

SOFA-2 versus conventional SOFA for mortality prediction in critically ill patients with multiple trauma: a retrospective cohort study of total and organ-specific scores

Hulya Kahraman Sah

drhulyakahraman@gmail.com

Antalya Şehir Hastanesi

Övgü Karataş

Antalya Şehir Hastanesi

Research Article

Keywords: multiple trauma, intensive care unit, Sequential Organ Failure Assessment, SOFA-2, mortality, prognosis, organ dysfunction

Posted Date: September 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10889120/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

The Sequential Organ Failure Assessment (SOFA) score is widely used to quantify acute organ dysfunction, but the original score reflects critical care practice from the 1990s. SOFA-2 was recently developed to incorporate contemporary organ-support strategies and revised thresholds. We compared conventional SOFA (SOFA-1) with SOFA-2 and evaluated their organ-specific components for prediction of intensive care unit (ICU) mortality in patients with multiple trauma.

Methods

This single-center retrospective cohort study included adult trauma patients admitted to the ICU of Antalya City Hospital, Türkiye, between January 2024 and July 2026 who remained in the ICU for at least 48 h. SOFA-1 and SOFA-2 were calculated from the worst values recorded during the first 24 h after ICU admission. The primary outcome was ICU mortality. Discrimination was assessed using receiver operating characteristic curves, and paired areas under the curve (AUCs) were compared using the DeLong method. Multivariable logistic regression models adjusted for age, sex, and traffic-accident mechanism. Bootstrap resampling was used for internal validation.

Results

Among 271 patients, 54 died in the ICU (19.9%). SOFA-1 showed excellent discrimination for ICU mortality (AUC 0.891, 95% CI 0.841–0.934), as did SOFA-2 (AUC 0.892, 95% CI 0.844–0.936). The paired difference was negligible (Δ AUC 0.001; DeLong $p = 0.883$). Exploratory thresholds of ≥ 6 for SOFA-1 and ≥ 5 for SOFA-2 yielded sensitivities of 87.0% and 85.2% and specificities of 77.0% and 79.7%, respectively. In adjusted models, each one-point increase was independently associated with ICU mortality for SOFA-1 (adjusted odds ratio [aOR] 1.69, 95% CI 1.47–1.96) and SOFA-2 (aOR 1.89, 95% CI 1.58–2.26). Neurological scores provided the greatest component-level discrimination; no clinically meaningful improvement in organ-specific discrimination was demonstrated with SOFA-2.

Conclusions

In this multiple-trauma ICU cohort, SOFA-2 preserved the excellent mortality discrimination of conventional SOFA but did not improve total-score AUC. These findings provide additional trauma-specific evidence on the prognostic performance of SOFA-2 while suggesting that its principal advantage in this population may be modernization and standardization of organ-dysfunction assessment rather than a major gain in mortality discrimination.

Background

Multiple trauma can cause rapidly evolving dysfunction across neurological, respiratory, cardiovascular, renal, hepatic, and coagulation systems. Quantifying the burden of organ dysfunction early after ICU admission is therefore clinically and epidemiologically important. The Sequential Organ Failure Assessment (SOFA) score was introduced in 1996 as a standardized six-organ framework for describing organ dysfunction in critically ill patients [1]. Subsequent work demonstrated a strong association between the magnitude and trajectory of SOFA scores and mortality [2].

SOFA has also shown prognostic value in trauma. In a multicenter analysis of 181 trauma patients, non-survivors had higher global SOFA scores and greater early respiratory, cardiovascular, and neurological dysfunction than survivors [3]. In a later single-center study of 241 severe trauma patients, SOFA achieved an AUC of 0.904 for mortality, exceeding the discrimination of APACHE II in that cohort [4]. These observations support the biological relevance of early multisystem dysfunction after major injury, but they were generated using the conventional SOFA framework.

Nearly three decades of changes in critical care have altered the drugs, devices, thresholds, and practices used to support failing organs. SOFA-2 was therefore developed through expert consensus, systematic evidence synthesis, and data-driven validation in more than 3.3 million ICU encounters across nine countries [5, 6]. The updated score retains the six original organ systems and 0–4 grading for each domain while revising variables, thresholds, and organ-support definitions. In international validation, SOFA-2 showed slightly greater overall discrimination for ICU mortality than SOFA-1 (AUROC 0.79 vs 0.77) [6]. Trauma-specific validation remains important because neurological injury, hemorrhagic shock, respiratory failure, and trauma-associated coagulopathy may contribute differently to mortality than in unselected ICU populations.

The primary objective of this study was to compare the ability of SOFA-1 and SOFA-2, calculated from the worst values during the first 24 h of ICU admission, to discriminate ICU mortality in patients with multiple trauma. Secondary objectives were to compare the corresponding organ-system components and to examine whether each score remained independently associated with mortality after adjustment for age, sex, and trauma mechanism.

Methods

Study design and setting

This single-center retrospective cohort study was conducted at Antalya City Hospital, Antalya, Türkiye. Adult trauma patients admitted to the ICU between January 2024 and July 2026 were retrospectively evaluated. Reporting was guided by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [7].

Participants

Patients were eligible if they were aged ≥ 18 years, were newly admitted to the ICU because of trauma, and had an ICU length of stay of at least 48 h. Patients admitted for diagnoses other than trauma and patients with an ICU stay < 48 h were excluded. The analyzed cohort consisted of patients with pedestrian traffic accidents and in-vehicle traffic accidents. The final dataset included 271 patients with complete SOFA-1, SOFA-2, organ-component, and ICU outcome data; therefore, no imputation of missing score-component or outcome data was required.

Data collection and SOFA assessment

Demographic and clinical data included age, sex, trauma mechanism, ICU length of stay, conventional SOFA (SOFA-1), updated SOFA (SOFA-2), and the corresponding organ-specific component scores. Sex was coded as female = 0 and male = 1; trauma mechanism was categorized as pedestrian traffic accident or in-vehicle traffic accident.

For each patient, SOFA-1 and SOFA-2 were calculated using the worst physiological, laboratory, and organ-support values documented during the first 24 h after ICU admission. SOFA-1 was scored according to the conventional six-domain system [1], and SOFA-2 according to the updated definitions and thresholds [5, 6]. The analyzed domains were neurological, respiratory, cardiovascular, liver, kidney, and hemostasis. Component variables represented the number of points contributed by each organ domain to the corresponding total score.

For both SOFA-1 and SOFA-2, the neurological component was based on the last reliable Glasgow Coma Scale (GCS) assessment obtained before initiation of sedation and/or endotracheal intubation. GCS values potentially depressed by sedative medication were therefore not used for neurological scoring. For SOFA-2, the updated neurological scoring rules were applied in accordance with the published SOFA-2 methodology [5, 6].

Outcome

The primary outcome was all-cause ICU mortality, coded as death = 1 and discharged alive = 0. ICU length of stay was described but was not entered as a baseline mortality predictor because it is determined after admission and may be affected by survival time.

Statistical analysis

Continuous variables were summarized as median and interquartile range (IQR), and categorical variables as counts and percentages. Survivor and non-survivor groups were compared using the Mann-Whitney U test for continuous variables and the chi-square test for categorical variables. Discrimination of SOFA-1, SOFA-2, and each organ-system component was quantified by the area under the receiver

operating characteristic curve (AUC) with 95% confidence intervals. Because SOFA-1 and SOFA-2 were measured in the same patients, AUCs were compared using the nonparametric paired method of DeLong et al. [8]. Exploratory cut points were selected by maximizing the Youden index and were not considered externally validated clinical decision thresholds.

Two parsimonious multivariable logistic regression models were fitted separately to avoid collinearity between SOFA-1 and SOFA-2. Each model included one SOFA score together with age, sex, and trauma mechanism. Associations are reported as odds ratios (ORs) with 95% confidence intervals. Model discrimination was quantified by AUC and overall prediction error by the Brier score. Internal validation of the multivariable models was performed using 1,000 bootstrap resamples to estimate optimism-corrected discrimination, consistent with contemporary prediction-model methodology [9]. Given 54 deaths, the number of adjustment variables was intentionally limited to reduce overfitting. Component-level analyses were considered secondary and exploratory. All tests were two-sided, and $p < 0.05$ was considered statistically significant. No formal prospective sample-size calculation was performed because of the retrospective design; model complexity was constrained by the observed number of deaths.

Statistical analyses were conducted using Statistical Package for the Social Sciences (SPSS) version 27.0 (IBM Corporation, Armonk, NY, USA). R software version 4.5.2 was used for bootstrap analysis.

Ethics approval

The study was approved by the Antalya City Hospital Scientific Research Ethics Committee (Decision No. 17/136, dated 27 November 2025). The requirement for individual informed consent was waived because of the retrospective nature of the study. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Results

Cohort characteristics

The study included 271 patients; 54 died in the ICU, corresponding to an ICU mortality of 19.9%. The cohort was predominantly male (218/271, 80.4%), and median age was 39 years (IQR 26–60). Non-survivors were older than survivors (median 53 vs 36 years, $p < 0.001$). Sex distribution and traffic-accident mechanism did not differ significantly between outcome groups. Median ICU length of stay was longer among non-survivors (16 vs 8 days, $p = 0.006$). Both SOFA-1 and SOFA-2 were substantially higher in non-survivors (Table 1).

Table 1

Baseline characteristics according to ICU outcome. IQR, interquartile range; ICU, intensive care unit; SOFA, Sequential Organ Failure Assessment.

Variable	Overall (N = 271)	Survivors (n = 217)	Non-survivors (n = 54)	p value
Age, years, median (IQR)	39 (26–60)	36 (25–57)	53 (35.75–67.75)	< 0.001
Male sex, n (%)	218 (80.4%)	175 (80.6%)	43 (79.6%)	0.866
In-vehicle traffic accident, n (%)	71 (26.2%)	54 (24.9%)	17 (31.5%)	0.324
ICU length of stay, days, median (IQR)	9 (4–18)	8 (4–15)	16 (5–46.75)	0.006
SOFA-1, median (IQR)	4 (2–6)	4 (2–5)	9 (7–12)	< 0.001
SOFA-2, median (IQR)	3 (1–5)	3 (1–4)	7 (5.25–10)	< 0.001

Discrimination and exploratory thresholds

SOFA-1 and SOFA-2 both demonstrated excellent discrimination for ICU mortality. The AUC was 0.891 (95% CI 0.841–0.934) for SOFA-1 and 0.892 (95% CI 0.844–0.936) for SOFA-2. The paired AUC difference was 0.001 and was not statistically significant (DeLong $p = 0.883$). The exploratory Youden threshold for SOFA-1 was ≥ 6 (sensitivity 87.0%, specificity 77.0%, positive predictive value 48.5%, negative predictive value 96.0%). For SOFA-2, the corresponding threshold was ≥ 5 (sensitivity 85.2%, specificity 79.7%, positive predictive value 51.1%, negative predictive value 95.6%).

At the organ-component level, neurological dysfunction provided the strongest discrimination in both scoring systems (SOFA-1 AUC 0.849; SOFA-2 AUC 0.861), followed by the cardiovascular component (0.763 and 0.773, respectively). Renal and respiratory components showed moderate discrimination, whereas liver and hemostasis components showed limited discrimination. No component demonstrated a clinically meaningful improvement with SOFA-2. Although the liver-domain AUC difference reached nominal statistical significance ($p = 0.015$), both AUCs were close to 0.5 and the absolute difference was small; this exploratory result should not be interpreted as clinically important (Table 2).

Table 2
Paired comparison of discrimination for total and organ-specific SOFA-1 and SOFA-2 scores. AUC, area under the receiver operating characteristic curve; GCS, Glasgow Coma Scale.

Domain	SOFA-1 AUC	SOFA-2 AUC	Δ AUC (2 – 1)	DeLong p
Total SOFA	0.891	0.892	0.001	0.883
Neurological (GCS)	0.849	0.861	0.012	0.277
Respiratory	0.685	0.693	0.008	0.481
Cardiovascular	0.763	0.773	0.010	0.321
Liver	0.512	0.526	0.013	0.015
Renal	0.643	0.661	0.018	0.179
Hemostasis	0.601	0.602	0.001	0.960

Multivariable analyses and internal validation

In the model containing SOFA-1, each one-point increase in SOFA-1 was associated with higher odds of ICU mortality after adjustment for age, sex, and trauma mechanism (aOR 1.69, 95% CI 1.47–1.96; $p < 0.001$). Age was also independently associated with mortality, whereas sex and trauma mechanism were not. The model AUC was 0.907, with a bootstrap optimism-corrected AUC of 0.898.

In the corresponding SOFA-2 model, each one-point increase was associated with an aOR of 1.89 for ICU mortality (95% CI 1.58–2.26; $p < 0.001$). The model AUC was 0.910 and the bootstrap optimism-corrected AUC was 0.901. The SOFA-2 model had a lower Brier score (0.086 vs 0.090) and lower Akaike information criterion (165.97 vs 172.21) than the SOFA-1 model; these differences were considered supportive rather than evidence of formal model superiority because the primary paired discrimination comparison was not significant. The multivariable results are summarized in Table 3.

Table 3

Multivariable logistic regression models for ICU mortality. OR, odds ratio; AUC, area under the receiver operating characteristic curve.

Model / variable	Adjusted OR (95% CI)	p value	Model AUC	Bootstrap-corrected AUC
SOFA-1 model			0.907	0.898
SOFA-1, per point	1.69 (1.47–1.96)	< 0.001		
Age, per year	1.032 (1.011–1.054)	0.003		
Male vs female	0.96 (0.36–2.54)	0.928		
In-vehicle vs non-vehicle accident	0.65 (0.25–1.70)	0.384		
SOFA-2 model			0.910	0.901
SOFA-2, per point	1.89 (1.58–2.26)	< 0.001		
Age, per year	1.033 (1.011–1.055)	0.003		
Male vs female	0.78 (0.29–2.08)	0.617		
In-vehicle vs non-vehicle accident	0.75 (0.29–1.93)	0.553		

Discussion

In this retrospective cohort of 271 critically ill patients with multiple trauma, both conventional SOFA and SOFA-2 demonstrated excellent discrimination for ICU mortality. The principal finding was the virtually identical performance of the two total scores, with an AUC of 0.891 for SOFA-1 and 0.892 for SOFA-2 (Δ AUC = 0.0009; DeLong p = 0.883). SOFA-2 nevertheless remained strongly and independently associated with ICU mortality after adjustment for age, sex, and trauma mechanism, and the adjusted model retained good discrimination after bootstrap internal validation. These findings suggest that SOFA-2 preserves the strong prognostic information provided by conventional SOFA in critically ill trauma patients but does not confer a measurable improvement in rank-order discrimination for ICU mortality in this cohort.

Our findings should be interpreted in the context of the original SOFA-2 development and validation study and subsequent external validations. In the international SOFA-2 study, which included more than 3.3 million ICU encounters, SOFA-2 demonstrated modestly higher discrimination for ICU mortality than SOFA-1 (AUROC 0.79 vs 0.77) [6]. Subsequent external validations have generally reported similar or improved discrimination with SOFA-2, but the magnitude of improvement has varied. In MIMIC-IV, SOFA-2

achieved an AUROC of 0.829 compared with 0.796 for SOFA-1 [10]. In the Japanese OneICU database, the difference was much smaller (0.859 vs 0.853), despite statistical significance in a very large cohort [11]. A large Chinese multicenter validation likewise reported higher discrimination with SOFA-2 (0.809 vs 0.767) [12]. Taken together, these findings indicate that the incremental discrimination gained with SOFA-2 is case-mix dependent. The absence of an AUC improvement in our cohort therefore does not contradict the validity of SOFA-2; rather, conventional SOFA already demonstrated high discrimination in this selected trauma population, leaving limited room for further improvement in AUC.

Our SOFA-1 performance is consistent with earlier trauma literature. Antonelli et al. demonstrated that SOFA reliably characterized organ dysfunction and identified trauma patients at increased risk of prolonged ICU stay or death [3]. Lee et al. subsequently reported an AUC of approximately 0.90 for mortality prediction with SOFA in critically ill trauma patients [4], closely resembling the AUC of 0.891 observed in the present cohort. Other trauma studies have likewise linked SOFA-defined multiple organ dysfunction with mortality and adverse outcomes [13–15]. A simplified SOFA approach has also been associated with 30-day trauma mortality in a resource-limited setting [16], while qSOFA studies using trauma registries support the prognostic relevance of early physiologic and neurologic derangement after injury [17, 18]. Collectively, these findings support the concept that organ dysfunction during the early post-injury period contains substantial prognostic information in severely injured patients.

Very recent trauma-specific data provide further context. In a four-center cohort of 2,162 severely injured patients, application of SOFA-2 reduced the proportion classified as having multiple organ dysfunction syndrome compared with SOFA-1 (61.6% vs 68.5%) and produced lower first-day total scores, predominantly because of changes in respiratory and cardiovascular scoring [19]. Higher SOFA-2 respiratory, cardiovascular, and renal dysfunction remained associated with critical care mortality. That study primarily addressed organ-dysfunction classification and was available as a preprint at the time of preparation of the present manuscript [19]. Our study addresses a complementary question by directly comparing paired mortality discrimination of the two total scores and their organ-specific components in a contemporary trauma ICU cohort.

The absence of a significant increase in AUC should not be interpreted as evidence that the SOFA-2 update lacks clinical value. SOFA was originally developed primarily as a standardized description of organ dysfunction rather than as a stand-alone mortality prediction model [1, 2], whereas SOFA-2 was designed to modernize this framework by incorporating contemporary organ-support practices, revised thresholds, alternative variables, and more explicit scoring rules [5, 6]. Improvements in measurement validity and contemporary clinical relevance do not necessarily translate into large increases in AUC, particularly when baseline discrimination is already high. Consistent with this interpretation, the SOFA-2 adjusted model in our cohort showed slightly lower prediction error and AIC than the corresponding SOFA-1 model, although these secondary findings should be considered supportive rather than evidence of statistical superiority.

The organ-specific analyses provide additional insight into the performance of the two scores in trauma. Neurological dysfunction showed the highest individual discrimination for ICU mortality in both systems (AUC 0.849 for GCS-1 and 0.861 for GCS-2), followed by cardiovascular dysfunction. This pattern is biologically plausible in a trauma population, in which traumatic brain injury, impaired consciousness, hemorrhagic shock, and circulatory failure are major determinants of early outcome. Previous trauma studies have demonstrated strong prognostic performance of GCS and other neurologic or physiologic scoring approaches [20, 21]. Respiratory, cardiovascular, renal, and hemostatic components showed no statistically significant differences between the two SOFA systems. Although the liver-domain comparison reached nominal statistical significance, both AUCs were close to 0.50 and the absolute difference was minimal; therefore, this finding is unlikely to represent clinically meaningful discrimination and may reflect multiple testing. Organ-component analyses should consequently be regarded as exploratory.

The exploratory cutoffs identified in this study should also be interpreted cautiously. SOFA-1 ≥ 6 and SOFA-2 ≥ 5 yielded negative predictive values exceeding 95%, suggesting that lower scores may identify a group at relatively low risk of ICU mortality. However, predictive values are prevalence-dependent, and these thresholds were derived and evaluated in the same cohort. They should therefore not be interpreted as validated triage or treatment thresholds [9]. Moreover, discrimination alone does not establish the reliability of individual predicted risks; calibration, external validation, and assessment of clinical utility remain essential before a prognostic score or threshold is used to support individual decision-making [22, 23].

Several features strengthen the present study. SOFA-1 and SOFA-2 were calculated in the same patients over the same first-24-h observation window, allowing a direct paired comparison without between-cohort differences. The analysis included all six organ-specific components, thereby examining not only overall discrimination but also the domains potentially responsible for differences between scoring systems. Multivariable models were deliberately parsimonious given the number of mortality events, and bootstrap internal validation demonstrated only modest optimism in model discrimination. This approach is consistent with contemporary recommendations that prediction-model performance should be assessed beyond apparent in-sample discrimination and that internal validation be used to quantify optimism [9, 22, 23].

Several limitations should nevertheless be emphasized. First, this was a retrospective single-center study, and the findings may not generalize to trauma systems with different case mix, referral patterns, or critical care practices. Second, eligibility required an ICU stay of at least 48 h. Patients who died or were discharged before 48 h were therefore excluded, introducing potential selection and survivorship bias. In particular, exclusion of patients with rapidly fatal injuries may have altered the observed mortality rate and the distribution and prognostic performance of neurological, cardiovascular, and other organ dysfunction components. Contemporary trauma data indicate that clinically important organ dysfunction and associated deaths may occur early after injury [15]. The present findings should consequently be interpreted as applying to trauma patients who survived and remained in intensive care for at least 48 h

rather than to all patients admitted after major trauma. Third, Injury Severity Score (ISS) and Abbreviated Injury Scale (AIS) data were unavailable, preventing adjustment for anatomical injury severity and limiting comparison with established trauma-specific prognostic approaches. This is important because ISS remains a foundational anatomical severity metric [24], and AIS-based models demonstrate the prognostic contribution of detailed anatomical injury information [25, 26]. Fourth, the cohort was restricted to the traffic-related trauma mechanisms represented in our database, which may further limit generalizability. Fifth, residual confounding from traumatic brain injury severity, comorbidity, transfusion burden, operative interventions, and treatment-limitation decisions cannot be excluded. Sixth, scores were based on the worst measurements recorded during the first 24 h after ICU admission; therefore, the study evaluates early ICU prognostication rather than prediction available immediately at admission. Seventh, although pre-sedation GCS values were used to minimize drug-related distortion of neurological scoring, retrospective documentation may still have introduced measurement error. Finally, the organ-component analyses and cutoffs were exploratory and internally derived, and independent external validation is required.

Conclusions

SOFA-2 demonstrated excellent discrimination for ICU mortality in critically ill patients with multiple trauma and performed almost identically to conventional SOFA at the total-score level. The updated score remained strongly associated with mortality after adjustment for age, sex, and traffic-accident mechanism, but it did not provide a statistically significant improvement in AUC. In trauma care, the principal advantage of SOFA-2 may therefore lie in contemporary and standardized organ-dysfunction assessment while preserving established prognostic performance. Multicenter external validation incorporating anatomical injury severity is warranted.

Abbreviations

AIS: Abbreviated Injury Scale; AUC: area under the receiver operating characteristic curve; CI: confidence interval; GCS: Glasgow Coma Scale; ICU: intensive care unit; IQR: interquartile range; ISS: Injury Severity Score; OR: odds ratio; ROC: receiver operating characteristic; SOFA: Sequential Organ Failure Assessment; STROBE: Strengthening the Reporting of Observational Studies in Epidemiology.

Declarations

Ethics approval and consent to participate

The study was approved by the Antalya City Hospital Scientific Research Ethics Committee (Decision No. 17/136, dated 27 November 2025). The requirement for individual informed consent was waived because of the retrospective nature of the study. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request, subject to institutional and ethical requirements.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Authors' contributions

H.K.S. contributed to the study design, data collection, literature review, manuscript drafting, critical revision of the manuscript, and approval of the final version. Ö.K. contributed to data collection and literature review. Both authors read and approved the final manuscript.

Use of artificial intelligence

During preparation of this manuscript, the authors used ChatGPT (OpenAI) as a generative artificial intelligence tool to assist with language editing, readability, and manuscript organization. The AI tool was not used to generate or modify the original study data or to make clinical decisions. All AI-assisted content was critically reviewed, verified, and revised by the authors, who take full responsibility for the accuracy, integrity, and final content of the manuscript.

Acknowledgements

Not applicable.

Clinical trial number

Not applicable.

References

1. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med.* 1996;22(7):707–10. 10.1007/BF01709751.
2. Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA.* 2001;286(14):1754–8. 10.1001/jama.286.14.1754.
3. Antonelli M, Moreno R, Vincent JL, et al. Application of SOFA score to trauma patients. *Sequential Organ Failure Assessment.* *Intensive Care Med.* 1999;25(4):389–94. 10.1007/s001340050863.
4. Lee MA, Choi KK, Yu B, et al. Acute Physiology and Chronic Health Evaluation II Score and Sequential Organ Failure Assessment Score as Predictors for Severe Trauma Patients in the Intensive Care Unit. *Korean J Crit Care Med.* 2017;32(4):340–6. 10.4266/kjccm.2017.00255.
5. Moreno R, Rhodes A, Ranzani O, et al. Rationale and Methodological Approach Underlying the Development of the Sequential Organ Failure Assessment (SOFA)-2 Score: A Consensus Statement. *JAMA Netw Open.* 2025;8(10):e2545040. 10.1001/jamanetworkopen.2025.45040.
6. Ranzani OT, Singer M, Salluh JIF, et al. Development and Validation of the Sequential Organ Failure Assessment (SOFA)-2 Score. *JAMA.* 2025;334(23):2090–103. 10.1001/jama.2025.20516.
7. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet.* 2007;370(9596):1453–7. 10.1016/S0140-6736(07)61602-X.
8. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837–45. 10.2307/2531595.
9. Efthimiou O, Seo M, Chalkou K, Debray T, Egger M, Salanti G. Developing clinical prediction models: a step-by-step guide. *BMJ.* 2024;386:e078276. 10.1136/bmj-2023-078276.
10. Bai S, Huang M, Chen H, Yang X, Chen L, Liu F, et al. SOFA-2 versus SOFA-1 for mortality prediction in critically ill patients. *J Intensive Care.* 2026;14:32. 10.1186/s40560-026-00875-6.
11. Iwasaki Y, Kinoshita T, Yoshimura J, Maruyama S, Ohshimo S, Murao S, et al. Admission-day SOFA-2 versus SOFA for mortality discrimination: external validation in a large multicenter ICU database. *Crit Care.* 2026;30:260. 10.1186/s13054-026-06020-x.
12. Liufu R, Cao M, Shi XQ, Fu XY, Shen F, Chen Y, et al. External validation of SOFA 2 score: a retrospective Chinese cohort study. *Ann Intensive Care.* 2026;16:100107. 10.1016/j.aicoj.2026.100107.
13. Ulvik A, Kvåle R, Wentzel-Larsen T, Flaatten H. Multiple organ failure after trauma affects even long-term survival and functional status. *Crit Care.* 2007;11(5):R95. 10.1186/cc6111.
14. Hutchings L, Watkinson P, Young JD, Willett K. Defining multiple organ failure after major trauma: a comparison of the Denver, Sequential Organ Failure Assessment, and Marshall scoring systems. *J Trauma Acute Care Surg.* 2017;82(3):534–41. 10.1097/TA.0000000000001328.

15. Cole E, Gillespie S, Vulliamy P, Brohi K. Organ Dysfunction in Trauma (ORDIT) study collaborators. Multiple organ dysfunction after trauma. *Br J Surg*. 2020;107(4):402–12. 10.1002/bjs.11361.
16. Rees CA, Xiao M, Wild J, et al. A simplified Sequential Organ Failure Assessment score is associated with 30-day trauma mortality: a novel tool for a low-resource practice setting. *J Trauma Acute Care Surg*. 2026;100(3):e22–6. 10.1097/TA.0000000000004804.
17. Miyamoto K, Shibata N, Ogawa A, Nakashima T, Kato S. Prehospital quick sequential organ failure assessment score to predict in-hospital mortality among patients with trauma. *Am J Emerg Med*. 2019;37(12):2165–70. 10.1016/j.ajem.2019.03.007.
18. Miyamoto K, Shibata N, Ogawa A, Nakashima T, Kato S. Prehospital and in-hospital quick Sequential Organ Failure Assessment (qSOFA) scores to predict in-hospital mortality among trauma patients: an analysis of nationwide registry data. *Acute Med Surg*. 2020;7(1):e532. 10.1002/ams2.532.
19. Krishna A, Rossetto A, Brohi K, Vulliamy P, Cole E. SOFA-2 reclassifies multiorgan dysfunction syndrome in major trauma patients. *medRxiv [Preprint]*. 2026. 10.64898/2026.08.05.26359771. Posted August 7, 2026.
20. Ghelichkhani P, Esmaeili M, Hosseini M, Seylani K. Glasgow Coma Scale and FOUR Score in predicting the mortality of trauma patients: a diagnostic accuracy study. *Emerg (Tehran)*. 2018;6(1):e42.
21. Khari S, Zandi M, Yousefifard M. Glasgow Coma Scale versus physiologic scoring systems in predicting the outcome of ICU admitted trauma patients: a diagnostic accuracy study. *Arch Acad Emerg Med*. 2022;10(1):e25. 10.22037/aaem.v10i1.1483.
22. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17(1):230. 10.1186/s12916-019-1466-7.
23. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55–63. 10.7326/M14-0697.
24. Baker SP, O'Neill B, Haddon W Jr, Long WB. The Injury Severity Score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma*. 1974;14(3):187–96.
25. Osler T, Glance L, Buzas JS, Mukamel D, Wagner J, Dick A. A trauma mortality prediction model based on the anatomic injury scale. *Ann Surg*. 2008;247(6):1041–8. 10.1097/SLA.0b013e31816ffb3f.
26. Wang M, Wu D, Qiu W, Wang W, Zeng Y, Shen Y. An injury mortality prediction based on the anatomic injury scale. *Med (Baltim)*. 2017;96(35):e7945. 10.1097/MD.0000000000007945.