumed to have no derangement, consistent with the approach recommended in the Sepsis 3 guidelines [6]. The fourth, missing data method, MI, uses multiple imputation by chained equations (MICE), also known as fully conditional specification using a regression-based approach. An MI approach is available in all major statistical software packages (i.e., SAS, SPSS, Stata and R). The goal of MI is not to assume data, but rather to allow all the data that are present to be used in analyses to achieve valid statistic inference, not perfect point prediction [14]. Essentially, MI assigns plausible values to missing data points based on the data that are available. Theoretically, MI is superior to median and zero imputation as it improves the accuracy of imputed data and is superior to complete case analysis since it allows all existing data to be used.

A total of 25 imputations were performed per model, with the point estimates and standard errors combined using Rubin's rules [15]. The variables used for MI include age, race, sex, payor group, the six component SOFA elements, log-transformed ICU length of stay, log-transformed total charges, as well as the outcome of in-hospital mortality. ICU length of stay and total charges were log transformed to approximately conform to normality.

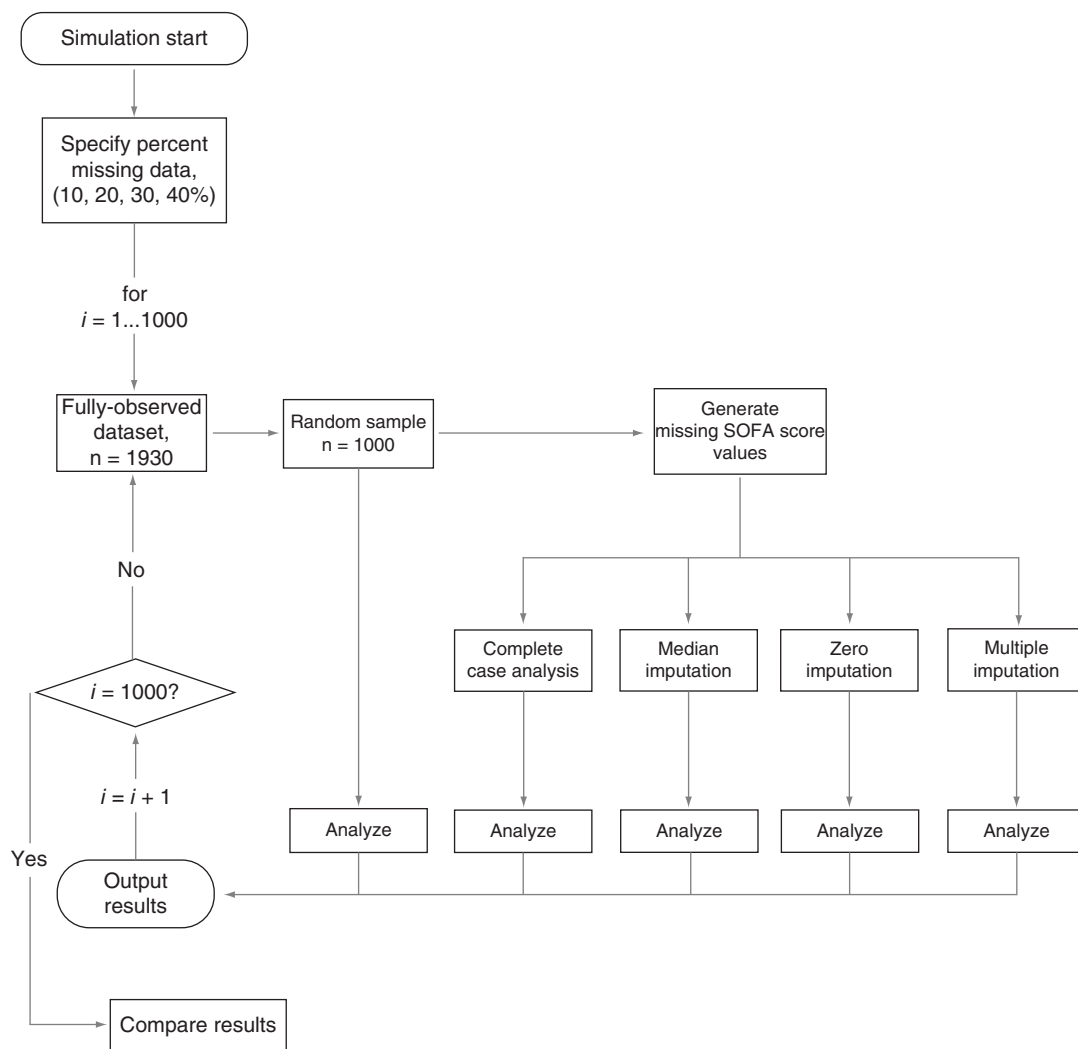


Figure 2. Algorithm showing the simulation process used in this study. From the dataset without missing measures, a simple random sample of 1000 observations were chosen. The simulation parameter, γ (set to 10, 20, 30 and 40%), denotes the percentage of observations within the dataset that will contain one or more missing Sequential Organ Failure Assessment component values. Missing data were then generated under the missing at random and missing not at random mechanisms by choosing γ percent of the observations to have a missing Sequential Organ Failure Assessment score value resulting in four new datasets with variable proportions of missing data. Next, four statistical methods for addressing missing data were applied to the four derivative datasets with missing data and a logistic regression model was developed to predict the outcome of in-hospital mortality. The simulation loops through 1000 times for each value of γ and each missing data mechanism (missing at random and missing not at random).

Statistical analysis

Each of the missing data methods were examined using a consistent multiple logistic regression model which included the same: covariates: age (in decades), sex, race (Black, White, Other/unknown); exposure variable: composite SOFA score; and outcome variable: in-hospital mortality. The first model was fit on the fully observed dataset (i.e., no missing SOFA score variables), then the same model was used in all of the simulations without subsequent model fitting (no covariates were removed). Statistical significance was determined *a priori* to be at the $\alpha = 0.05$ level.

Two summary statistics were used to assess the impact of each of the missing data approaches on bias. The first summary statistic is the pooled odds ratios for in-hospital mortality and 95% CIs across the 1000 simulations, comparing these to the population parameter of the odds ratio (from the complete dataset) for the effect of SOFA score on in-hospital mortality. These odds ratios were pooled using Rubin's rules [15], and are presented graphically

as forest plots. When a missing data method's 95% confidence interval does not contain the true odds ratio, this is indicative of a biased estimate of SOFA score's effect on in-hospital mortality.

The second summary statistic is relative bias. Relative bias is calculated as $\frac{\hat{\beta}_i - \beta}{\beta}$, where $\hat{\beta}_i$ represents the parameter estimate for the SOFA score for each of the ($i = 1, \dots, 1000$) simulation runs in the logistic regression model, and β represents the population parameter of the SOFA score from the fully observed dataset. Ideally, this number will be 0% – meaning there is no difference in parameter estimates between the simulated regressions and the regression performed on the sample that has no missing data. Means and 95% CIs of the relative bias for each missing data technique at each percentage of missingness were calculated, allowing the observation of the magnitude and direction of bias that each missing data technique introduces. Relative bias was calculated for each simulation run and missing data method.

Results

A total of 4,384 adult admissions to an ICU who had VDRF and were intubated for at least 96 h (see [Figure 1](#)) were assessed for eligibility. As mentioned previously, only the first admission to an ICU for each patient (removing 292 admissions) were retained, and those who had missing start/end ventilation times ($n = 303$) were excluded. In total, 3789 patients were eligible for inclusion into this study. Next, the ability to calculate a complete SOFA score using the EHR data was examined. Only 1930 of these patients (50.9%) had all variables present within the EHR to calculate a complete SOFA score. Among those for whom a SOFA score could not be calculated because of missing items, the most commonly missing variables were bilirubin (27.5% of records), Glasgow Coma Scale score (19.9%) and PaO₂/FiO₂ (5.9%); the cardiovascular SOFA component (mean arterial pressure and vasopressor usage) was the most infrequently missing (3.0%).

The demographics and characteristics of the VDRF patient cohort used for this simulation study, grouped on whether or not all the variables were present to calculate a complete SOFA score is presented in [Table 1](#). These two groups did not differ significantly with respect to age, sex, race, insurance status, Charlson score, length of stay, nor total charges. However, the in-hospital mortality rate was higher among patients without missing data compared with those who were missing SOFA component data (36.5 vs 27.3%, $p < 0.0001$) suggesting the possibility of an MNAR mechanisms. Groups did differ by in-hospital mortality rate.

Given that a higher SOFA score positively correlates with greater patient acuity within the intensive care setting, as well as, poor outcomes following the ICU stay, it is probable that the SOFA scores for those patients whose SOFA scores could not be calculated as a result of one or more missing components would have, on average, lower SOFA scores than those with complete data. However, this finding is given with caution as other studies have shown a single mechanism is unlikely to be the sole cause of missingness, and empirical distinction between the two missing data mechanisms is impossible [16,17].

The pooled odds ratios for in-hospital mortality for a dataset with 40% data MAR is shown in [Figure 3](#). The true odds ratio (i.e., from the complete data set) for the association of SOFA score and in-hospital mortality was 1.21. MICE provided a completely unbiased estimate while zero imputation resulted in an odds ratio that was biased against an association between SOFA score and in-hospital mortality (95% confidence interval did not include true estimate). Complete case analysis yielded less statistical power as evidenced by widened confidence intervals but did not introduce significant bias to the association. Median imputation resulted in a trend toward a reduced odds ratio, however, the 95% confidence interval still contained the true odds ratio. Similar patterns were observed when the simulations were run on data sets missing 30, 20 and 10% ([Supplementary Figure 1A](#)) although, as expected, bias introduced by zero imputation decreased as the amount of missing data decreased.

[Figure 4](#) demonstrates the pooled odds ratios for in-hospital mortality in datasets where 40% of the data is MNAR. Once again, MICE provided an unbiased estimate of the true association between SOFA score and in-hospital mortality. Zero imputation again underestimated the odds ratio; however, in this non-random dataset the 95% confidence interval did contain the true odds ratio. Complete case analysis and median imputation performed similarly between the MAR and MNAR datasets. Each imputation method performed similarly across datasets with varying levels of missing data ([Supplemental Figure A1](#)).

Discussion

This study sought to examine the effect of different approaches to handling missing SOFA score variables. The SOFA score is a widely used and important risk adjustment tool in critical care research, and missing data frequently

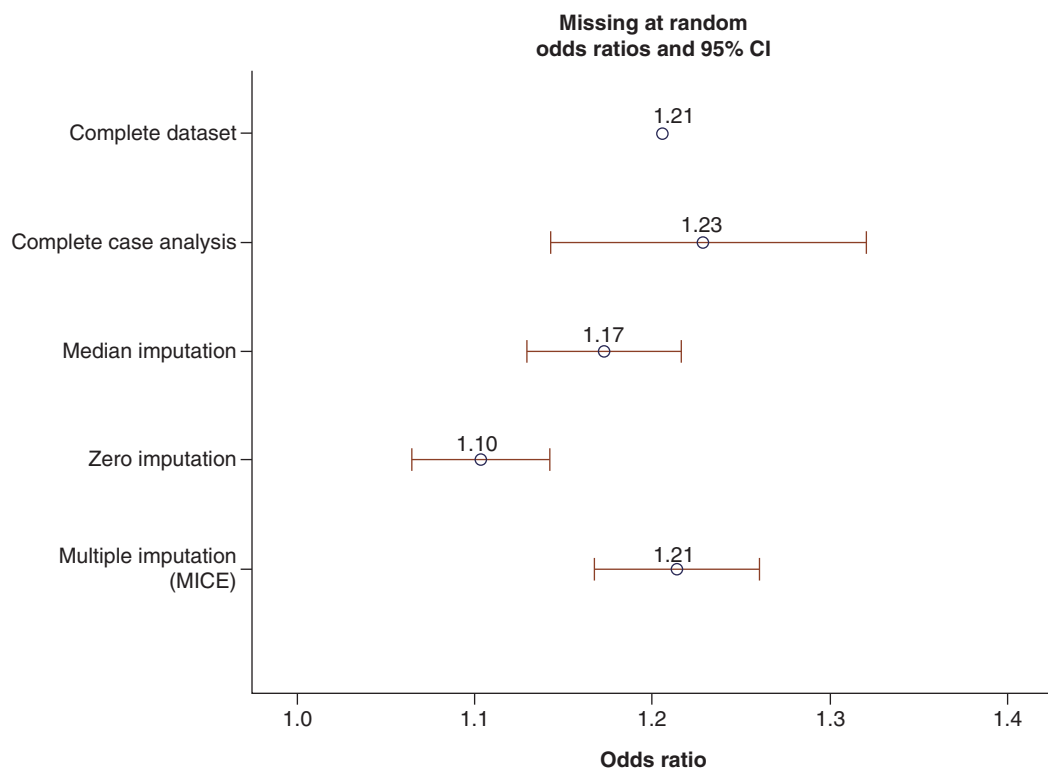


Figure 3. Pooled odds ratios for in-hospital mortality and 95% CIs across 1000 simulations for the missing at random missing data mechanism at 40% missing Sequential Organ Failure Assessment score components.

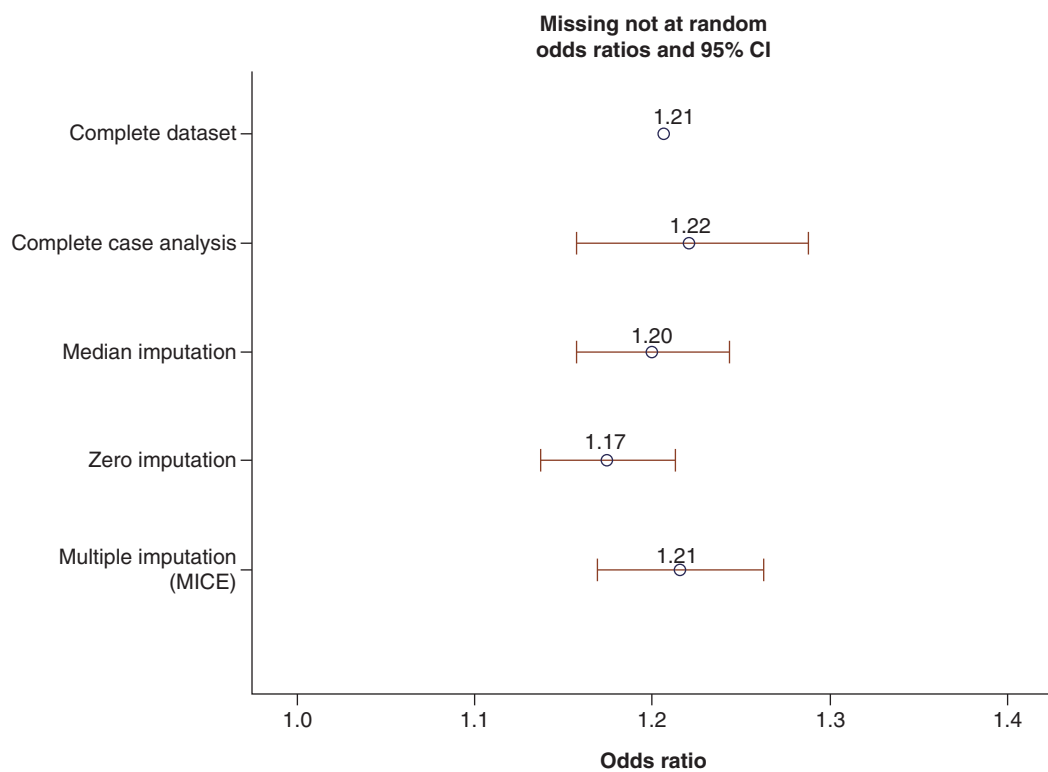


Figure 4. Pooled odds ratios and 95% CIs for in-hospital mortality across 1000 simulations for the missing not at random missing data mechanism at 40% missing Sequential Organ Failure Assessment score variables.

Table 1. Characteristics of ventilator dependent respiratory failure patient cohort.

	SOFA score present?		p-value [†]
	Yes (n = 1930)	No (n = 1859)	
Age (years)	56.6 ± 17.1	55.6 ± 17.7	0.0844
Male	1126 (58.3)	1099 (59.1)	0.6277
Race			0.1631
Black	794 (41.1)	735 (39.5)	
White	1042 (54.0)	1051 (56.5)	
Other/unknown	94 (4.9)	73 (3.9)	
Insurance			0.9739
Commercial	598 (31.0)	577 (31.0)	
Medicare/Medicaid	1078 (55.9)	1042 (56.1)	
Other/unknown	254 (13.1)	240 (12.9)	
Charlson score	3.0 ± 3.0	3.2 ± 2.8	0.0919
Length of stay [‡]			
ICU	10.0 ± 10.7	9.9 ± 11.3	0.8137
Overall	17.1 ± 21.6	17.3 ± 19.4	0.7547
Total charges	\$198,539 ± 219,757	\$200,896 ± 225,994	0.7449
Died	704 (36.5)	507 (27.3)	<0.0001
SOFA score	8.8 ± 4.1 8 [6] [§]		
SOFA components [¶]			
CNS	2 [0–4]		
Cardiovascular	1 [0–4]		
Coagulation	0 [0–4]		
Hepatic	0 [0–4]		
Renal	1 [0–4]		
Respiratory	3 [0–4]		

All values are expressed as mean ± S.D., n (%), or as otherwise indicated.

[†]p-Values were calculated using the Wilcoxon Mann-Whitney U test for continuous measures, and the χ^2 or Fisher's Exact tests for categorical measures (as appropriate). Statistically significant comparisons at the $\alpha = 0.05$ level are given in bold.

[‡]Expressed in days.

[§]Median (interquartile range).

[¶]Median (range).

Other/unknown race is comprised of Asian, Hawaiian, Indian/Alaskan and where this value was missing from the original dataset.

leads to incomplete scores in large datasets. Currently, there is no standard methodology for addressing these missing data, although the recently published Sepsis-3 guidelines have advocated for the use of zero imputation (cite). Our results suggest that MI by chained equations (MICE) provides superior statistical adjustment for missing SOFA score data and should be the recommended technique for addressing this critical problem. By comparison, alternate techniques either introduced bias (zero imputation) or reduced statistical power (complete case analysis) when calculating the association between SOFA scores and mortality. These errors were less exaggerated with median imputation. In comparison, two of the alternatives tested herein – median and zero imputation, both deterministic imputation techniques – yielded tighter confidence intervals than they should due to lack of accounting for the missing data by invoking a smaller variance. Conversely, MI techniques create multiple datasets which Rubin (1987, p. 2) described as “representing a distribution of possibilities” [15]. As previously stated, the goal of MI is not to make up data, but rather to allow all the data that are present to be used in analyses to achieve valid statistical inference, not perfect point prediction [14].

With the advent and widespread adoption of EHR systems, large, clinical databases amenable to pragmatic outcomes research are now ubiquitous. However, missing data is an inherent limitation to these “real world” databases requiring statistical approaches in order to maximize database utility. The goal of MI is not to arbitrarily assign value to a data point, but rather to allow all the available data to be used for valid inferences of its value. By comparison, median and zero imputation, both deterministic imputation techniques, yield estimates that are not always associated with existing data elements and, therefore, may introduce bias. In our analysis, this was particularly

evident with zero imputation, which resulted in negative bias in nearly all examined scenarios. Likewise, complete case analysis excludes any subject with a missing data element resulting in unnecessary loss of statistical power. While these three alternative approaches are simpler and easier to implement in a clinical setting, our data suggest that MICE should be used to address missing SOFA data in research whenever possible.

Based on our findings and review of the existing literature, we offer three recommendations for performing MI on missing SOFA scores. First, use as many variables as possible in the MI process, including at least all the variables you might use in your later analysis [18]. Second, ensure you include all outcomes of analysis – in-hospital mortality in this study – in the MI model [19,20]. Third, use at least the default number of imputations in your statistical package – which in the case of SAS is 25, or one imputation for every 1% of missing data [21]. Finally, for additional best-practice recommendations, we highlight additional helpful references [14,21,22].

Limitations of this study include only using data from one academic medical center in the southeastern United States, which possibly limits generalizability due to local practice patterns. Similarly, this study included only a specific clinical cohort – those with VDRE. We had an initial high rate of missingness in our dataset (49.1% of SOFA scores included one or more missing component values) and it is possible that other studies using EHR data may encounter less missing data. Finally, our dataset contained a limited SOFA score range (0-22; max 24).

Future research should externally validate the findings of this study using a larger dataset, with ideally a smaller percent of missing data from which to sample for the Monte Carlo simulations. Additionally, it is important to test the generalizability of this study's findings, specifically their geographic and historical transportability [23], to determine if these findings remain consistent in different populations. Consideration should also be given to applying this evaluation framework to other ICU severity adjustment tools.

Conclusion

A simulation study was conducted to test several missing data methods at varying amount of missingness, using two of the most likely missing data mechanisms. The result of this study showed that the Sepsis-3 consensus paper's method for handling missing component SOFA score values (imputing a zero; assuming no organ derangement) yields biased estimates, and is, therefore, not recommended. Conversely, MI worked well – yielding unbiased estimates of the association between SOFA score and in-hospital mortality even when data were MNAR. This study suggests that researchers use MI when component SOFA score items are missing.

Summary points

- The Sequential Organ Failure Assessment (SOFA) score is an essential risk adjuster in critical care research.
 - SOFA score ranges from 0 to 24, and is composed of 6 subscores.
 - The subscores have a range of 0–4 points being assigned to each of six organ systems: respiratory, hematologic, hepatic, cardiac, neurologic and renal.
 - A higher score represents a higher level of dysfunction, and thus greater severity.
- Although the elements which comprise the SOFA score are routinely collected in the intensive care unit, they often are missing – preventing accurate calculation of the score.
- When data are missing, the Sepsis-3 guidelines suggest assuming no organ derangement – which is to impute a zero.
- This study tested four common missing data methods for the SOFA score on the outcome of death, finding multiple imputation to perform quite well in amount of up to 40% missingness.
- The other methods tested introduced varying amounts of bias or decreased the sample size.
- We recommend multiple imputation be part of the *a priori*-defined statistical analysis plan to handle missing SOFA score component data.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/cer-2021-0079

Author contributions

All authors were responsible for the design of the study, interpretation of the findings, and critical revisions of the written findings. DL Brinton is responsible for the analysis and drafting of the written findings. All authors meet the ICMJE's 4 criteria for authorship.

Financial & competing interests disclosure

This publication was supported by the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) as part of the National Telehealth Center of Excellence Award (U66 RH31458-01-00). The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement, by HRSA, HHS, or the U.S. Government. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors affirm they have obtained appropriate institutional review board approval by the Medical University of South Carolina for this research, and this research has been designated as non-human subjects research.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

Papers of special note have been highlighted as: ● of interest; ●● of considerable interest

- Vincent JL, De Mendonca A, Cantraine F *et al.* Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. *Crit. Care Med.* 26(11), 1793–1800 (1998).
- **Validation study of the SOFA score**
- Strand K, Flaatten H. Severity scoring in the ICU: a review. *Acta Anaesthesiol. Scand.* 52(4), 467–478 (2008).
- Buyse S, Teixeira L, Galicier L *et al.* Critical care management of patients with hemophagocytic lymphohistiocytosis. *Intensive Care Med.* 36(10), 1695–1702 (2010).
- Neto AS, Barbas CSV, Simonis FD *et al.* Epidemiological characteristics, practice of ventilation, and clinical outcome in patients at risk of acute respiratory distress syndrome in intensive care units from 16 countries (PROVENT): an international, multicentre, prospective study. *Lancet Resp. Med.* 4(11), 882–893 (2016).
- Ferreira FL, Bota DP, Bross A, Melot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA* 286(14), 1754–1758 (2001).
- Singer M, Deutschman CS, Seymour CW *et al.* The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 315(8), 801–810 (2016).
- **Sepsis-3 guidelines, which state one is to assume no organ derangement unless the patient has a known organ dysfunction**
- Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* 2(7872), 81–84 (1974).
- Sessler CN, Gosnell MS, Grap MJ *et al.* The Richmond Agitation–Sedation Scale: validity and reliability in adult intensive care unit patients. *Am. J. Respir. Crit. Care Med.* 166(10), 1338–1344 (2002).
- Sessler CN, Grap MJ, Brophy GM. Multidisciplinary management of sedation and analgesia in critical care. Presented at: *Semin. Respir. Crit. Care Med.* (2001).
- Ely EW, Inouye SK, Bernard GR *et al.* Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). *JAMA* 286(21), 2703–2710 (2001).
- Charlson ME, Pompei P, Ales KL, Mackenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J. Chronic Dis.* 40(5), 373–383 (1987).
- Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J. Clin. Epidemiol.* 45(6), 613–619 (1992).
- Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat. Med.* 25(24), 4279–4292 (2006).
- Rubin DB. Multiple Imputation after 18+ years. *J. Am. Stat. Assoc.* 91(434), 473–489 (1996).
- **Review of multiple imputation (MI) framework and gives response to criticism of MI, comparing alternative strategies.**
- Rubin DB. *Multiple imputation for nonresponse in surveys.* John Wiley & Sons, NY, USA (1987).
- **Seminal article on MI in research.**
- Molenberghs G, Beunckens C, Sotito C, Kenward MG. Every missingness not at random model has a missingness at random counterpart with equal fit. *J. Roy. Stat. Soc. Ser. B. (Stat. Method.)* 70(2), 371–388 (2008).

17. Bell ML, Fairclough DL, Fiero MH, Butow PN. Handling missing items in the Hospital Anxiety and Depression Scale (HADS): a simulation study. *BMC Res. Notes* 9(1), 479 (2016).
18. Schafer JL. *Analysis of Incomplete Multivariate Data*. Chapman & Hall, Boca Raton, FL. (1997).
19. Von Hippel PT. Regression with missing Ys: an improved strategy for analyzing multiply imputed data. *Sociological Methodol.* 37, 83–117 (2007).
20. White IR, Daniel R, Royston P. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. *Comput. Stat. Data Anal.* 54(10), 2267–2275 (2010).
21. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat. Med.* 30(4), 377–399 (2011).
22. Li P, Stuart EA, Allison DB. Multiple imputation: a flexible tool for handling missing data. *JAMA* 314(18), 1966–1967 (2015).
- **A very approachable primer on MI and missing data mechanisms**
23. Yourman LC, Lee SJ, Schonberg MA, Widera EW, Smith AK. Prognostic indices for older adults: a systematic review. *JAMA* 307(2), 182–192 (2012).