



Published in final edited form as:

J Med Syst. ; 42(10): 199. doi:10.1007/s10916-018-1060-0.

Validation of a Sequential Organ Failure Assessment Score using Electronic Health Record Data

Luis E. Huerta, MD¹, Jonathan P. Wanderer, MD, MPhil^{2,3}, Jesse M. Ehrenfeld, MD, MPH^{2,3,4,5}, Robert E. Freundlich, MD, MS^{2,3}, Todd W. Rice, MD, MSc¹, Matthew W. Semler, MD, MSc¹, SMART Investigators*, and Pragmatic Critical Care Research Group

¹Division of Allergy, Pulmonary, and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN

²Department of Anesthesiology, Vanderbilt University Medical Center, Nashville, TN

³Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN

⁴Department of Surgery, Vanderbilt University Medical Center, Nashville, TN

⁵Department of Health Policy, Vanderbilt University Medical Center, Nashville, TN

Abstract

The sequential organ failure assessment (SOFA) score is a scoring system commonly used in critical care to assess severity of illness. Automated calculation of the SOFA score using existing electronic health record data would broaden its applicability. We performed a manual validation of an automated SOFA score previously developed at our institution. A retrospective analysis of a random subset of 300 patients from a previously published randomized trial of critically ill adults was performed, with manual validation of SOFA scores from the date of initial intensive care unit admission. Spearman's rank correlation coefficient, weighted Cohen's kappa, and Bland-Altman plots were used to assess agreement between manual and electronic versions of SOFA scores and between manual and electronic versions of their individual components. There was high agreement between manual and electronic SOFA scores (Spearman's rank correlation coefficient = 0.90, 95% CI 0.87 – 0.93). Renal and respiratory components had lower agreement (weighted Cohen's kappa = 0.63, 95% CI 0.53 – 0.73 for renal; weighted Cohen's kappa = 0.77, 95% CI 0.70 – 0.84 for respiratory). The area under the receiver operating characteristic curve (AUC) for 30-day in-hospital mortality was 0.77 (95% CI 0.68 – 0.84) for manual SOFA scores and 0.75 (95% CI 0.66 – 0.83) for automated SOFA scores. Automatic calculation of SOFA scores from the electronic health record is feasible and correlates highly with manually calculated SOFA scores. Both have similar predictive value for 30-day in-hospital mortality.

Keywords

Automation; Sepsis; Critical Care; Decision Support Techniques

* A full list of the SMART Investigators may be found in the appendix.

Corresponding Author: Luis E. Huerta, MD, Vanderbilt University Medical Center, 1161 21st Ave S., T-1218 MCN, Nashville, TN 37232-2650, USA, luis.e.huerta@vumc.org, 615-322-3412.

Conflicts of Interest: All authors completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

INTRODUCTION

Objectively quantifying severity of illness and risk of death is an important challenge in critical care clinical research. Scoring systems developed with this aim include the acute physiology and chronic health evaluation (APACHE) score [1], the simplified acute physiology score (SAPS) [2], and the sequential organ failure assessment (SOFA) score [3]. Of these, the SOFA score is the simplest to calculate, as it requires the fewest variables, all of which have clear definitions and are potentially extractable from the electronic medical record, and, after calculation of each individual SOFA component, the final score requires only addition of its 6 components. Although the SOFA score was initially developed to quantify organ failure, more recently it has been used to predict mortality [4–6]. Furthermore, an increase in a patient’s SOFA score by 2 or more points was recently included as a component of the Sepsis-3 definition, widening the score’s clinical applicability [7]. Despite being simpler to calculate than other scores, manual calculation of SOFA scores can be time-consuming, limiting its application to clinical care and its use in large-scale clinical research.

A SOFA score calculated automatically from data available in the electronic health record could eliminate the burden of manual SOFA score calculation and potentially increase its use. Such a tool, however, must be validated against the traditional, manual approach to SOFA score calculation to confirm its accuracy before being applied to research or clinical practice. An electronic SOFA (E-SOFA) score calculator had been previously developed at our institution but had not been rigorously validated [8]. Therefore, we used data on manually calculated SOFA scores collected as part of a recently completed randomized trial to examine the performance characteristics of the E-SOFA score [9]. We hypothesized that the E-SOFA score would correlate closely with the manually collected SOFA score and would predict in-hospital mortality with similar accuracy.

METHODS

Study Design and Oversight

We performed a retrospective analysis of a subset of patients enrolled in the Isotonic Solutions and Major Adverse Renal Events Trial (SMART). SMART was a 15,802-patient, pragmatic, single-center, cluster-randomized, multiple-crossover trial comparing balanced crystalloids versus saline among critically ill adults [9]. Both the initial trial and the current study were approved by the Vanderbilt University Medical Center Institutional Review Board with a waiver of informed consent. Portions of this research were previously presented in abstract form [10].

Patients

Complete inclusion and exclusion criteria for the SMART trial were previously published [9,11]. Briefly, all adults (age 18 years or older) admitted to five intensive care units (medical, surgical, neurologic, trauma, and cardiovascular) at Vanderbilt University Medical Center during the study period were enrolled at the time of initial intensive care unit (ICU)

admission. For the current study, a subset of 300 patients was selected using simple computer-generated randomization from the overall population of SMART participants.

Algorithm Development

The SOFA score contains 6 primary components: cardiovascular, respiratory, renal, hepatic, neurologic, and coagulation [3]. Patients receive between 0 and 4 points for each component, depending on the results of routinely collected clinical and laboratory values, with higher scores signifying increasing severity of illness. The maximum possible SOFA score is 24 points.

The E-SOFA score examined in this study was previously developed in a separate cohort of ICU patients with central line infections [8]. The E-SOFA score was calculated by electronic extraction of all necessary data elements from the institutional electronic data warehouse – a repository generated from a subset of data elements contained within the electronic health record. Within the data warehouse, a list of all study participants was uploaded and all elements necessary for the calculation of a SOFA score were selected. Using a preexisting function, every value for each necessary data element over a specified date range was downloaded. A file containing the data was uploaded into R Statistical Software Version 3.4.1 (R Foundation for Statistical Computing, Vienna, Austria). Once uploaded, an algorithm in R calculated the individual E-SOFA score components and summed them into a final E-SOFA score.

The E-SOFA score contains all values found in the traditional SOFA score, with the exception that the partial pressure of oxygen in arterial blood (PaO₂) to fraction of inspired oxygen (FiO₂) ratios are estimated using oxygen saturation (SpO₂) to fraction of inspired oxygen (FiO₂) ratios if no PaO₂ values are available [3,12,13]. For patients receiving supplemental oxygen via nasal cannula, FiO₂ is estimated based on the oxygen flow in liters per minute (Supplemental Table 1). In our electronic medical record, FiO₂ and oxygen delivery device are clearly labeled in patients receiving supplemental oxygen through invasive or noninvasive mechanical ventilation. Most other noninvasive oxygen delivery devices (including nonrebreather mask, Venturi mask, facemask, and high-flow nasal cannula) are not labeled in a manner that is readily extracted from the EHR, although an FiO₂ is often explicitly recorded. For the purposes of calculating an SpO₂ to FiO₂ ratio, any FiO₂ listed for these patients was assumed to be accurate. For each parameter, the most abnormal value during a calendar day is used to calculate the E-SOFA score. Missing data are assumed to be normal.

As a sensitivity analysis, we examined a modified version of the algorithm which calculated E-SOFA scores using a recently-published modification of the cardiovascular SOFA score component [14]. The modified cardiovascular component incorporates several parameters, such as lactic acid, shock index (the ratio of heart rate to systolic blood pressure), and additional vasoactive agents (e.g., vasopressin), not included in the traditional SOFA score. It outperformed the traditional cardiovascular SOFA score in predicting ICU mortality, in-hospital mortality, and 28-day mortality [14]. Another sensitivity analysis was performed examining a simplified SOFA score excluding vasopressor dose and urine output data due to

concerns that these data might be harder to extract at some institutions (Supplemental Table 2).

Outcomes

The primary outcome in this study was the agreement between the E-SOFA score on the date of ICU admission for each patient and a manual SOFA score for the same date calculated by physician manual review of the electronic health record using a standardized case report form. Discrepant scores were reviewed individually to determine the etiology of the discrepancy.

Secondary outcomes included the agreement between each of the individual manual and E-SOFA score components, between modified manual and E-SOFA scores incorporating the modified cardiovascular component, and between simplified manual and E-SOFA scores. Death prior to the first of hospital discharge or 30 days (30-day in-hospital mortality) was the outcome used to assess the predictive performance of the manual and E-SOFA scores.

Data collection

Demographics and clinical outcomes were obtained from the SMART trial database, which used data recorded in the electronic health record during routine clinical care. Manual review of the medical record with collection of all values needed to manually calculate a SOFA score was performed in all patients in the validation cohort using a standardized form (provided in the electronic supplement). Manual SOFA scores were calculated by the first author, who was blinded with respect to patients' E-SOFA scores at the time of data collection. Further details regarding the manual data extraction process are also provided in the electronic supplement. E-SOFA scores were automatically calculated for patients in the validation cohort, as described previously.

Statistical Methods

Categorical values were described using numbers and percentages, and continuous variables were described using means and standard deviations or medians and interquartile ranges, as appropriate. Agreement between manually collected SOFA scores and E-SOFA scores was evaluated using Spearman's rank correlation coefficient. Agreement between each individual component of the manual and E-SOFA scores was evaluated using a weighted Cohen's kappa [15]. For SOFA score components with lower agreement between the manually and electronically collected versions, agreement between the manually and electronically collected versions of the raw data used to calculate them was measured using Spearman's rank correlation coefficient. This analysis was performed to determine if one specific clinical or laboratory variable was primarily responsible for any discrepancies.

Bland-Altman plots were generated to illustrate the agreement between manual and E-SOFA scores, between each of their individual components, and, for those components with lower agreement, between the manual and electronic versions of the raw data used to generate those components [16].

Unadjusted logistic regression models were fit to evaluate the association between manual and E-SOFA scores with 30-day in-hospital mortality. The area under the receiver operating characteristic (ROC) curve was calculated to determine the discriminative ability of both manual and E-SOFA scores for 30-day in-hospital mortality. Statistical significance was set at a 2-sided P-value less than 0.05. All statistical analyses were performed with R Version 3.4.1 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Of 15,802 patients in the SMART trial dataset [9], 300 were randomly selected for the current study. Among these, the median age was 59 years [IQR 45–71], and 57% were male (Table 1). Thirty-six percent of patients were admitted to the medical ICU. The median manually calculated SOFA score was 5 [IQR 2–8]. The median E-SOFA score was 5 [IQR 3–8]. A total of 39 patients (13.0%) experienced 30-day in-hospital mortality.

The manual SOFA and E-SOFA scores for each patient were highly correlated (Spearman's rank correlation coefficient 0.90; 95% CI 0.87 – 0.93). The mean difference between manual and E-SOFA scores was 0.34 (95% CI –2.47 to 3.14) (Table 2, Fig. 1). The mean absolute value of the difference between the manual and E-SOFA scores was 0.80 (95% CI 0.67 – 0.94), and the median was 0 [IQR 0–1] (Table 2). The largest difference between an individual patient's manual and E-SOFA scores was 7, which occurred in one patient. Two patients differed in their manual and E-SOFA scores by 6 points, one by 5 points, and the remaining 296 patients had manual and E-SOFA scores within 4 points of each other (Fig. 1).

Manual and E-SOFA scores were highly correlated for each of the cardiovascular, hepatic, coagulation, and neurologic components, with all 4 having a weighted Cohen's kappa greater than 0.95 (Table 2, Supplemental Figs. 1–4). Agreement between the manual and E-SOFA scores was lower for the respiratory and renal components (weighted Cohen's kappa 0.77; 95% CI 0.70 – 0.84 for respiratory and weighted Cohen's kappa 0.63; 95% CI 0.53 – 0.73 for renal components) (Table 2, Supplemental Figs. 5–6). Etiologies of discrepancies are described in Supplemental Table 3.

When analyzing the values used to calculate the renal component, creatinine values collected manually and electronically were highly correlated (Spearman's rank correlation coefficient = 1.00; 95% CI 1.00 – 1.00) (Supplemental Fig. 7). Agreement between manual and electronically collected values for daily urine output, however, was lower (Spearman's rank correlation coefficient = 0.72; 95% CI 0.63 – 0.79) (Supplemental Fig. 8). Manual review of discrepancies between manually and electronically collected urine output values demonstrated failure of the electronically collected values to distinguish anuria from improperly recorded urine output among patients who were not anuric. Twelve patients had missing electronically collected urine output data that were able to be successfully calculated manually (Supplemental Table 3). Forty-nine patients had discrepancies due to electronically collected urine output not being adjusted for the time of admission.

For the values used to calculate the respiratory component of the SOFA score, a high correlation was observed between electronic and manual PaO₂ to FiO₂ ratios (Spearman's rank correlation coefficient = 0.88; 95% CI 0.79 – 0.95) (Supplemental Fig. 9). Correlation was lower between manually and electronically collected SpO₂ to FiO₂ ratios (Spearman's rank correlation coefficient = 0.62; 95% CI 0.51 – 0.74) (Supplemental Fig. 10). Discrepancies primarily derived from difficulty capturing FiO₂ values at the same time that an SpO₂ value was recorded for some patients (Supplemental Table 3). Also, 24 patients had arterial blood gases which were not captured by our data extraction technique. These arterial blood gases were primarily those obtained in the perioperative setting prior to the patient returning to the intensive care unit (n = 16) (Supplemental Table 3).

The manual and electronic versions of the modified cardiovascular SOFA score component exhibited excellent agreement (weighted Cohen's kappa = 0.99; 95% CI 0.97 – 1.00) (Supplemental Table 4, Supplemental Fig. 11). Similarly, manual and E-SOFA scores incorporating the modified cardiovascular component were highly correlated (Spearman's rank correlation coefficient = 0.91; 95% CI 0.88 – 0.94) (Supplemental Table 4, Supplemental Fig. 12). Simplified manual and E-SOFA scores excluding the vasopressor dose and urine output data were also highly correlated (Spearman's rank correlation coefficient = 0.97; 95% CI 0.95 – 0.98) (Supplemental Table 4, Supplemental Fig. 13), as were the simplified manual and electronic cardiovascular and renal components (Supplemental Table 4, Supplemental Figs. 14–15).

In separate logistic regression models, manual SOFA scores (p < 0.001; OR per point 1.26; 95% CI 1.16 – 1.36) and E-SOFA scores (p < 0.001; OR per point 1.24; 95% CI 1.14 – 1.35) were each significantly associated with 30-day in-hospital mortality. The area under the ROC curve (AUC) for 30-day in-hospital mortality was 0.77 (95% CI 0.68 – 0.84) for manual SOFA scores and 0.75 (95% CI 0.66 – 0.83) for E-SOFA scores (Fig. 2). The simplified E-SOFA score was also significantly associated with 30-day in-hospital mortality (p < 0.001; OR per point 1.31; 95% CI 1.18 – 1.44) and had similar predictive ability for that outcome (AUC = 0.76, 95% CI 0.67 – 0.84) (Supplemental Fig. 16).

DISCUSSION

This retrospective analysis of a random subset of patients from a randomized trial demonstrated high correlation between manually calculated SOFA scores and SOFA scores calculated using data automatically extracted from the electronic health record. The manual SOFA and E-SOFA scores demonstrated similar predictive ability for 30-day in-hospital mortality. These findings have short-term implications for clinical research and potential long-term implications for the monitoring and care of hospitalized patients.

Several characteristics of the E-SOFA score contributed to its success. First, it incorporated all SOFA components. Second, the comprehensive nature of our electronic health record meant that we were able to extract nearly all necessary data, and efforts are ongoing to refine extraction of the remainder. Third, as some variables for which a numeric value was needed were occasionally recorded as text (e.g., an FiO₂ of 0.21 recorded as “room air”), our algorithm was specifically designed to capture this information.

While our E-SOFA score validated well, the accuracy of the urine output and SpO₂ to FiO₂ ratio could be improved. Accurate urine output was difficult to calculate for two reasons. First, two distinct populations frequently lacked urine output data: anuric patients and ambulatory patients urinating in the restroom. The current algorithm assumes all missing data are normal and cannot distinguish them. Imputation of missing urine output might improve its predictive power. Second, because E-SOFA scores were calculated for each calendar day, patients admitted later in the day had a lower proportion of their daily urine output recorded, resulting in a falsely low urine output on the date of admission. Adjusting urine output for time of admission may be more accurate. Calculating E-SOFA scores for the first 24 hours after admission, as opposed to the first calendar day after admission, would also fix this issue.

Discrepant SpO₂ to FiO₂ ratios arose from difficulties in determining an accurate FiO₂ at the time SpO₂ was measured. Due to frequent FiO₂ titration in some patients, only values recorded at the time of an SpO₂ measurement are currently used to calculate an E-SOFA score. In some cases, however, FiO₂ can be inferred during manual chart review from values recorded near the time of SpO₂ measurement. As a result, using FiO₂ values within a specified time range of a measured SpO₂ to calculate an SpO₂ to FiO₂ ratio may improve the algorithm's accuracy.

Several prior studies have described the development of electronic SOFA scores automatically collected from the medical record [17–21]. Prior electronic SOFA scores did not always validate the scores generated by electronic extraction against the reference-standard of a manually reviewed SOFA score [18,19]. Additionally, prior studies frequently modified or excluded significant components of the original SOFA score [18,19]. Of the prior studies that did validate against a manual SOFA score, one study included only 50 patients [17] and two studies did not provide detailed data regarding their validation process [17,20]. Our study provides a detailed validation against a manually calculated SOFA score among a relatively large number of patients and incorporates all SOFA score components.

In the prior study most similar in design to the current work, Harrison et al. manually validated an automated electronic SOFA score and its individual components in a patient population similar to ours [21]. They reported a mean difference between manual and electronically calculated SOFA scores of 0.29 and a standard deviation of 1.75, similar to our results. Harrison et al. also reported a higher rate of discrepancies in the respiratory and renal components of the manual and automated SOFA scores, as was noted in our study, although the rate of discrepancies in these components was significantly improved in a later study prospectively validating an updated, near real-time version of their algorithm [22]. An important difference in findings was that Harrison et al. reported difficulty collecting some GCS scores electronically, which we did not observe in our study, although again, the authors appear to have resolved this issue by the time a later study was published [22].

This study has several strengths. First, it is larger than most previously reported studies of electronically generated SOFA scores. Our study also includes a broad population of both medical and surgical ICU patients, increasing its generalizability. Our E-SOFA score uses all six SOFA score components, without simplification or exclusion of any component.

Furthermore, it incorporates recent modifications to the SOFA score intended to improve its accuracy and widen its applicability. Specifically, use of the SpO₂ to FiO₂ ratio in cases where the PaO₂ to FiO₂ ratio is not available has been utilized by only one other group validating electronic SOFA scores, to our knowledge [21,22]. In addition to the traditional SOFA score, we have also validated our ability to accurately extract a modified cardiovascular component of the SOFA score. To our knowledge, validation of the modified cardiovascular SOFA score component has not been reported outside of the institution that developed the score [14], a notable result given the modified cardiovascular score's improved ability to predict mortality. Finally, we validated a simplified SOFA score for use in healthcare systems with limitations on extractable data. A validated severity of illness score that is easier to automatically collect may assist in the performance of clinical research across a broader range of institutions.

In addition, our study has important limitations. It is a single-center study, limiting generalizability, particularly given that the accuracy of the E-SOFA calculator is dependent on the accuracy of the data input, which may vary by institution. A single investigator performed all manual SOFA score calculations, introducing the potential for errors in manual data collection. This validation was performed retrospectively and even the manually calculated SOFA scores relied on data recorded in the electronic health record. Prospective data collection by study personnel at the bedside might be more accurate than SOFA scores calculated from manual chart review. FiO₂ was estimated in some patients by oxygen flow in liters per minute, which may introduce significant variability [23,24]. GCS scores in sedated patients were not able to be excluded, which may affect the accuracy of the neurologic component. At least one single-center study reported a near real-time SOFA score calculator for immediate clinical use [22]. Our current SOFA score calculator is not currently designed for such rapid turnaround and such near real-time use of our E-SOFA score was therefore not evaluated. Finally, our calculator is not currently integrated into our medical record system, requiring extraction of the data prior to calculation of the E-SOFA score. A fully integrated calculator would broaden the potential applications of the score.

CONCLUSION

Calculation of SOFA scores from data automatically extracted from the electronic health record is feasible and correlates highly with manually calculated SOFA scores. Future studies should focus on tighter integration of automated severity of illness scores into the electronic health record and whether provision of a near real-time E-SOFA score to clinicians alters clinical management or outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Source of Funding: Financial support for the study was provided by the Vanderbilt Institute for Clinical and Translational Research (UL1 TR000445 and UL1TR002243 from NCATS/NIH).

L.E.H. was supported in part by a National Institute of Allergy and Infectious Diseases (NIAID) T32 award (5T32AI095202). M.W.S. was supported in part by the National Heart, Lung, and Blood Institute (NHLBI) (K12HL133117). R.E.F was supported in part by a grant from the Vanderbilt Faculty Research Scholars (KL2TR002245). The funding institutions had no role in: conception, design, or conduct of the study; collection, management, analysis, interpretation, or presentation of the data; preparation, review, or approval of the manuscript; or the decision to submit for publication.

T.W.R. reported serving on an advisory board for Avisa Pharma, LLC and as the Director of Medical Affairs for Cumberland Pharmaceuticals, Inc. R.E.F reported receiving funding from Medtronic and serving as a consultant to Medtronic.

REFERENCES

1. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE (1981) APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 9:591–7 10.1097/00003246-198108000-00008 [PubMed: 7261642]
2. Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, et al. (1984) A simplified acute physiology score for ICU patients. *Crit Care Med* 12:975–7 10.1097/00003246-198411000-00012 [PubMed: 6499483]
3. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. (1996) The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med* 22:707–10 10.1007/s001340050156 [PubMed: 8844239]
4. Jones AE, Trzeciak S, Kline JA (2009) The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation. *Crit Care Med* 37:1649–54 10.1097/CCM.0b013e31819def97 [PubMed: 19325482]
5. Vincent J-L, de Mendonca A, Cantraine F, Moreno R, Takala J, Suter PM, et al. (1998) Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units. *Crit Care Med* 26:1793–800 10.1097/00003246-199811000-00016 [PubMed: 9824069]
6. Minne L, Abu-Hanna A, de Jonge E (2009) Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. *Crit Care* 12:R161 10.1186/cc7160
7. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. (2016) The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 315:801 10.1001/jama.2016.0287 [PubMed: 26903338]
8. Huerta LE, Nelson GE, Rice TW (2017) The Effect of Antimicrobial Treatment Duration on Mortality and Recurrence in Hospital-Acquired Central Line-Associated Bloodstream Infections. In: D24 CRITICAL CARE: THE OTHER HALF OF THE ICU - UPDATE IN MANAGEMENT OF NON-PULMONARY CRITICAL CARE American Thoracic Society; p. A7148 10.1164/ajrccm-conference.2017.195.1_MeetingAbstracts.A7148
9. Semler MW, Self WH, Wanderer JP, Ehrenfeld JM, Wang L, Byrne DW, et al. (2018) Balanced Crystalloids versus Saline in Critically Ill Adults. *N Engl J Med* 378:829–39 10.1056/NEJMoa1711584 [PubMed: 29485925]
10. Huerta LE, Semler MW (2018) Validation of an Electronic Sequential Organ Failure Assessment Score in Intensive Care Unit Patients. In: B104 CRITICAL CARE: BIG DATA IN HEALTH CARE - PREDICTIVE ANALYTICS, CLINICAL DECISION SUPPORT, AND RAPID RESPONSE American Thoracic Society; p. A4289–A4289 10.1164/ajrccm-conference.2018.197.1_MeetingAbstracts.A4289
11. Semler MW, Self WH, Wang L, Byrne DW, Wanderer JP, Ehrenfeld JM, et al. (2017) Balanced crystalloids versus saline in the intensive care unit: Study protocol for a cluster-randomized, multiple-crossover trial. *Trials* 18:1–13 10.1186/s13063-017-1871-1 [PubMed: 28049491]
12. Brown SM, Duggal A, Hou PC, Tidswell M, Khan A, Exline M, et al. (2017) Nonlinear Imputation of PaO₂/FIO₂ From SpO₂/FIO₂ Among Mechanically Ventilated Patients in the ICU. *Crit Care Med* 45:1317–24 10.1097/CCM.0000000000002514 [PubMed: 28538439]
13. Pandharipande PP, Shintani AK, Hagerman HE, St Jacques PJ, Rice TW, Sanders NW, et al. (2009) Derivation and validation of SpO₂/Fio₂ ratio to impute for Pao₂/Fio₂ ratio in the respiratory

- component of the Sequential Organ Failure Assessment score*. Crit Care Med 37:1317–21 10.1097/CCM.0b013e31819cefa9 [PubMed: 19242333]
14. Yadav H, Harrison AM, Hanson AC, Gajic O, Kor DJ, Cartin-Ceba R (2015) Improving the accuracy of cardiovascular component of the sequential organ failure assessment score. Crit Care Med 43:1449–57 10.1097/CCM.0000000000000929 [PubMed: 25785522]
 15. Cohen J (1968) Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychol Bull 70:213–20 10.1037/h0026256 [PubMed: 19673146]
 16. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–10 10.1016/S0140-6736(86)90837-8 [PubMed: 2868172]
 17. Thomas M, Bourdeaux C, Evans Z, Bryant D, Greenwood R, Gould T (2011) Validation of a computerised system to calculate the sequential organ failure assessment score. Intensive Care Med 37:557 10.1007/s00134-010-2083-2 [PubMed: 21152902]
 18. Nates JL, Cárdenas-Turanzas M, Wakefield C, Kish Wallace S, Shaw A, Samuels JA, et al. (2010) Automating and simplifying the SOFA score in critically ill patients with cancer. Health Informatics J 16:35–47 10.1177/1460458209353558 [PubMed: 20413411]
 19. Junger A, Engel J, Benson M, Böttger S, Grabow C, Hartmann B, et al. (2002) Discriminative power on mortality of a modified Sequential Organ Failure Assessment score for complete automatic computation in an operative intensive care unit. Crit Care Med 30:338–42 10.1097/00003246-200202000-00012 [PubMed: 11889305]
 20. Zygun DA, Laupland KB, Fick GH, Sandham JD, Doig CJ (2005) Limited ability of SOFA and MOD scores to discriminate outcome: a prospective evaluation in 1,436 patients. Can J Anaesth 52:302–8 10.1007/BF03016068 [PubMed: 15753504]
 21. Harrison AM, Yadav H, Pickering BW, Cartin-Ceba R, Herasevich V (2013) Validation of Computerized Automatic Calculation of the Sequential Organ Failure Assessment Score. Crit Care Res Pract 2013:1–8 10.1155/2013/975672
 22. Aakre C, Franco PM, Ferreyra M, Kitson J, Li M, Herasevich V (2017) Prospective validation of a near real-time EHR-integrated automated SOFA score calculator. Int J Med Inform 103:1–6 10.1016/j.ijmedinf.2017.04.001 [PubMed: 28550994]
 23. Bazuaye EA, Stone TN, Corris PA, Gibson GJ (1992) Variability of inspired oxygen concentration with nasal cannulas. Thorax 47:609–11 <http://www.ncbi.nlm.nih.gov/pubmed/1412117>. [PubMed: 1412117]
 24. Markovitz GH, Colthurst J, Storer TW, Cooper CB (2010) Effective inspired oxygen concentration measured via transtracheal and oral gas analysis. Respir Care 55:453–9 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20406513. [PubMed: 20406513]

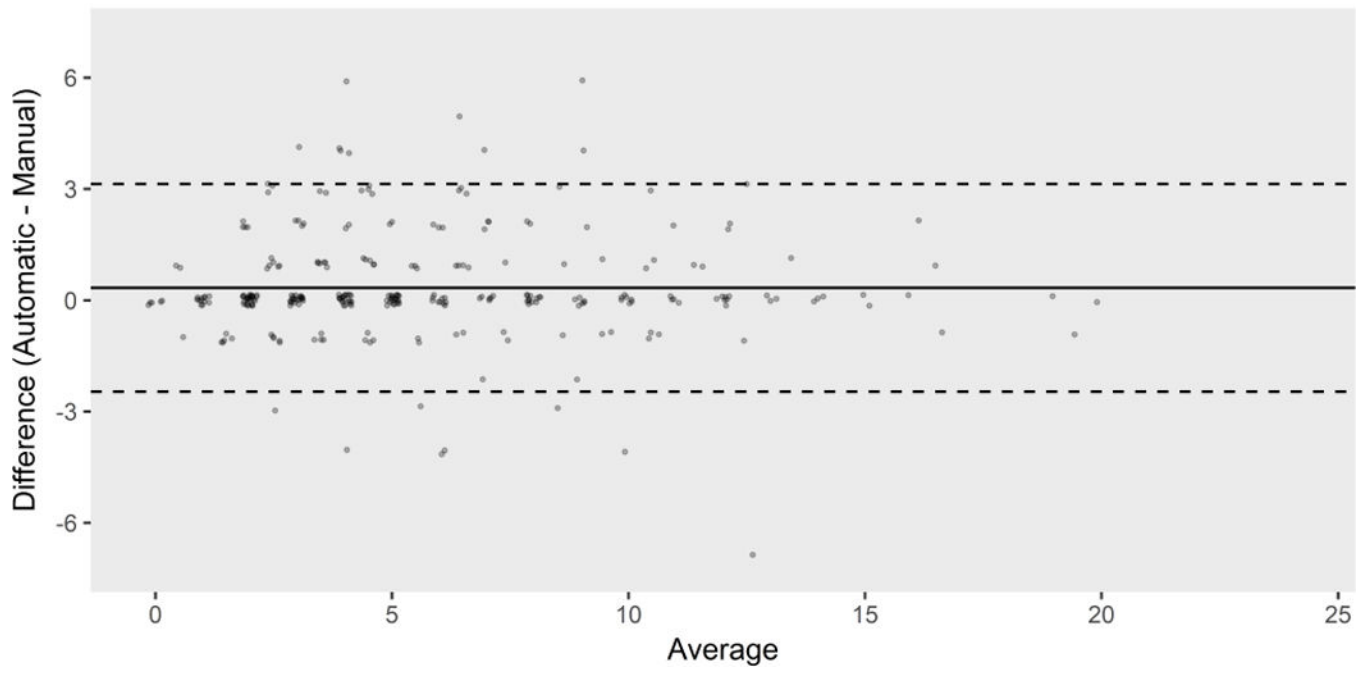


Fig. 1. Bland-Altman plot of manual and E-SOFA scores (N = 300). The mean difference between the scores was 0.34 (95% CI -2.47 – 3.14). E-SOFA = electronic sequential organ failure assessment

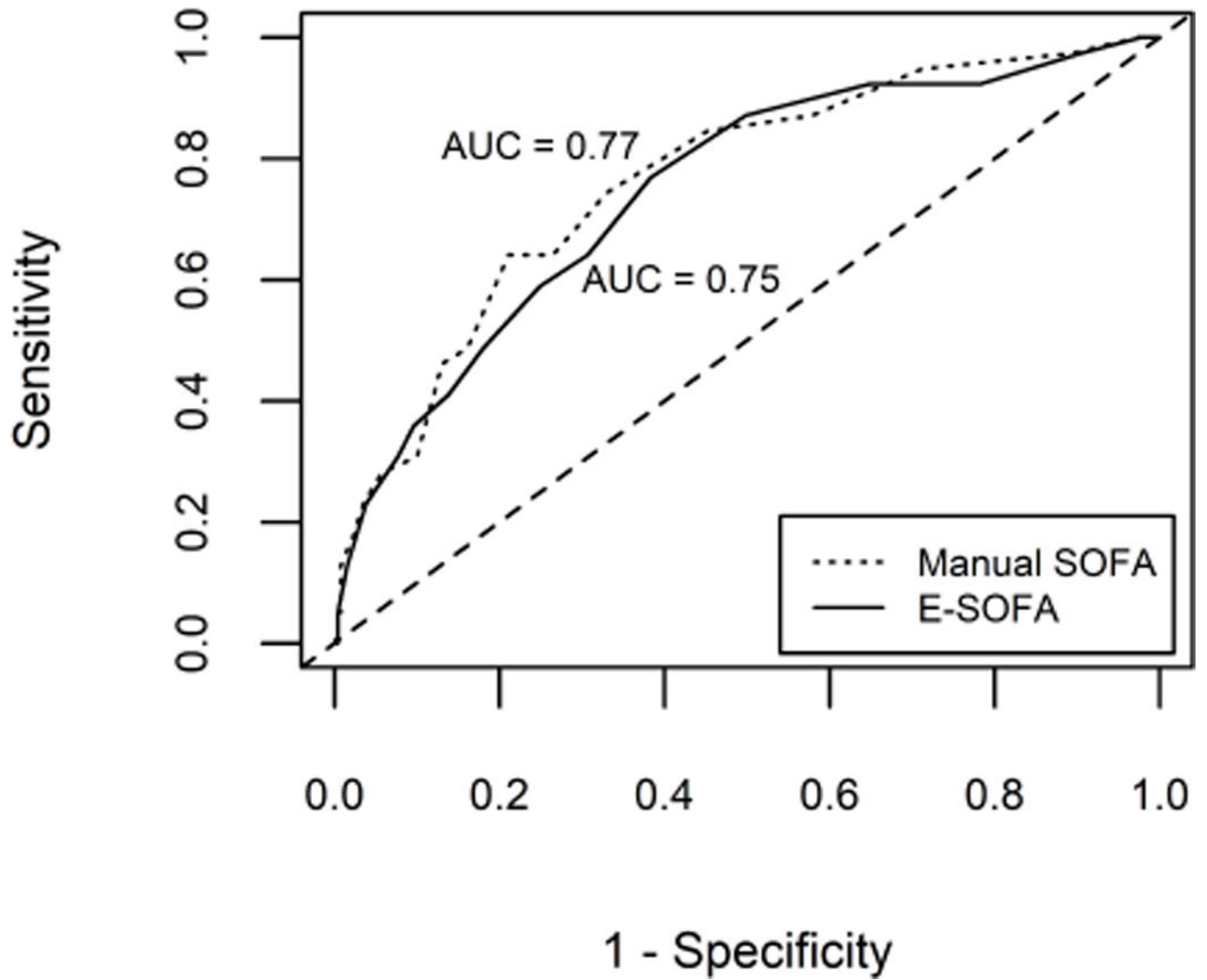


Fig. 2. ROC curves for manual and E-SOFA scores for 30-day in-hospital mortality (N = 300). ROC = receiver operating characteristic; E-SOFA = electronic sequential organ failure assessment; AUC = area under the receiver operating characteristic curve

Table 1:

Baseline Characteristics

Characteristics	Validation Cohort N = 300
Age	59.1 [45.0–71.1]
Male Sex	171 (57)
White Race	245 (81.7)
Enrollment ICU	
Medical	108 (36)
Surgical	24 (8.3)
Neurologic	56 (18.7)
Cardiac	55 (18.3)
Trauma	56 (18.7)
Source of Admission	
Emergency Department	158 (52.7)
Operating Room	50 (16.7)
Transfer from Another Hospital	50 (16.7)
Hospital Ward	24 (8.0)
Other	18 (6.0)
Manual SOFA Score on ICU Admission Date	5 [2–8]
E-SOFA Score on ICU Admission Date	5 [3–8]
ICU Length of Stay	2.2 [1.2–4.6]
Hospital Mortality	39 (13)

ICU = intensive care unit; SOFA = sequential organ failure assessment

Values are presented as number (percentage) or median [interquartile range] as appropriate

All outcomes censored 30 days after enrollment (date of ICU admission)

Table 2:

Agreement Between Manual and E-SOFA Scores

SOFA Score Component (N = 300)	Mean Difference ^a (95% Confidence Interval)	Mean Absolute Difference ^b (95% Confidence Interval)	Agreement ^c (95% Confidence Interval)
Cardiovascular	-0.02 (-0.64 – 0.60)	0.05 (0.02 – 0.09)	0.97 (0.93 – 1.00)
Respiratory	0.01 (-1.33 – 1.35)	0.32 (0.25 – 0.39)	0.77 (0.70 – 0.84)
Renal	0.33 (-1.95 – 2.61)	0.52 (0.39 – 0.64)	0.63 (0.53 – 0.73)
Hepatic	0.00 (-0.11 – 0.12)	0.00 (0.00 – 0.01)	1.00 (0.99 – 1.00)
Neurologic	0.00 (-0.11 – 0.12)	0.00 (0.00 – 0.01)	1.00 (1.00 – 1.00)
Coagulation	0.01 (-0.24 – 0.26)	0.01 (0.00 – 0.02)	0.99 (0.97 – 1.00)
Full SOFA Score	0.34 (-2.47 – 3.14)	0.80 (0.67 – 0.94)	0.90 (0.87 – 0.93)

SOFA = sequential organ failure assessment

^a mean difference = mean difference between E-SOFA score component and manual SOFA score component

^b mean absolute difference = mean of the absolute value of the difference between E-SOFA score component and manual SOFA score component

^c agreement was measured by Spearman's rank correlation coefficient for the full SOFA score, and by weighted Cohen's Kappa for individual SOFA components