

Data from clinical database on septic shock are valid

Lars Grønlykke, Sofie Louise Rygaard Brandstrup & Anders Perner

ABSTRACT

INTRODUCTION: Clinical databases have become important tools in intensive care. Disease severity and organ dysfunction scoring systems are registered in the databases, including the Simplified Acute Physiology Score II (SAPS II) and the Sequential Organ Failure Assessment (SOFA) score. The purpose of this study was to evaluate the reliability and accuracy of a clinical database on intensive care unit (ICU) patients.

MATERIAL AND METHODS: Data were extracted from the clinical database, Critical Information System (CIS). We included all adult patients admitted to one of seven Danish ICUs between 1 January 2008 and 31 December 2010 diagnosed with septic shock. Validation of the diagnosis of septic shock and SAPS II and SOFA scores were obtained on every tenth patient by comparing data entries in CIS with the source data stored in the unit.

RESULTS: A total of 1,353 patients were identified and data on 142 patients were selected for validation. All but one patient (99%, 95% confidence interval (CI): 95-100) fulfilled the diagnostic criteria for septic shock. We found less than 10% variation in SAPS II in 78% (95% CI: 73-86) and less than 10% variation in SOFA scores in 80% (95% CI: 72-85) of the cases. The average bias between the registered and corrected SAPS II according to the Bland-Altman plot was -1.8 (limits of agreement: -10.1 to 6.6). Furthermore, the average bias between the registered and corrected SOFA score according to the Bland-Altman plot was -0.2 (limits of agreement: -2.4 to 2.0).

CONCLUSION: The accuracy of the diagnosis of septic shock was high and both SAPS II and SOFA scores were reliable and accurately recorded in the ICU database.

FUNDING: not relevant.

TRIAL REGISTRATION: not relevant.

Clinical databases have become important tools in intensive care units (ICUs) where data on patients are registered and used for clinical research and quality improvement.

Disease severity scoring systems are used to compare case mix over time and between units and as baseline variables in clinical and research databases [1, 2]. The scoring systems are increasingly used to calculate the standardized mortality ratio (SMR), which is the ratio of observed mortality and the mortality predicted from severity scoring, with a view to assessing the quality of care and overall performance of ICUs [3].

A widely used disease severity scoring system is the Simplified Acute Physiology Score II (SAPS II), which calculates a score based on two patient characteristics, three disease variables and 12 clinical variables obtained within the first 24 hours of ICU admittance. From SAPS II, the risk of in-hospital mortality can be estimated for cohorts of ICU patients, but SAPS II cannot be used to predict the mortality in individual patients [4, 5].

Another scoring system often used in clinical databases in ICUs is organ dysfunction scoring such as the Sequential Organ Failure Assessment (SOFA) score, which has proven to be a simple, but useful tool for the description of organ failure in ICU patients [6, 7].

Furthermore, patients' diagnoses may be registered in the databases to enable clinicians and researchers to extract and review data of specific groups of patients.

It is essential to know the quality of the data to make proper use of the disease severity and organ dysfunction scoring systems registered in the database for clinical research and quality-improvement.

The aim of our study was therefore to evaluate the quality of a clinical ICU database, specifically the registration of the septic shock diagnosis and of the SAPS II and the SOFA scoring.

MATERIAL AND METHODS

This study was conducted in seven general ICUs in Denmark (at Rigshospitalet and the hospitals of Gentofte, Herlev, Hillerød, Kolding, Odense and Vejle) using the administrative diagnoses coding and clinical scoring systems Critical Information System (CIS) by Daintel, Copenhagen, Denmark. CIS is an ICU-specific electronic medical record system that holds the patient's national identification number, admission characteristics and notes, diagnoses, daily organ-specific status and raw data for SAPS II and SOFA scoring. In all seven ICUs, clinical doctors prospectively entered all of the following data into the CIS for all patients admitted to the ICU: primary and secondary diagnoses and raw data for SAPS II and SOFA scoring from patient files, laboratory result notes and observation charts. All data must be registered before a patient can be discharged, which ensures that datasets are complete. CIS data were filed in local databases at each hospital and CIS has a search function allowing system administrators to search and extract patient specific data in Excel format.

The CIS database at each ICU was searched and

ORIGINAL ARTICLE

Department of
Intensive Care,
Rigshospitalet

Dan Med J
2012;59(10):A4522

TABLE 1

Patients in categories at each intensive care unit (ICU). The values are n (%).

Category	ICU 1	ICU 2	ICU 3	ICU 4	ICU 5	ICU 6	ICU 7	Average
Medical	24 (49)	5 (36)	3 (43)	6 (30)	8 (38)	9 (90)	10 (48)	65 (46)
Elective surgical	1 (2)	0 (0)	1 (14)	0 (0)	2 (10)	0 (0)	0 (0)	4 (3)
Acute surgical	24 (49)	9 (64)	3 (43)	14 (70)	11 (52)	1 (10)	11 (52)	73 (51)

TABLE 2

Patients with accurate data in the four categories in each intensive care unit and total in all intensive care units. The values are n (%).

Shock criteria fulfilled	49 (100)	14 (100)	7 (100)	20 (100)	21 (100)	10 (100)	20(95)	141(99)
SIRS criteria fulfilled	49 (100)	14 (100)	7 (100)	20 (100)	21 (100)	10 (100)	20(95)	141(99)
SAPS II with < 10% variation	43 (88)	13 (93)	6 (86)	14 (70)	11 (52)	7 (70)	17 (81)	111 (78)
SOFA with < 10% variation	39 (80)	9 (64)	7 (100)	17 (85)	19 (91)	8 (80)	14 (67)	114 (80)
GCS accuracy	45 (92)	13 (93)	6 (86)	14 (70)	16 (76)	6 (60)	16 (76)	116 (82)

GCS = Glasgow Coma Scale; SAPS II = Simplified Acute Physiology Score II; SIRS = Systemic Inflammatory Response Syndrome score; SOFA = Sequential Organ Failure Assessment score.

data were extracted by the local administrator following written and verbal instructions from LG and SLRB. In the search, we included all patients aged 18 years or older admitted to one of the seven ICUs between 1 January 2008 and 31 December 2010 having the diagnosis of septic shock as the primary or as a secondary diagnosis in the coding system. One ICU started using CIS in the beginning of 2009, so no patients from 2008 were included in this unit. At each ICU, we listed the included patients by admission date and selected every tenth patient in the full study period. Validation of CIS data entries was performed for these patients by comparing their data with the source data specified by the ICU (patient files, laboratory result notes, observation charts, etc.) This was done by LG and SLRB. If the patient had died or had been discharged within the first 24 hours of admission, no SAPS II or SOFA score was registered for that patient. These patients were excluded from the study population.

We reviewed SAPS II data and first-day SOFA scoring and the septic shock diagnosis, i.e. evidence of infection, two positive systemic inflammatory response syndrome (SIRS) criteria and hypotension (mean arterial pressure < 70 mmHg) or use of vasopressor treatment (infusion of noradrenalin or dopamine) after initial fluid treatment [8].

We also registered the type of admission recorded in the CIS (medical, elective surgical or acute surgical).

Furthermore, we made a specific evaluation of the Glasgow Coma Scale (GCS) scoring recorded in SAPS II

and SOFA. We compared the doctors' CIS description of patients' neurological status upon admission and recorded whether there was a discrepancy between the source data and the GCS-registered data as part of the SAPS II and SOFA scores. We did not include any inaccuracy of GCS scoring in the final evaluation of SAPS II and SOFA accuracy because of expected difficulties in defining the true GCS score from patient files.

Each ICU was invited to provide information on their introduction and training programme for doctors entering data into the CIS.

The study was approved by the Danish National Board of Health and by the Danish Data Protection Agency.

Statistics

Data were given as means (standard deviation, SD) or number (percentages); 95% confidence intervals (CI) were given where appropriate. A priori, we defined an acceptable variation as below 10% difference in numeric scores between the registered and the validated SAPS II and SOFA scores. We analyzed the influence of the observed data errors by calculating the differences between SAPS II and SOFA scores before and after correction by paired Wilcoxon signed rank test. We used Bland-Altman plots to analyze the bias and limits of agreement between the scores before and after correction. All data were analyzed using SPSS statistics v. 20. A p-value < 0.05 was considered statistically significant.

Trial registration: not relevant.

RESULTS

In the three-year study period, 1,353 adult patients were diagnosed with septic shock in the seven ICUs. A total of 168 unique patient records were selected for study. Twenty-six of these patients had either died or were discharged within 24 hours of admission before SAPS II and SOFA scoring could be done and were therefore excluded from the study leaving 142 unique patients' records for validation. There were some differences in patient characteristics between ICUs; one ICU had almost exclusively medical admissions (90% medical and 10% acute surgical), whereas the others had more even distributions between surgical and medical patients (Table 1).

All patients except one were correctly diagnosed with shock (99%), being either hypotensive or treated with vasopressors. Furthermore, all patients except one had a minimum of two positive SIRS criteria (99%). The majority of the patients (83%) had a documented focus of infection, and the remaining (17%) had an unknown focus, but all were treated with antibiotics. Taken together, 99% (95% CI: 95-100) of the patients fulfilled the

diagnostic criteria for septic shock with no variation between units (Table 2).

The mean SAPS II was 54.3 (SD 16.7) in the registered data, which changed to 56.1 (SD 16.5) after the validation against source data ($p < 0.001$). The change in SAPS II after validation corresponds to a change in predicted mortality from 55% to 60%. For the SAPS II scoring, 78% (95% CI: 73-86) of the paired values (registered versus validated) had a variation below the predefined acceptable level of 10%. There was some variation (range 52% to 93%) between the ICUs (Table 2). The most common data entry error was that mechanical ventilation was not accounted for, which led to a falsely low SAPS II score. Another error was discrepancy between the laboratory values registered in SAPS II and those found in the source data, which led to both falsely lower and falsely higher scores.

The average bias between the registered and corrected SAPS II according to the Bland-Altman plot was -1.8 (limits of agreement: -10.1 to 6.6) (Figure 1).

The mean SOFA score in our population was 10.7 (SD 4.1) before validation and 10.9 (SD 4.0) after validation ($p = 0.03$). For SOFA scoring, 80% (95% CI: 72-85) had a variation below the predefined acceptable threshold of 10% with some variation between ICUs (64%-100%) (Table 2). The most common error was not to include doses of vasopressors or inotropes in the calculation of the sub-score for circulatory failure of the SOFA score.

The average bias between the registered and the corrected SOFA score according to the Bland-Altman plot was -0.2 (limits of agreement: -2.4 to 2.0) (Figure 2).

We found an accuracy of GCS registration of 82%. There was a tendency to GCS score patients after sedation, which led to a falsely low GCS score.

All of the ICUs had a CIS introduction programme for new doctors. It consisted of either one or two training sessions with a doctor experienced in using CIS. One ICU had a follow-up session 1-2 days after the initial CIS introduction.

DISCUSSION

We have shown that the data of the clinical database used in seven Danish ICUs have a high accuracy for the diagnosis of septic shock and a reasonable accuracy and reliability for the disease severity and organ failure scores. The doctors entering the data may not have gone through extensive training, but it is possible that the mandatory entry of raw data reduced the error rates; at least there were no missing values.

During the last year of our study period (from 23 December 2009), a large multicentre study on septic shock patients was initiated in six of the seven ICUs in-

FIGURE 1

Brandt-Altman plot of SAPS II scores before and after correction. Bias (mean difference —) and 95% limits of agreement (± 1.96 standard deviation - - -) are shown.

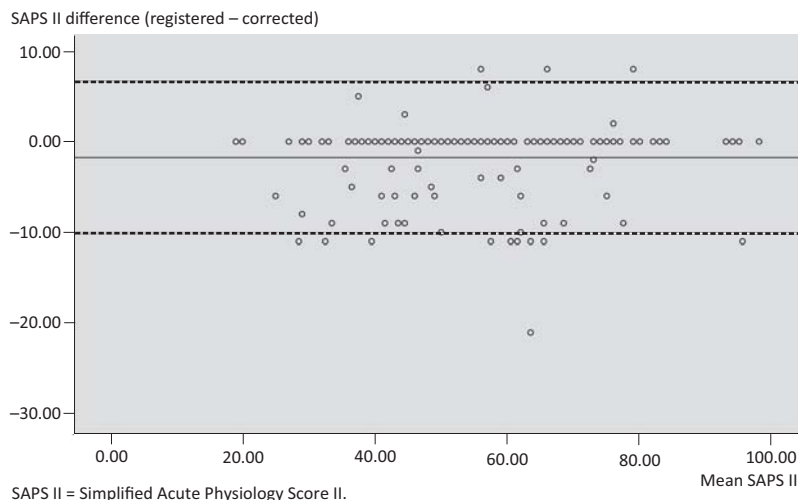
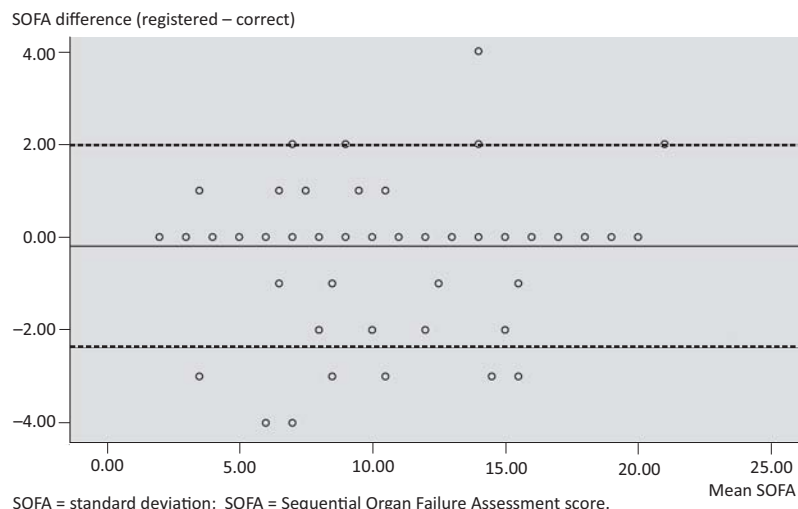


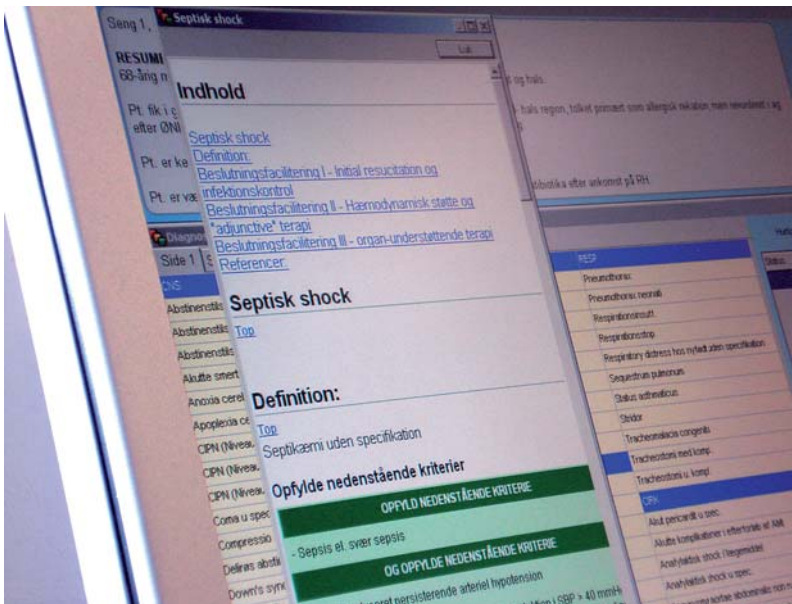
FIGURE 2

Brandt-Altman plot of SOFA scores before and after correction. Bias (mean difference —) and 95% limits of agreement (± 1.96 standard deviation - - -) are shown.



cluded in this study [9]. This may have resulted in an increased focus on septic shock patients leading to a higher accuracy for this particular diagnosis.

The typical error in SAPS II scoring was failure to account for mechanical ventilation and in the SOFA score failure to include inotrope/vasopressor administration. To further improve the quality of the data, we propose the introduction of an automatic message box in CIS to ensure that the user records correctly whether the patient is mechanically ventilated and/or receives vasopressor treatment. With the increasing use of elec-



More and more clinical databases become electronic.

tronic data capture, this may be done automatically to improve data.

The Bland-Altman plot shows that both SAPS II and SOFA scores were lower in the database than in the validated data. Even though the average bias was insignificant, this means that the clinicians generally underestimated SAPS II and SOFA in CIS compared with the corrected data. We believe that this apparent systematic error is acceptable, also when translated into SMRs. On the other hand, the limits of agreement around the mean of predicted mortality in the present cohort ranged from 37% to 73%. This represents a potential 36 percentage point difference in SMRs, which should be acknowledged and discussed when SMRs are used as a tool to assess the quality of care. Previous studies showed that SAPS II has limitations when used for quality of care assessment due to low calibration [10-12], but that mortality prediction can be improved when using customized models of SAPS II for septic shock patients [13].

A growing amount of data has shown that there can be substantial variation in inter-observer reliability when using the scoring systems; especially GCS scoring has been reported to suffer from low accuracy and reliability [14-17].

Studies have investigated if a training course or a refresher course could improve the quality of the data. In a single Finnish ICU, Tallgren et al [17] showed that a short refresher course in SOFA only improved the accuracy slightly. However, a study by Arts et al [14] showed that it is possible to generate scoring systems with a high level of reliability and accuracy. The authors credited the accuracy to the implementation of obligatory

training sessions for data collectors. In an experimental study, Arts et al also showed that the quality of the data improved with training [18].

We included no discrepancies between the GCS score and the recorded central nervous system (CNS) status in the final evaluation of SAPS II and SOFA because of limitations in obtaining the gold standard measure. When patients were sedated before arriving to the ICU, a description of their habitual mental status was seldom recorded by the clinicians. Moreover, all three parameters of GCS were rarely described individually in the medical records, which made it impossible to validate the GCS scoring in SOFA and SAPS II. Furthermore, the GCS was reportedly lower in SOFA than in SAPS II in a few cases. These problems rendered it impossible to define a gold standard for GCS scoring in our dataset.

In accordance with the existing knowledge, including corrected GCS in the evaluation of SAPS II and SOFA would presumably result in a lower accuracy than observed in this study. In order to achieve a higher accuracy and reliability in GCS scoring, we propose training sessions combined with written instructions as a possible solution. Tallgren et al [17] also proposed that an alternative and simpler neurological scoring tool be used in the ICU.

This study was based on seven different ICUs of both university and non-university hospitals in the whole country and included data from three consecutive years. This improves the external validity of our study. However, the fact that the study was retrospective is a weakness because it prevented us from verifying the source data. Moreover, all units used the same database system, so results comparable with other database systems may not be obtained. Additionally, we cannot know whether there were patients with septic shock in the units during the study period who were never coded as such in the database and therefore not subjected to validation.

CONCLUSION

In conclusion, the databases at the ICUs were of good accuracy for the diagnosis of septic shock and of reasonable reliability and accuracy for disease severity and organ dysfunction scores. The data in databases are therefore reliable tools for clinical research and quality-improvement as long as the potential variability is acknowledged and discussed.

CORRESPONDENCE: Lars Grønlykke, Blegdamsvej 78A, 3. th., 2100 Copenhagen, Denmark. E-mail: larsgroenlykke@gmail.com

ACCEPTED: 23 August 2012

CONFLICTS OF INTEREST: Disclosure forms provided by the authors are available with the full text of this article at www.danmedj.dk.

LITERATURE

1. Poulsen JB, Møller K, Kehlet H et al. Long-term physical outcome in patients with septic shock. *Acta Anaesth Scand* 2009;53:724-30.

2. Schrøder MA, Poulsen JB, Perner A. Acceptable long-term outcome in elderly intensive care unit patients. *Dan Med Bul* 2011;58(7):A4297.
3. Breslow MJ, Badawi O. Severity scoring in the critically ill: part 1 – interpretation and accuracy of outcome prediction scoring systems. *Chest* 2012;141:245-52.
4. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270:2957-63.
5. Aegerter P, Boumendil A, Retbi A et al. SAPS II revisited. *Int Care Med* 2005;31:416-23.
6. 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;22:707-10.
7. Vincent JL, de Mendonca A, Cantraine F et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine. *Crit Care Med* 1998;26:1793-800.
8. Levy MM, Fink MP, Marshall JC et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Critical Care Med* 2003;31:1250-6.
9. Perner A, Haase N, Guttormsen AB et al. Hydroxyethyl starch 130/0.42 versus Ringer's acetate in severe sepsis. *N Engl J Med* 2012;367:124-34.
10. Le Gall JR, Neumann A, Hemery F et al. Mortality prediction using SAPS II: an update for French intensive care units. *Crit Care* 2005;9:R645-52.
11. Juneja D, Singh O, Nasa P et al. Comparison of newer scoring systems with the conventional scoring systems in general intensive care population. *Minerva Anesth* 2012;78:194-200.
12. Capuzzo M, Valpondi V, Sgarbi A et al. Validation of severity scoring systems SAPS II and APACHE II in a single-center population. *Int Care Med* 2000;26:1779-85.
13. Arabi Y, Al Shirawi N, Memish Z et al. Assessment of six mortality prediction models in patients admitted with severe sepsis and septic shock to the intensive care unit: a prospective cohort study. *Crit Care* 2003;7:R116-22.
14. Arts D, de Keizer N, Scheffer GJ et al. Quality of data collected for severity of illness scores in the Dutch National Intensive Care Evaluation (NICE) registry. *Int Care Med* 2002;28:656-9.
15. Chen LM, Martin CM, Morrison TL et al. Interobserver variability in data collection of the APACHE II score in teaching and community hospitals. *Crit Care Med* 1999;27:1999-2004.
16. Arts DG, de Keizer NF, Vroom MB et al. Reliability and accuracy of Sequential Organ Failure Assessment (SOFA) scoring. *Crit Care Medicine* 2005;33:1988-93.
17. Tallgren M, Backlund M, Hynninen M. Accuracy of Sequential Organ Failure Assessment (SOFA) scoring in clinical practice. *Acta Anaesth Scand* 2009;53:39-45.
18. Arts DG, Bosman RJ, de Jonge E et al. Training in data definitions improves quality of intensive care data. *Crit Care* 2003;7:R179-R184.