Reliability and Accuracy of Sepsis-related Organ Failure Assessment Scoring among Emergency Physicians

Yun-Xia Chen, Yi-Xian Li, Shu-Bin Guo, Xue Mei
Emergency Department, Beijing Chao-Yang Hospital Affiliated to Capital Medical University, Beijing 100020, China

To the Editor: With the growing importance of Sepsis-related Organ Failure Assessment (SOFA), clinicians are increasingly concerned about its reliability and accuracy. In Intensive Care Units, SOFA scoring was found to be insufficiently accurate among both physicians and nurses. Few studies have reported the performance of emergency department (ED) physicians in manually assigning the SOFA score. We therefore conducted a single-center prospective investigation to examine that issue.

The study was conducted in May 2017 in the ED of Beijing Chao-Yang Hospital affiliated to Capital Medical University, Beijing, China. This investigation assessed the performance of ED physicians in SOFA scoring using hypothetical cases after a training program. The study did not involve actual patients, and there were no ethical considerations.

Thirty cases were developed by one of the authors (Yun-Xia Chen). The cases covered the most common diseases encountered in the ED including sepsis, shock, stroke, severe hemorrhage, acute heart failure, acute coronary syndrome, and acute pancreatitis. The mean age of these cases was 56 ± 17 years and male was 73%. The mean score values of SOFA and Acute Physiology and Chronic Health Evaluation (APACHE) II score were 9 ± 3 and 22 ± 9, respectively.

Error rates were calculated for the raw data value, scores of all the 12 variables, scores for the six organ systems, and total SOFA scores. We estimated the reliability of the total SOFA score by calculating the intraclass correlation coefficient (ICC) and the 95% confidence interval (CI) based on a single-measure, one-way random-effects model. To evaluate the accuracy of the total SOFA score, agreement between participants’ scores and the gold standard was assessed using Bland–Altman plots. It is recommended that 95% of data points should lie within ±1.96 SD of the mean differences.

In all, 40 ED physicians participated in the training and investigation including 18 residents (45%) and 22 specialists (55%). The mean age of participants was 37 ± 6 years, the mean working years was 11 ± 8 years, and male was 77.5%. Of the participants, 80% had experience with APACHE II scoring and 37.5% with SOFA scoring.

The ICC of the total SOFA score was 0.873 (95% CI, 0.805–0.928). The accuracy of the raw data value was moderate (72–84%) for PaO2/FiO2, respiratory support, Glasgow Coma Scale (GCS) score, and urine output (UO); it was excellent (>90%) for all other variables. The accuracy in assigning scores for the individual variables was good (80–82%) for the GCS score and UO and excellent (>94%) for the others. The accuracy of scores for the individual organ systems was good (80–88%) for respiration, central nervous system, and renal systems and excellent (>90%) for coagulation, liver, and cardiovascular systems. The accuracy of the total SOFA score was 54%, and there was no difference between residents and specialists (53% vs. 55%, P = 0.785).

Figure 1 shows the Bland–Altman plots of the total SOFA scores for the gold standard and participants. The mean bias between the participants’ and the gold standard total SOFA scores was 0.480 ± 1.899 (95% CI, −3.2 to 4.2). Of all the total SOFA scores, 243 (97%) were within the 95% CI of the mean difference [Figure 1a and 1b]. Bias occurred in both the low and high values of the total SOFA score [Figure 1b].

In total, 258 errors were found. The first three most frequent errors were not using the worst value for 24 h, incorrect GCS score, and extrapolating UO per 24 h. Infrequent errors included calculating errors, using incorrect variables, failure to consider all variables in a single-organ system, and clerical errors. There were 51 (21%) major errors with the total SOFA score.

Address for correspondence: Prof. Shu-Bin Guo, Emergency Department, Beijing Chao-Yang Hospital Affiliated to Capital Medical University, Beijing 100020, China E-Mail: guoshubinyy@163.com

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms. © 2018 Chinese Medical Journal ǀ Produced by Wolters Kluwer - Medknow

Received: 29-09-2017 Edited by: Qiang Shi
How to cite this article: Chen YX, Li YX, Guo SB, Mei X. Reliability and Accuracy of Sepsis-related Organ Failure Assessment Scoring among Emergency Physicians. Chin Med J 2018;131:247-8.
This investigation found that for the total SOFA score, reliability was good but the accuracy was not ideal. Human factors are the main influence on accuracy in SOFA scoring. Several studies also found good reliability, with the ICC ranging from 0.870 to 0.889, and poor accuracy (48–53%) with the total SOFA score.\[2,3,5\]

In our investigation, the ICC of the total SOFA was 0.873, and 97% of the differences were within the 95% CI of mean difference. Statistically, this result indicates good reliability of the total SOFA score among our participants. However, we also observed that the 95% CI was relatively broad, ranging from −3.2 to 4.2. Together with the low accuracy of the total SOFA score (54%), we considered that these variations would not significantly impact the prognostic value of SOFA, but they would essentially decrease its accuracy in diagnosing sepsis. Since the SOFA score acutely increasing by 2 or more points is defined as sepsis, variations in total SOFA score between raters should be less than that range. In our study, the major error rate in the total SOFA score was 21%; that includes scores both greater and less than the gold standard. If that situation occurred in clinical practice, overdiagnosis and misdiagnosis of sepsis would be the consequence, and that could result in unfavorable outcomes. As a diagnostic criterion, SOFA scoring should be highly precise. Systematic, rigorous, repeated training is the conventional means of improving the accuracy of SOFA scoring. Automated SOFA is attractive because of advantages in time saving, reliability, accuracy, and frequency.

**Financial support and sponsorship**

Nil.

**Conflicts of interest**

There are no conflicts of interest.

**REFERENCES**

1. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, 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;22:707-10. doi: 10.1007/s001340050156.
2. Arts DG, de Keizer NF, Vroom MB, de Jonge E. Reliability and accuracy of Sequential Organ Failure Assessment (SOFA) scoring. Crit Care Med 2005;33:1988-93. doi: 10.1097/01.CCM.0000178178.02574.AB.
3. Baykara N, Gökduman K, Hoşten T, Solak M, Toker K. Comparison of Sequential Organ Failure Assessment (SOFA) scoring between nurses and residents. J Anesth 2011;25:839-44. doi: 10.1007/s00540-011-1232-2.
4. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155-63. doi: 10.1016/j.jcm.2016.02.012.
5. Tallgren M, Bäcklund M, Hynminen M. Accuracy of Sequential Organ Failure Assessment (SOFA) scoring in clinical practice. Acta Anaesthesiol Scand 2009;53:39-45. doi: 10.1111/j.1399-6576.2008.01825.x.