Published in final edited form as:

Crit Care Med. 2017 December ; 45(12): 2014–2022. doi:10.1097/CCM.0000000000002708.

Serial daily organ failure assessment beyond ICU day 5 does not independently add precision to ICU risk-of-death prediction

Andre L. Holder, MD MSc¹, Elizabeth Overton, MSPH¹, Peter Lyu, MSPH², Jordan A Kempker, MD MSc¹, Shamim Nemati, PhD³, Fereshteh Razmi³, Greg S. Martin, MD MSc^{1,4}, Timothy G. Buchman, MD PhD⁴, and David J. Murphy, MD PhD^{1,4}

¹Division of Pulmonary, Critical Care, Allergy and Sleep Medicine; Emory University School of Medicine

²Graduate School of Arts and Sciences, Harvard University

³Department of Biomedical Informatics; Emory University School of Medicine

⁴Emory Critical Care Center, Emory Healthcare

Abstract

OBJECTIVE—To identify circumstances in which repeated measures of organ failure would improve mortality prediction in ICU patients.

DESIGN—Retrospective cohort study, with external validation in a de-identified ICU database.

SETTING—Eleven ICUs in three university hospitals within an academic healthcare system in 2014.

PATIENTS—Adults (18 years or older) who satisfied the following criteria: (1) 2 of 4 systemic inflammatory response syndrome criteria plus an ordered blood culture, all within 24 hours of hospital admission; and (2) ICU admission for at least 2 calendar days, within 72 hours of emergency department presentation.

INTERVENTION—None

MEASUREMENTS AND MAIN RESULTS—Data were collected until death, ICU discharge, or the seventh ICU day, whichever came first. The highest SOFA score from the ICU admission day (ICU day 1) was included in a multivariable model controlling for other covariates. The worst SOFA scores from the first 7 days after ICU admission were incrementally added and retained if they obtained statistical significance ($p < 0.05$). The cohort was divided into seven subcohorts to facilitate statistical comparison using the integrated discriminatory index (IDI). Of the 1290 derivation cohort patients, 83 patients (6.4%) died in the ICU, compared with 949 of the 8441 patients (11.2%) in the validation cohort. Incremental addition of SOFA data up to ICU day 5

Corresponding author: Andre L. Holder, MD, MS, Assistant Professor of Medicine, Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Emory University School of Medicine, Grady Memorial Hospital, 80 Jesse Hill Jr. Drive SE, Rm 2D012, Atlanta, GA 30303, Phone: 404.616.0823, Fax: 404.616.8455, andre.holder@emory.edu.

Financial support: none

Conflicts of interest: The remaining authors have disclosed that they do not have any remaining conflicts of interest.

improved the IDI in the validation cohort. Adding ICU day 6 or 7 SOFA data did not further improve model performance.

CONCLUSIONS—Serial organ failure data improves prediction of ICU mortality, but a point exists after which further data no longer improves ICU mortality prediction of early sepsis.

Keywords

SOFA; organ failure; sepsis; prognostication; ICU; mortality

INTRODUCTION

Worsening organ dysfunction is associated with higher mortality in septic patients who are critically ill. [1] The sequential organ failure assessment (SOFA) score is a validated tool used to temporally track changes in physiologic derangements, providing real time organ dysfunction data. [1–3] Predictive modeling using repeated measurements over time may improve prognostication, but this comes at a cost. Understanding the dynamics of organ failure and how they are impacted by therapeutic interventions requires more data collection, processing, and time to wait for these changes to occur.

After a certain point, repeated organ failure assessment may be neither practical nor necessary. Some suggest that only very early changes in organ function responses (within 1 day) are truly associated with organ failure data. [4] The objective of this study is to identify a point in the ICU stay after which mortality prediction no longer improves with additional organ failure data collected over time. What predictive value do clinicians, patients and families gain from the data, and what is sacrificed to achieve that gain?

MATERIALS AND METHODS

Study Population

Patients aged 18 years or older with suspected sepsis in the ICU, defined by the following: 2 out of 4 systemic inflammatory response syndrome (SIRS) criteria [5] within 4 hours of each other and had a blood culture ordered (indicating suspicion of infection), all within 24 hours. They must have been admitted to the ICU for at least 2 calendar days from ICU admission (to compare SOFA scores from at least two different days). Patients were excluded for any of the following: (1) they were transferred from an outside hospital; (2) they were not admitted through the ED, or that information was unknown; or (3) they were admitted to the ICU greater than 72 hours from ED presentation (a surrogate to exclude inpatient ward transfers to the ICU). Only the first ICU stay within a single hospital admission was included in the study. Data was obtained on patients for the first 7 days of their ICU stay, or until ICU discharge or death, whichever came first.

Statistical Analysis

Though the primary model outcome was ICU mortality (with hospital mortality was a secondary model outcome), the primary *objective* was to note independent *improvement* in ICU mortality prediction with incremental addition of daily SOFA scores to a multivariate

model. The primary “exposures” of interest were the SOFA scores from ICU calendar days 1 through 7.

The SOFA score for a given ICU day was based on the sum of the worst scores for all component organ failures in that calendar day. Evidence of neurologic, respiratory, cardiovascular, renal, hematologic, and hepatic dysfunction were extracted every other day for the first seven days of ICU admission. Our dataset did not include doses of vasoactive agents, therefore cardiovascular dysfunction over a score of 1 was based on whether a drug was administered regardless of dose. All other organ failure calculations were based on the original SOFA score definitions (Appendix Table 1). [6] Component organ scores were assigned a value of 0 if the raw data from which they were derived was missing or nonsensical (e.g. a pO₂ of 1 mmHg).

Demographic characteristics, comorbid illness data, and other hospital course variables were also examined. Patient demographic covariates included age, sex, and race. Chronic comorbid illnesses were examined using the Charlson-Deyo Comorbidity Index [CCI]. [7, 8] Hospital course data included the type of ICU(s) to which patients were admitted, and abnormal SIRS parameters at ICU admission.

A derivation cohort was selected to construct all models. Bivariate analyses were conducted to assess the relationship between each candidate predictor and mortality. All continuous variables were assessed for normality and analyzed using Student’s t-test or Wilcoxon rank-sum test, as appropriate. Categorical variables were analyzed using Pearson’s chi-squared. Only candidate variables associated with mortality on bivariate analysis with a $p < 0.25$ were eligible for inclusion as covariates in multivariable logistic regression models.

The cohort was stratified into 6 “subcohorts” (Figure 1) to account for patient attrition and statistically assess the effect of adding SOFA scores from days following ICU admission. (See Appendix Tables 3 & 4 for patient characteristics of each subcohort.) Six corresponding “base models” were developed to evaluate the independent association between the SOFA score on ICU day 1 (ICU admission) and ICU mortality in each subcohort. Age, CCI, ICU type (surgical, nonsurgical, or mixed), and the SOFA score on ICU day 1 were forced *a priori* in this, and all models regardless of statistical significance. The base models were used to build other models by incrementally adding SOFA scores from ICU days 2, 3, and so on, up to ICU day 7. All models were then regularized to prevent overfitting. In particular, an elastic net regularization method was used, and an internal 10-fold cross-validation was employed to find the optimal value of the regularization parameter (lambda). [9] This method ensures generalizability of the model to out-of-sample data (validation cohort).

Models were compared to each other using the integrated discrimination index (IDI), a reclassification score that assesses improvement in model performance when a new biomarker is added to an existing model. In this example, the IDI is the sum of the improvement in the average sensitivities and the reduction in the average false positive rates as a SOFA score is added to a subcohort model. (More positive values indicate improvement in performance.) IDIs were only used to compare models within a subcohort (i.e., those derived from samples of equal size). A set of predetermined comparisons assessed the

incremental improvement in mortality prediction with the addition of a single SOFA score up to the complete model for that subcohort. (Appendix Table 5 & 6) *Post hoc* comparisons were also performed within subcohorts to note improvement in model performance with addition of more than one SOFA score. (All reported IDIs were multiplied by 100.)

The results of all model comparisons were confirmed in an external validation cohort to assess the generalizability of those results. Bonferroni correction was applied in the validation group to account for multiple IDI comparisons (21 prespecified comparisons with ICU mortality as outcome, and 21 prespecified comparisons with hospital mortality as outcome=42 comparisons; $p=0.05/42=0.0012$).

All statistical analyses were performed with STATA 15 (StataCorp LP, College Station, TX, 2016). A two-sided p -value <0.05 was used to determine statistical significance in all tests, unless otherwise stated. This study was approved by the Emory Institutional Review Board.

Please see methods supplement for more details on study settings and data sources.

RESULTS

After all exclusions applied to the derivation cohort, 1290 patients were included for analysis (Figure 2). A total of 83 patients (6.4%) died during their ICU stay, and 122 (9.4%) died during hospitalization. (Table 1) Patients who died in the derivation cohort had a higher percentage of abnormal white blood cell counts on ICU admission (73.5% vs. 56.3%; $p<0.01$), had longer median days on the ventilator (2 days vs. 0 days; $p<0.01$), longer median ICU lengths of stay (5 days vs. 3 days; $p=0.01$), and shorter median hospital lengths of stay (5 days vs. 10 days; $p<0.01$). Patients in the validation group who died tended to be older (median age 74 vs. 65; $p<0.01$) with a lower proportion of African Americans (9.4% vs. 14.6%; $p<0.01$); neither relationship was seen in the derivation cohort. Compared to the external validation group (8441 patients), the derivation cohort had a lower percentage of whites (31.5% vs. 73.5%), less MICU patients (53.9% vs. 66.3%), and had higher maximum and mean SOFA scores (8 vs. 5, and 5.2 vs. 3.5, respectively) but lower ICU and hospital mortality (6.4% vs. 11.3%, and 9.4% vs. 20.1%). (Table 1 & Appendix Table 2)

In the derivation cohort, there was an independent incremental improvement (IDI with $p<0.05$) in the addition of daily SOFA up to ICU day 4, but only in subcohorts 1 through 4. (Table 2, bolded items in top panel; Figure 3, top panel). There was statistically significant improvement ($p<0.05$) in the IDI of all subcohorts in the adjusted models up to ICU day 4 when aggregated SOFA scores over multiple consecutive days were added to the analysis post hoc. (Appendix Figure 1, top panel)

Incremental addition of daily SOFA scores in the validation cohort showed no improvement in adjusted ICU mortality prediction after ICU day 5 in the predetermined analysis (Table 2, bolded items in bottom panel; Figure 3, bottom panel). The point of minimal improvement in ICU mortality was extended to ICU day 6 when aggregate SOFA data were added to adjusted models in post hoc analysis (Appendix Figure 1, bottom panel).

Data after ICU day 3 no longer showed improvements in IDI with additional SOFA data to predict hospital mortality in the derivation cohort (Appendix Table 7, top panel; Appendix Figure 2, top panel). However, this point did not exist in the validation cohort; serial SOFA score data collection improved hospital mortality prediction in the validation cohort throughout the analysis period, or up to 7 days post ICU admission. (Appendix Table 7, bottom panel; Appendix Figure 2, bottom panel)

DISCUSSION

In this study, we sought to identify the conditions in which organ failure data collected over time would improve ICU mortality prediction. We found statistically significant improvement in ICU mortality prediction when daily SOFA scores from ICU admission to day 5 were incrementally added to a multivariate model applied to our validation cohort. No such demonstrable improvement in ICU mortality prediction occurred by adding SOFA data after ICU day 5. Serial addition of organ failure data consistently improved hospital mortality prediction throughout our data collection period.

To our knowledge, this is the first study to demonstrate appropriate conditions for serial organ failure assessment. Though SOFA score is a useful prognostic tool in the ICU, prognostic capabilities could be improved by further organ failure data. [2, 3, 10, 11] This is the first study to demonstrate a point after which serial daily ICU data no longer improves ICU risk prediction. We used a large cohort to assess the independent relationship between additional organ failure and ICU mortality prediction, and confirmed this in a much larger external cohort. Our study accounts for attrition bias by performing analyses in six subcohorts of the larger dataset. This has the added effect of seeing how length of stay affects the amount of SOFA data needed before it no longer impacts ICU mortality prediction. Those who stayed in the ICU past ICU day 6 in the derivation and validation cohorts did not experience any statistically significant improvement with incremental addition of daily SOFA scores. (Table 2)

This study used the IDI to assess the added value of serial SOFA scores to ICU risk prediction. The IDI provides more relevant clinical information over other summary metrics such as the AUC since it assesses improvements in the sensitivity and specificity regardless of the risk threshold. IDI-based comparisons proved similar to AUC comparisons in one sepsis-related study. [12] However, improvements in the AUC can be nominal in magnitude, even when a new biomarker adds clinically meaningful predictive power to model performance. [13] Furthermore, the IDI may be easier to interpret since it directly relates to sensitivity and specificity. In our study, an IDI of 5.38% with the addition of SOFA day 2 to an adjusted model with SOFA day 1 (for validation subcohort 1) means that there was a collective improvement in sensitivity or specificity of 5.38%. This was associated with a corresponding increase in the AUC from 0.69 to 0.77. (Table 2)

The appropriateness of how and at what point to use serial SOFA assessment has important implications for clinicians, patients and family, administrators, and researchers. Clinicians can use SOFA score in a validated risk prediction tool to decide if it is appropriate to recommend aggressive care in the setting of persistent or worsened organ failure despite

appropriate therapy. Goals of care discussions may be more timely with a prediction model incorporating incremental organ failure data, since clinicians would have all the information needed by ICU day 5.

Hospital administrators may find our results useful towards accelerating value. Sepsis is the most expensive condition treated in U.S. hospitals, and the cost is rising annually. [14] Others have suggested using changes in SOFA throughout the hospital course as a way of streamlining care. [15] For example, an EMR alert could prompt clinicians to have early goals of care discussions if risk of death is above a certain point, based on SOFA score dynamics during the first few days of ICU care. Strategic healthcare application may reduce spending without sacrificing care quality.

Once other investigators apply our findings to create ICU mortality risk assessment models, they may add relevant biomarker data to refine them. The AUCs of our adjusted models were quite modest, the largest of which was 0.82 (Table 2). Perhaps the addition of newer organ failure biomarkers with proven prognostic value, such as neutrophil gelatinase-associated lipocalin (NGAL), [16, 17] interleukin-1 (IL-2) and angiopoietin-2 [18] could further improve the predictive accuracy of our models and make them useful as a risk prediction tool.

The diminishing prognostic returns we found with early organ failure data has been established in prior studies, but our study is the first to use data as it is collected which could eventually be used in real-time for ICU mortality prediction. Others have used summary SOFA measures that require knowledge of data from the entire ICU stay. [2] Levy and colleagues [4] showed that changes after ICU day 1 had little impact on 28-day mortality, but they did not use raw SOFA values. While death is likely determined by an interaction between acute illness severity, age, comorbidities and appropriate care, early therapeutic response seems to influence likelihood of ICU death in our sample.

Our post hoc analysis suggests prognostic improvement with aggregate SOFA scores from consecutive days when there was no improvement with any single score. It extended the utility of ICU organ failure data to day 6 in the validation cohort (Appendix Figure 1, subcohorts 5 & 6). This is expected since those who stay longer in the ICU are more likely to have subtle day-to-day changes in organ failure as a response to initial therapy (antibiotics, fluid resuscitation, etc.) which would be more clinically meaningful when observed over a longer time.

While the derivation and validation cohort had much in common, there were some potentially important differences. The overall ICU and hospital mortality in the derivation group was lower than that of the validation group, even though the former group had higher median SOFA scores. This is likely due to a difference in recruitment periods, with the earliest MIMIC patients recruited in 2001 when sepsis mortality was higher. [19] There was also a difference in the distribution of SOFA score utility. In the derivation cohort, the most recent SOFA score (up to day 4) provided the most useful information, but all scores (up to ICU day 5) in the validation cohort seemed to have equal importance. (Table 2) These patterns were not explained by differences in mortality, total attrition rates or attrition deaths

(Figure 3, Appendix Figure 2), hospital or ICU lengths of stay (data not shown). Despite these differences, the general outcome was the same: indefinite serial organ failure data collection does not lead to consistent improvement in ICU mortality prediction.

There are limitations to our study. First, we did not have access to relevant clinical and laboratory information because of absent data (e.g. serum lactate, medication and antibiotic administration) or recording inaccuracies (e.g. amount of fluids given). We were only able to obtain laboratory values used to calculate the SOFA score. However, it is unlikely that uncollected laboratory data would have been enough of a confounder to erase the incremental benefit of serial SOFA data given the established relationship between SOFA and mortality. [1, 2] Second, this was a retrospective sample in which we assumed clinical suspicion of infection based on ordered blood cultures. If a clinician orders a blood culture, that action indirectly confers a suspicion of infection. This surrogate clinical measure has been used in other relevant prospective sepsis studies. [20] Third, we performed multiple comparisons, which could have biased the results with too many false model improvements from further organ failure data. However, we used Bonferroni adjustment to deal with this, one of the more conservative statistical methods available. [21] Fourth, the IDI doesn't tell us whether any improvement seen in model performance is from better sensitivity or specificity. This could be answered in future studies. Finally, the median length of stay was 3 days, with only about 36% of patients in the ICU past day 5 (the point of minimal return for further organ failure data) in the validation cohort. However, this group comprised over 2000 patients, which should be large enough to detect value of serial SOFA data past ICU day 5 if it existed.

Further study is needed to validate and expand upon our findings. The goal of our study was not to derive a prediction model, but to demonstrate the point at which predictive accuracy no longer improves with further organ failure data. Any variables used in our models were used to remove confounding effects on prognostic improvements with serial addition of SOFA scores. Ultimately, all pertinent clinical and laboratory data should be calibrated specifically for use in a predictive tool. Other novel physiologic and laboratory biomarkers can be incorporated into this predictive algorithm, preferably using a prospectively collected cohort of patients.

CONCLUSIONS

Our data suggests that the incremental value of daily organ failure data for predicting ICU mortality is lost after ICU day 5. Perhaps this point of minimal return is extended to ICU day 6 if aggregate SOFA data from consecutive days are added to existing ICU mortality prediction models. The addition of early daily SOFA data always improved prediction of hospital mortality prediction during our data collection period (from ICU admission to ICU day 7). These findings are useful for clinical decision-making, hospital administration, and research. Future testing is required to derive high-performing prediction models for clinical use.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

sources of funding: Dr. Holder received support from CR Bard. Dr. Kempker receives support from the National Center for Advancing Translational Sciences of the National Institutes of Health (UL1TR000454 and KL2 TR000455). Dr. Nemati is supported by the National Institute of Environmental Health Sciences of the NIH (K01 ES025445). Dr. Martin received funding from Bard, Grifols (Medical Advisory Boards) and received support for article research from the NIH. His institution received funding from the NIH, Food and Drug Administration (FDA), and Baxter Healthcare. He is supported by the National Institute for General Medical Sciences (R01 GM113228), and the National Center for Advancing Translational Sciences of the NIH (UL1 TR000454). Dr. Buchman is an investigator in the Surgical Critical Care Initiative (<http://www.SC2i.org>), a federally funded activity administered through the Henry M. Jackson Foundation for the Advancement of Military Medicine (<http://www.hjfm.org>). He is also the Editor-in-Chief of *Critical Care Medicine*.

Copyright form disclosure: Dr. Holder received funding from Bard Medical. Dr. Kempker's institution received funding from National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH) under Award Number UL1TR000454 and KL2 TR000455, and he received funding from the NIH. Dr. Nemati is supported by the National Institute of Environmental Health Sciences of the NIH (K01 ES025445). Dr. Martin's institution received funding from the NIH (NCATS UL1 TR000454), the Food and Drug Administration, and Baxter Healthcare; he received funding from Bard Medical, Edwards, Grifols, and the Society of Critical Care Medicine (SCCM) (Project Help); and he received support for article research from the NI. Dr. Buchman's institution received funding from Henry M Jackson Foundation supporting the Surgical Critical Care Initiative (www.sc2i.org), SCCM (services as Editor-in-Chief of Critical Care Medicine), and the Coulter Foundation (grant support). He received funding from Philips Corporation paid to a physician organization in the Republic of Korea as an unrestricted educational grant (this defrayed travel costs for him to present findings of eICU work).

References

1. Vincent JL, 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*. 1998; 26(11):1793–800. [PubMed: 9824069]
2. Minne L, Abu-Hanna A, de Jonge E. Evaluation of SOFA-based models for predicting mortality in the ICU: A systematic review. *Crit Care*. 2008; 12(6):R161. [PubMed: 19091120]
3. Sakr Y, et al. Patterns and early evolution of organ failure in the intensive care unit and their relation to outcome. *Crit Care*. 2012; 16(6):R222. [PubMed: 23158219]
4. Levy MM, et al. Early changes in organ function predict eventual survival in severe sepsis. *Crit Care Med*. 2005; 33(10):2194–201. [PubMed: 16215369]
5. Bone RC, et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest*. 1992; 101(6):1644–55. [PubMed: 1303622]
6. Vincent JL, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med*. 1996; 22(7):707–10. [PubMed: 8844239]
7. Charlson ME, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987; 40(5):373–83. [PubMed: 3558716]
8. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol*. 1992; 45(6):613–9. [PubMed: 1607900]
9. Zou H, Hastie T, Tibshirani R. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*. 2006; 15(2):265–286.
10. Jones AE, Trzeciak S, Kline JA. 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*. 2009; 37(5):1649–54. [PubMed: 19325482]

11. Moreno R, et al. The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study. Working Group on Sepsis related Problems of the ESICM. *Intensive Care Med.* 1999; 25(7):686–96. [PubMed: 10470572]
12. Mayaud L, et al. Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension. *Crit Care Med.* 2013; 41(4):954–62. [PubMed: 23385106]
13. Pencina MJ, et al. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med.* 2008; 27(2):157–72. discussion 207–12. [PubMed: 17569110]
14. Torio, CM., Andrews, RM. National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011: Statistical Brief #160, in Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Agency for Health Care Policy and Research (US); Rockville, MD: 2011.
15. Cabre L, et al. Multicenter study of the multiple organ dysfunction syndrome in intensive care units: the usefulness of Sequential Organ Failure Assessment scores in decision making. *Intensive Care Med.* 2005; 31(7):927–33. [PubMed: 15856171]
16. Zhang A, et al. Diagnosis and prognosis of neutrophil gelatinase-associated lipocalin for acute kidney injury with sepsis: a systematic review and meta-analysis. *Crit Care.* 2016; 20:41. [PubMed: 26880194]
17. Shapiro NI, et al. A prospective, multicenter derivation of a biomarker panel to assess risk of organ dysfunction, shock, and death in emergency department patients with suspected sepsis. *Crit Care Med.* 2009; 37(1):96–104. [PubMed: 19050610]
18. Terpstra ML, et al. Plasma biomarkers for acute respiratory distress syndrome: a systematic review and meta-analysis*. *Crit Care Med.* 2014; 42(3):691–700. [PubMed: 24158164]
19. Martin GS, et al. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med.* 2003; 348(16):1546–54. [PubMed: 12700374]
20. Shapiro N, et al. The association of sepsis syndrome and organ dysfunction with mortality in emergency department patients with suspected infection. *Ann Emerg Med.* 2006; 48(5):583–90. 590 e1. [PubMed: 17052559]
21. Narum SR. Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics.* 2006; 7(5):783–787.

Figure 1. Subcohort schematic

Each subsequent subcohort is a subsample of prior subcohorts. Model analysis was performed by subcohort to statistically compare the incremental benefit of daily serial SOFA data. As SOFA data from serial days were added to multivariate logistic regression models, the sample size stayed constant since days were added for each subcohort only up to the point of patient attrition. Note that patient attrition occurs at different rates depending on the subcohort. (Not drawn to scale.)

Figure 2. Patient selection (derivation cohort)

ICU=intensive care unit; ED=emergency department; L&D=labor and delivery; LTAC=long-term acute care facility

a-Defined as 2 of 4 systemic inflammatory response syndrome (SIRS) criteria with 4 hours + blood cultures drawn, and all within 24 hours of each other.

b-Patients came from the following areas: interventional radiology lab, interventional pulmonary lab, cardiac catheterization lab, or from the operating room for a scheduled case.

c-Patients came from the following areas: direct admission from registration office, outpatient clinic, echocardiographic lab, or electrophysiology lab.

d-Used as a surrogate for ward transfers to the ICU, since specific unit locations prior to ICU admission were unavailable.

Figure 3. Effect of incrementally adding daily SOFA scores on ICU mortality prediction
 Demonstrates the effect of adding a single SOFA value to a model for mortality prediction. Right panel shows effect of adding one or more SOFA scores (post hoc analysis). All IDIs that were not statistically significant ($p < 0.05$ in derivation; $p < 0.001$ in validation) were assigned a value of 0 for simplification. The two figures demonstrate the effects of additional SOFA on each subcohort, represented by the six plots on each figure. The sample size, ICU mortality, number of patients lost and percentage of those who died on any given day is

listed in a table under each figure. If present, a circle marks the last day after which additional SOFA data no longer contributes to ICU mortality.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Patient characteristics by ICU mortality

	Derivation cohort					Validation cohort			
Variables	Total (n=1290)	Survivors (n=1207)	Non- survivors (n=83) ^a	p- value	Total (n=8441)	Survivors (n=7492)	Non- survivors (n=949) ^a	p- value	
Demographics									
Age (years), median (IQR)	62 (51,73)	62 (51,73)	64 (53,74)	0.30	65 (51,79)	65 (51,79)	74 (60,83)	<0.01	
Female, n (%)	642 (49.8)	595 (49.3)	47 (56.6)	0.20	4027 (47.7)	3575 (47.7)	452 (47.6)	0.96	
Race/ethnicity, n (%)				0.93				<0.01	
Caucasian	407 (31.5)	381 (31.6)	26 (31.3)		6204 (73.5)	5495 (73.3)	709 (74.7)		
African-American	826 (64.0)	772 (64.0)	54 (65.1)		1186 (14.0)	1097 (14.6)	89 (9.4)		
Other	57 (4.4)	54 (4.5)	3 (3.6)		1051 (12.4)	900 (12.0)	151 (15.9)		
CCI, median (IQR)	3 (2,5)	3 (2,5)	4 (2,6)	0.34	2 (1,4)	2 (1,4)	2 (1,4)	<0.01	
ICU type, n (%)				0.67				0.06	
Medical ICU	696 (53.9)	650 (53.8)	46 (55.4)		5595 (66.3)	4994 (66.7)	601 (63.3)		
Surgical ICU	232 (18.0)	215 (17.8)	17 (20.5)		1102 (13.1)	977 (13.0)	125 (13.2)		
Mixed medical/surgical ICU	99 (7.7)	97 (8.0)	2 (2.4)		-	-	-		
Coronary care unit	111 (8.6)	103 (8.5)	8 (9.6)		985 (11.7)	850 (11.3)	135 (14.2)		
Neurovascular ICU	60 (4.6)	57 (4.7)	3 (3.6)		-	-	-		
Cardiothoracic ICU	53 (4.1)	49 (4.1)	4 (4.8)		759 (9.0)	671 (9.0)	88 (9.3)		
Multiple	39 (3.0)	36 (3.0)	3 (3.6)		-	-	-		
SOFA, median (IQR)									
ICU Admission	6 (4,9)	6 (4,9)	11 (7,13)	<0.01	4 (2,6)	4 (2,6)	6 (4,9)	<0.01	
+1 day (ICU Day 2)	5 (3,9)	5 (3,9)	12 (8,15)	<0.01	4 (2,6)	3 (2,5)	7 (4,10)	<0.01	
+2 days (ICU Day 3)	5 (3,9)	5 (3,8)	11 (7,14)	<0.01	3 (2,6)	3 (2,5)	6 (4,9)	<0.01	
+3 days (ICU Day 4)	6 (3,9)	5 (3,8)	10 (8,14)	<0.01	4 (2,6)	3 (1,5)	6 (4,9)	<0.01	
+4 days (ICU Day 5)	6 (3,9)	6 (3,8)	10.5 (6.5,13.5)	<0.01	4 (2,6)	3 (2,5)	6 (4,9)	<0.01	
+5 days (ICU Day 6)	6 (3,9)	6 (3,8)	8.5 (6,15)	<0.01	4 (2,6)	3 (2,5)	6 (3,9)	<0.01	
+6 days (ICU Day 7)	6 (4,9)	6 (3,9)	9.5 (7,12)	<0.01	4 (2,6)	3 (1,6)	6 (4,9)	<0.01	

Variables	Derivation cohort			p-value	Validation cohort			p-value
	Total (n=1290)	Survivors (n=1207)	Non-survivors (n=83) ^a		Total (n=8441)	Survivors (n=7492)	Non-survivors (n=949) ^a	
Maximum ICU SOFA	8 (5.11)	7 (5.11)	14 (12.17)	<0.01	5 (3.8)	5 (3.7)	9 (7.12)	<0.01
Mean ICU SOFA	5.2 (3.3,7.5)	5 (3.7)	11 (8.14)	<0.01	3.5 (2.5,3)	3.2 (2.4,7)	6.7 (4.7,9)	<0.01
<i>ICU/Hospital course</i>								
Systemic inflammatory parameters at admission								
Temperature >38°C or <36°C, n (%)	813 (63.0)	752 (62.3)	61 (73.5)	0.04	6146 (72.8)	5335 (71.2)	811 (85.6)	<0.01
WBC > 12K or <4K, n (%)	740 (57.4)	679 (56.3)	61 (73.5)	<0.01	5322 (63.7)	4619 (62.2)	703 (75.5)	<0.01
Maximum heart rate, median (IQR)	116 (99,132)	115 (99,132)	119 (98,136)	0.30	109 (96,123)	108 (96,121)	117 (101,132)	<0.01
Maximum respiratory rate, median (IQR)	28 (24,34)	27 (24,34)	28 (24,34)	0.27	28 (24,32)	27 (24,32)	30 (26,36)	<0.01
Length of stay, median (IQR)								
Ventilator	0 (0,3)	0 (0,3)	2 (1,8)	<0.01	0 (0,1)	0 (0,1)	1 (0,5)	<0.01
ICU	3 (2,7)	3 (2,6)	5 (3,9)	0.01	3 (2,6)	3 (2,6)	5 (3,6)	<0.01
Hospital	9 (6,17)	10 (6,17)	5 (3,10)	<0.01	7 (4,12)	7 (4,12)	3 (1,8)	<0.01

^a -Includes only those who died after ICU day 2 (1 day after ICU admission)

ICU=intensive care unit; CCI=Charlson-Deyo comorbidity index; IQR=interquartile range (25th & 75th percentiles); SOFA=sequential organ-failure assessment score; WBC=white blood cell count

Table 2

Effect of serial SOFA data on ICU mortality prediction using IDI^a

	ICU adm	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Derivation cohort ^b							
Subcohort 1 (n=1290)	(0.77)	4.76 (0.82)					
Subcohort 2 (n=898)	(0.74)	1.80 (0.76)	2.76 (0.79)				
Subcohort 3 (n=641)	(0.73)	-0.17 (0.73)	2.09 (0.75)	3.34 (0.79)			
Subcohort 4 (n=506)	(0.72)	-0.68 (0.72)	1.33 (0.73)	4.38 (0.78)	-1.34 (0.77)		
Subcohort 5 (n=406)	(0.73)	-1.23 (0.71)	2.43 (0.74)	<u>4.01</u> (0.79)	-0.77 (0.79)	2.72 (0.81)	
Subcohort 6 (n=331)	(0.67)	-3.67 (0.61)	4.33 (0.68)	<u>6.20</u> (0.74)	-0.18 (0.74)	0.84 (0.75)	3.96 (0.79)
Validation cohort ^b							
Subcohort 1 (n=8441)	(0.69)	5.38 (0.77)					
Subcohort 2 (n=5974)	(0.67)	4.69 (0.73)	2.30 (0.76)				
Subcohort 3 (n=4172)	(0.64)	3.58 (0.69)	2.70 (0.72)	2.32 (0.75)			
Subcohort 4 (n=3040)	(0.64)	2.29 (0.67)	1.80 (0.69)	2.95 (0.73)	1.99 (0.75)		
Subcohort 5 (n=2331)	(0.64)	1.58 (0.66)	<u>1.68</u> (0.68)	1.75 (0.71)	1.50 (0.73)	<u>1.11</u> (0.74)	
Subcohort 6 (n=1841)	(0.64)	1.61 (0.66)	<u>1.56</u> (0.68)	1.21 (0.70)	0.82 (0.71)	<u>1.36</u> (0.73)	2.06 (0.75)

^a -To account for patient attrition, the cohort was split and analyzed into 6 subcohorts" to allow for statistical comparison across days in which the sample size stayed constant. For example, subcohort 1 includes patients who were in the ICU 1 day after ICU admission (ICU day 2), subcohort 2 includes patients who were in the ICU 2 days after ICU admission (ICU day 3), and so on. (Each subsequent cohort is a subsample of those that precede it.) The integrated discriminatory index (IDI=100*[(average new sensitivity-average old sensitivity)+(average new false positive rate-average old false positive rate)]) is provided for each model comparison. Each IDI assesses how incremental addition of the SOFA score from that day (column) effects prognostication. (AUCs are provided in parentheses as a reference.) A bold IDI value represents a statistically significant improvement (p<0.05 for derivation cohort; p<0.001 for validation cohort after Bonferroni correction for multiple comparisons) in model performance by adding the SOFA score for a given day, compared to the model with all SOFA scores immediately preceding (but excluding) that day. An underlined IDI value represents a post hoc assessment of statistically significant improvement (p<0.05 derivation cohort; p<0.001 in validation cohort) between a model and the model with the most recent SOFA score addition that achieved statistical significance (the closest bolded or underlined item to its left). This was to account for any improvement in model performance by addition of more than one day. (Note: the actual values that are underlined still correspond to comparison of adjacent models.) AUCs are not statistically compared.

^b -Adjusted for other variables including age, Charlson comorbidity index, ICU type (medical or medical subspecialty, surgical or surgical subspecialty, or mixed), gender, and the presence of an abnormal white blood cell count at hospital admission. (See Appendix Table 5)

SOFA=sequential organ failure assessment; ICU=intensive care; AUC=area under the receiver operator curve; IDI=integrated discriminatory index; ICU Adm = Day of ICU Admission