J Anesth (2011) 25:839–844
DOI 10.1007/s00540-011-1232-2

ORIGINAL ARTICLE

Comparison of sequential organ failure assessment (SOFA) scoring between nurses and residents

Nur Baykara · Kaan Gökduman · Tülay Hoşten · Mine Solak · Kamil Toker

Received: 10 May 2011 / Accepted: 29 August 2011 / Published online: 20 September 2011

Abstract
Purpose We aimed to evaluate differences in the interobserver reliability and accuracy of sequential organ failure assessment (SOFA) scoring between nurses and residents.
Methods Eight nurses and eight residents independently scored 24 randomly selected patients. Intraclass correlation coefficients (ICCs) for the reliability of total SOFA scoring were calculated. The residents' and nurses' SOFA scores were compared with a gold standard to assess accuracy.
Results The overall ICC of the total SOFA score was 0.87 (nurses 0.89, residents 0.86) for a single measurement. Residents tended to assign higher total SOFA scores than did nurses, without a statistically significant difference (7.01 ± 4.43 vs. 6.72 ± 4.27, $P > 0.05$). The mean bias between the nurses' and the gold standard total SOFA scores was -0.16 ± 1.86 and the 95% confidence limit of agreement was -3.8 to $+3.49$. The mean bias between the residents' and the gold standard total SOFA scores was -0.39 ± 1.81, and the 95% confidence limit of agreement was -3.95 to $+3.16$. The percentage of accurate data for the total SOFA score was 47.4% for nurses and 51% for residents ($P > 0.05$). Although not statistically significant, the major error rate ($\geq$2 point deviation from the gold standard score) was higher for nurses than for residents (29.16 and 23.43%, $P > 0.05$). Accuracy of scoring individual organ systems was similar for the two groups; however, the major error rate in the cardiovascular system score was higher for nurses.
Conclusion Interobserver reliability was good and mean SOFA scores were not significantly different between nurses and residents. The accuracy of SOFA scoring was moderate for both groups; however, although the difference was not statistically significant, the major error rate was higher for nurses than for residents.

Keywords SOFA · Nurse · Resident · Interobserver reliability

Introduction

Prognostic scoring systems have become an important part of critical care practice, and play important roles in the prediction of outcome, evaluation of the effects of therapy, allocation of resources, and comparison between medical centers. A number of organ-dysfunction scores have been developed for use in critically ill patients, and one of the most commonly used is the sequential organ failure assessment (SOFA) score. To describe organ failure quantitatively, easily, and as objectively as possible, the SOFA system was developed by a consensus conference initiated by the European Society of Intensive Care Medicine [1]. The SOFA score has been validated prospectively in multiple studies [2–5].

The total SOFA score is calculated by summing the worst scores for each of the separate daily scores for the respiratory, renal, cardiovascular, coagulation, and hepatic systems, and the central nervous system (CNS). The total score ranges from 0 to 24, based on the scoring of each organ from 0 to 4, with a larger number indicating more severe failure. SOFA is calculated based on the most abnormal value in a 24-h period [1].

N. Baykara · K. Gökduman · T. Hoşten · M. Solak · K. Toker
Department of Anesthesiology and Reanimation,
Faculty of Medicine, University of Kocaeli,
Kocaeli, Turkey

N. Baykara (✉)
Çamlı Sok., 29/3 Suadiye, Istanbul, Turkey
e-mail: Baykaranur@yahoo.com

The reliability of a severity scale is the extent to which replicate observations give similar results [6]. Reapplication of the scale by the same rater can be investigated (intrarater reliability), and the consistency of response among different raters using the same scale can be assessed (interrater reliability). It is generally considered that quantifying agreement between different raters is the more rigorous test of reliability for severity scoring systems [6].

The interrater reliability of SOFA scoring has been assessed in a group of physicians and was found to be good [7]. The accuracy of SOFA scoring has also been assessed in physicians in two previous studies [7, 8]; however, the severity scoring systems are frequently used by different raters with different backrounds and levels of training. Nurses play an increasing role in the collection of severity score data. To our best knowledge, no study so far has compared SOFA scoring between nurses and physicians. The main objective of this study was to evaluate the interobserver reliability and accuracy of SOFA scoring when performed by nurses and residents.

Methods

The study protocol was approved by the local ethics committee of the University of Kocaeli. We conducted the study in a closed 12-bed, adult medical-surgical intensive care unit (ICU) at the medical center of Kocaeli University. This unit cares for any patient older than 17 years of age who requires intensive care, but not for patients who require cardiovascular surgery. Twenty-four patients were randomly selected from those who had been admitted to the ICU between August 2008 and January 2009. For each of the 24 patients, SOFA scoring was performed for the second day of their ICU stay. Each patient was scored by eight nurses and eight residents. This produced 16 SOFA scores for each individual patient.

Each nurse who participated in this study had at least 6 months' experience in the ICU. All of the residents in this study were in their second or third year of anesthesia residency, and had at least 6 months experience in the ICU. Even though all of the medical staff in our ICU were experienced in SOFA scoring, prior to the commencement of the study, all residents and nurses participating in the study underwent a training program during which SOFA data definitions and scoring rules were explained according to the recommendations in the original publication [1]. In the case of sedated patients, the presedation Glasgow Coma Scale (GCS) value had to be used when calculating the CNS score. Each rater was provided with a handout including special forms for the documentation of SOFA scores along with patient records. Each patient record contained a copy of the patient's chart, as well as laboratory and monitoring data. Each rater was also provided with a handout that included a conversion table to determine the correct fraction of the inspired oxygen concentration for patients who were not mechanically ventilated. Raters were blinded to the ratings of the other rating participants. The SOFA scores were calculated manually.

Patient records in our ICU are kept manually by residents and nurses. Mean arterial pressure, heart rate, and oxygen saturation are recorded at least once every hour by nurses. Laboratory tests are performed on a daily basis. Blood gas analyses are performed as required. All new laboratory test results are recorded in the patient's chart by the residents every morning. Residents and nurses work in 8-h shifts, and a complete report of each patient's condition is noted during each shift. All residents and nurses are instructed to keep extensive records in a uniform way. The charts are inspected by the staff intensivist during grand rounds.

To obtain the best possible gold-standard measure, the SOFA scores of these 24 patients were also assessed by the two of the authors who are attending physicians in the ICU. Both of these authors have had more than 5 years of ICU experience as senior physicians. The scores obtained following the consensus of these two authors were accepted as the gold standard.

We calculated a sample size of 24 patients to test whether a reliability of 0.9 exceeded a reliability of 0.8 [9]. The differences in mean SOFA scores between the two observer groups were compared using the paired *t*-test. We calculated intraclass correlation coefficients (ICCs) for a single measurement for the reliability of the total SOFA score. To interpret the ICCs, we used benchmarks suggested by Cicchetti and colleagues [10] ($\geq$0.75 excellent, 0.6–0.74 good, 0.4–0.59 fair, $<$0.4 poor).

To evaluate the accuracy of the total SOFA score, the agreement between residents' scores and the gold standard, and the agreement between nurses' scores and the gold standard were assessed using Bland–Altman plots. The Bland–Altman plot is a scatterplot in which variable means are represented on the horizontal axis, and the differences are represented on the vertical axis. This plot shows the magnitude of disagreement between the two measurements, and demonstrates how this disagreement relates to the magnitude of the measurements [11]. It is recommended that 95% of data points should lie within the $\pm$1.96 standard deviation (SD) of the mean differences [11]. If the differences observed in this plot are not deemed clinically significant (a decision not based on any *P* value), this is a confirmation of agreement. In the present study, a major error in the total SOFA score was defined as a $\geq$2 point deviation of the recorded values from the gold standard scores.

The percentage of accurate data and absolute deviation from gold standard scores were presented for total SOFA and component scores. The percentage of accurate data was

calculated by dividing the number of correctly recorded data items by the total number of data items that had been recorded. Differences in percentages of accurate data, or in deviations from gold standard scores between the two groups were tested using McNemar's test. Bland–Altman analysis was performed using GraphPad Prism software (version 5.0 for Windows; GraphPad Software, La Jolla, CA, USA). Other statistical analysis was performed using SPSS (version 13.0 for Windows; SPSS, Chicago, IL, USA).

Results

Patient clinical characteristics are presented in Table 1. The mean number of years that the nursing participants had worked in the ICU was 1.28 ± 1.56, and the mean number of years that the residents had worked in the ICU was 0.95 ± 0.28 ($P > 0.05$).

The overall ICC of the total SOFA score was 0.87 (nurses 0.89, residents 0.86) for a single measurement.

Table 1 Patient characteristics

Age, years	56.5 (23–78)
Sex, male/female	13/11
GCS (gold standard)	
15	10
10–14	5
<9	9
SOFA score, gold standard	6 (1–18)
Reason for ICU admission	
Major surgery	6
Multitrauma	7
Sepsis	2
GIS bleeding	1
Drug toxicity	1
Metabolic event	3
Cerebrovascular accident	3
Cardiopulmonary arrest	1
Mechanical ventilation	
Present	15 (62.5%)
Absent	9 (37.5%)
Sedative medication	
Administered	10 (41.6%)
Not given	14 (58.3%)
Inotropic agent	
Administered	9 (37.5%)
Not given	15 (62.5%)

Data are expressed as medians (ranges) or numbers (percentages) of patients

GCS Glasgow Coma Scale, *SOFA* sequential organ failure assessment, *ICU* intensive care unit, *GIS* gastrointestinal system

Even though residents tended to assign higher scores than did the nurses, there was no significant difference in mean total SOFA scores (residents 7.01 ± 4.43, nurses 6.72 ± 4.27, $P > 0.05$). The mean bias between the residents' and the gold standard total SOFA scores was −0.39 ± 1.81, and the 95% confidence limit of agreement was −3.95 to +3.16 (Fig. 1a). The mean bias between the nurses' and the gold standard total SOFA scores was −0.16 ± 1.86, and the 95% confidence limit of agreement was −3.8 to +3.49 (Fig. 1b).

The percentage of accurate data for the total SOFA score was 47.4% for the nurses, and 51% for the residents (Table 2, $P > 0.05$). Although not statistically significant, the difference in the major error rate (≥2 point deviation from the gold standard score) was greater for nurses than for residents (29.1 and 23.4%, respectively, $P > 0.05$).

The numbers and percentages of accurate values are presented, stratified by organ system, in Table 2. Both nurses and residents correctly evaluated coagulation, liver, and renal systems in more than 90% of the patients (Table 2). Cardiovascular system (CVS) scores were correct in at least 80% of patients for both groups (Table 2). Even though absolute accuracy was similar for nurses and residents for the CVS score, the major error rate (deviation

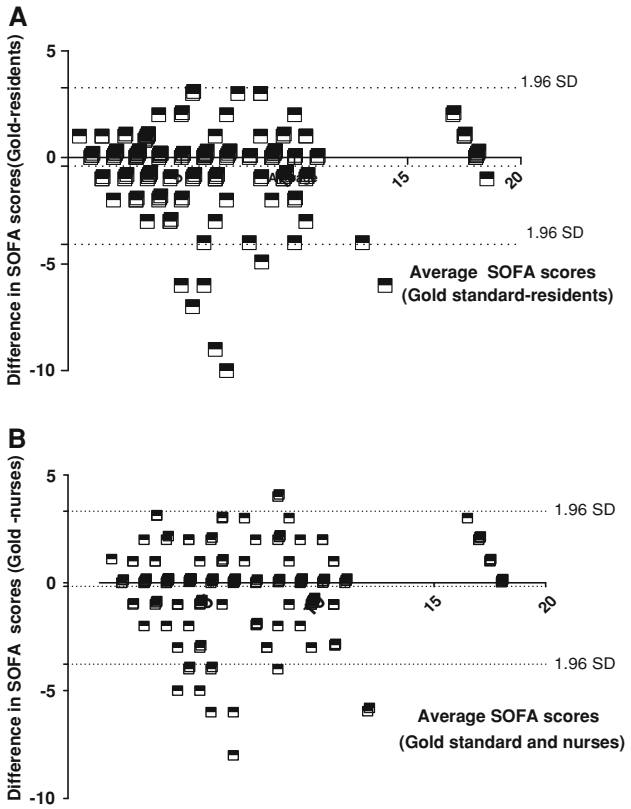

Fig. 1 Bland-Altman plots of sequential organ failure assessment (*SOFA*) scores for gold standard and residents (**a**) and for gold standard and nurses (**b**)

Table 2 Distribution of cases according to the absolute difference of the nurse/resident score minus the gold standard score

	Nurses			Residents		
	0	1	≥2	0	1	≥2
Respiration	141 (73.4%)	39 (20.3%)	12 (6.25%)	134 (69.8%)	49 (25.5%)	9 (4.7%)
Coagulation	188 (97.9%)	4 (2.08%)	–	187 (97.3%)	5 (2.6%)	–
Hepatic	190 (98.9)	2 (1.04%)	–	188 (97.6%)	4 (2.08%)	–
Cardiovascular	159 (82.9%)	22 (11.4%)	11 (5.7%)	165 (86%)	25 (13%)	2 (1.04%)*
Neurological	145 (75.5%)	45 (23.4%)	2 (1.04%)	153 (79.6%)	31 (16.1%)	8 (4.16%)
Renal	173 (90.1%)	15 (7.8%)	4 (2.08%)	180 (93.7%)	11 (5.72%)	1 (0.52%)
Total SOFA score	91 (47.4%)	45 (23.4%)	56 (29.1%)	98 (51.0%)	49 (25.5%)	45 (23.4%)

Data are expressed as numbers (percentages) of cases

SOFA sequential organ failure assessment

* Significantly different ($P < 0.05$), between nurses and residents

Table 3 The most frequent errors in SOFA scoring

	Nurses	Residents
Failing to choose the worst PaO_2/FiO_2 value	40 (20.8%)	42 (21.8%)
Not ignoring the effect of sedation	25 (13.0%)	15 (7.8%)
Scoring CVS as 0 instead of 1 despite MAP <70 mmHg	15 (7.8%)	15 (7.8%)
Calculation errors in computing GCS or total SOFA scores	16 (8.3%)	10 (5.2%)
Failing to take into consideration the use of inotropic agents	11 (5.7%)	5 (2.6%)
Failing to take into consideration low urine output	12 (6.2%)	7 (3.6%)
Using the wrong FiO_2 value	11 (5.7%)	11 (5.7%)
Not using the lowest values	6 (3.1%)	10 (5.2%)
Incorrect calculation of inotropic agent dose	6 (3.1%)	5 (2.6%)
Carelessness, selecting the wrong score, despite using the correct data	2 (1.04%)	11 (5.7%)*

Data are expressed as numbers (percentages) of cases

CVS Cardiovascular system, *MAP* mean arterial pressure, *GCS* Glasgow Coma Scale, *SOFA* sequential organ failure assessment

* Significantly different ($P < 0.05$), between nurses and residents

from the gold standard ≥2) was greater among nurses (Table 2, $P < 0.05$). The frequency of correct scoring was about 70% for the respiratory system and CNS for both nurses and residents.

The most frequent errors in SOFA scoring are presented in Table 3. Although the types of errors made by residents and nurses were similar in the two groups, some differences were found. For example, residents made more careless errors than did nurses ($P < 0.05$).

Discussion

Accurate data collection is an essential component of high-quality clinical research. Assessing interrater agreement between nurses and physicians is important, because they work in collaboration in clinical practice and research. In a previous study [12], the interrater reliability of nurses and residents collecting APACHE II (Acute Physiology and Chronic Health Evaluation II) data was determined. The ICC was found to be 0.95 for these two observer groups, and residents were more accurate data collectors than nurses [12]. In the present study, the interobserver reliability of the total SOFA score was investigated for raters who were nurses and residents. The overall ICC for a single measurement was found to be 0.87 (nurses 0.89, residents 0.86). In a previous study [7] in which the study design was similar to that of our study, Arts et al. investigated interobserver reliability of SOFA for physicians. The ICC was found to be 0.889 for the total SOFA score [7].

In the present study, although residents tended to assign higher scores than did nurses, the difference in mean SOFA scores between the two groups was not statistically significant. The percentage of accurate data for the total SOFA score was 47.4% for the nurses, and 51% for the residents ($P > 0.05$). The major error rate (≥2 point deviation from the gold standard score) was 29.1% for the nurses, and 23.4% for the residents. Similarly, in their experimental study, Arts et al. [7] showed that SOFA scores assigned by physicians were accurate in 53% of the cases, and that 19% of the SOFA scores assigned by physicians deviated by ≥2 points from the gold standard scores. The accuracy of SOFA scores recorded by physicians was also evaluated in a clinical study, which demonstrated that only half of the scores determined by

physicians were accurate in clinical practice [8]. Our study results support the notion that the accuracy of SOFA is not as high as expected. However the accuracy of SOFA is higher than that of APACHE II. Some studies have reported the accuracy of APACHE II to be as low as 14%, even for trained clinical trial personnel [13].

Even though the major error rate of the total SOFA score was higher for the nurses in our study, the difference betwen the two groups did not reach statistical significance; however, the difference might be significant in some situations, such as during data collection processes for clinical trials. The major error rate in the cardiovascular system (CVS) score was also higher for nurses than for residents ($P < 0.05$). This was mostly due to the nurses failing to take into consideration the inotropic agent dose. Due to the shortage of well-educated specialized personnel, nurses play an increasing role in the collection of severity score data. We believe that nurses should receive one-on-one education by an experienced intensivist before calculating the SOFA score of patients.

In our study, both the nurses and the residents correctly evaluated the coagulation, liver, and renal systems in more than 90% of cases. A relatively small number of accurate values were recorded for the respiratory system and CNS in both groups. This is an expected result because it has been shown that severity measures that depend on the extraction of data from medical records had higher reliability than those that depend on calculation, judgement, or computing scores from an algorithm [6, 7].

Although the types of errors made by residents and nurses in our study were similar for the two groups, some differences were found. For example, residents made more careless errors than did nurses. It is hard to explain why residents made more careless errors than did nurses in this study. As it is in many other clinics, residency is a stressful, overwhelming period in our clinic. Due to the stresses of resident training, including excessive workload, the large body of clinical knowledge to master, sleep deprivation, difficult patient problems, and the challenges of balancing work and home life, burnout and depression are highly prevalent among residents worldwide and across specialties [14, 15]. Some studies have found prevalence rates of burnout in residents to be between 41 and 76%, and rates of depression range from 7 to 56% [14, 15]. Even though burnout syndrome is high in all ICU healthcare workers, burnout syndrome is more common among ICU physicians than among nurses (one-half of the intensivists vs. one-third of the nurses) [16]. We speculate that fatigue, distress, probable burnout, and depression might have made residents more prone to careless mistakes than nurses. Residents may also accept small rating errors unintentionally instead of pursuing strict accuracy because they understand that scoring in acute medicine places special emphasis on simplicity and convenience of use rather than complexity and accuracy. To minimize careless errors, it may be necessary to employ specialist staff engaged exclusively in data collection.

In our study, similar to the data reported by a previous study [7], the most frequent error in both groups was made in choosing the worst PaO_2/FiO_2 ratio. This type of error was made especially in patients who required frequent arterial blood gas measurements within 24 h prior to SOFA scoring. In the present study, the PaO_2/FiO_2 ratio was calculated by hand. In situations in which several blood gas analyses are performed for the same patient, manual determination of the worst PaO_2/FiO_2 ratio requires much calculation and close attention, and is time-consuming. Using an ICU data system that automatically calculates the PaO_2/FiO_2 ratios, Tallgren et al. [8] reported 87% accuracy of the SOFA respiratory component score. Thus, this type of error can be decreased by using an ICU data system to calculate the PaO_2/FiO_2 ratio automatically, although the error cannot be completely abolished.

The CNS component of SOFA is based on the GCS. The frequency of correct scoring for the GCS was relatively low (about 70%) in both of the groups in our study. Our study design was experimental. The participants in this study did not see the real patients, a factor which could have contributed to the relatively low accuracy of the GCS. However; similar to our results, previous studies reported low accuracy and reliability of determining the GCS in both clinical and experimental circumstances [7, 17, 18].

Although the accuracy of the total SOFA score was not perfect for either nurses or residents in our study, inter-observer reliability was excellent for both groups. This might have been due to raters having made similar errors, such as ignoring sedation when calculating the GCS, or ignoring the use of an inotropic agent when mean arterial pressure (MAP) was >70 mmHg, or scoring CVS as 0 instead of 1 when MAP was <70 mmHg.

To improve the accuracy of severity scoring systems, some measures have been suggested, such as the application of a training program, and strict guidelines [19, 20]. It was shown that the interobserver variability of APACHE II, a severity of illness score more complicated than the SOFA, decreased, especially among more experienced physicians, when regular training and strict guidelines were applied [19]; however, in Tallgren's study [8], even though the major error rate decreased, the accuracy of SOFA scoring improved only slightly after a refresher course. In the present study, although the medical staff who took part in our study were experienced in SOFA scoring and also underwent a rigorous training program, the accuracy of SOFA scoring was still moderate. It may be more helpful if the medical staff who score the patients were to receive one-on-one education by experienced intensivists, or if

very well-educated specialized personnel were employed to score the patients; however, the reliability and accuracy of SOFA scoring depends to some degree on the reliability and accuracy of the GCS, another severity scoring system which has been shown previously to have low accuracy and reliability [7, 17, 18]. CNS scoring is based on the GCS in all severity of illness scores used in the ICU. The Simplified Motor Scale, which has been developed recently, predicts mortality equally well as the GCS in patients with traumatic brain injury, and with better interrater reliability [21]. We recommend that other neurological scales, such as the Simplified Motor Scale, should be taken into consideration instead of the GCS for neurological classification when SOFA is updated in the future.

In our retrospective single-center study, ınterobserver variability of SOFA was found to be excellent for nurses and residents, and the accuracy for both groups was moderate. Prospectively collected data may be more reliable and more accurate than retrospectively collected data [22]. Different results might be obtained from prospective, multicenter studies, but this remains to be investigated.

In conclusion, although residents tended to assign higher scores than did nurses, the difference between the two groups did not reach statistical significance. The accuracy of SOFA scoring was not good enough for either residents or nurses. Although the difference was not statistically significant, the major error rate was higher for the nurses than for the residents.

Acknowledgments The authors thank Mrs. Laura Danner for editing the English translation of the manuscript.

References

1. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996;22:707–10.
2. Bota DP, Melot C, Ferreira FL, Ba VN, Vincent JL. The multiple organ dysfunction score (MODS) versus the sequential organ failure assessment (SOFA) score in outcome prediction. Intensive Care Med. 2002;28:1619–24.
3. Vincent JL, De Mendonça A, Cantraine F, Moreno R, Takala J, Suter PM, Sprung CL, Colardyn F, Blecher S. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicentric, prospective study. Crit Care Med. 1998;26:1793–800.
4. Moreno R, Vincent JL, Matos R, Mendonça A, Cantraine F, Thijs L, Takala J, Sprung C, Antonelli M, Bruining H, Willatts S. The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective multicenter study. Intensive Care Med. 1999;25:686–96.
5. Lopes Ferreira F, Peres Bota D, Bross A, Melot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome. JAMA. 2001;286:1754–8.
6. Shiels C, Eccles M, Hutchinson A, Gardiner E, Smoljanovic L. The inter-rater reliability of a generic measure of severity of illness. Fam Pract. 1997;14:466–71.
7. 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.
8. Tallgren M, Backlund M, Hynninen M. Accuracy of sequential organ failure assessment (SOFA) scoring in clinical practice. Acta Anaesthesiol Scand. 2009;53:39–45.
9. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med. 1998;7:101–10.
10. Cicchetti DV, Volkmar F, Sparrow SS, Cohen D, Fermanian J, Rourke BP. Assessing the reliability of clinical scales when data have both nominal and ordinal features: proposed guidelines for neuropsychological assessments. J Clin Exp Neuropsychol. 1992;14:673–86.
11. Bland JM, Altman DJ. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;8:307–10.
12. Holt AW, Bury LK, Bersten AD, Skowronski GA, Vedig AE. Prospective evaluation of residents and nurses as severity score data collectors. Crit Care Med. 1992;20:1688–91.
13. Booth FV, Short M, Shorr AF, Arkins N, Bates B, Qualy RL, Levy H. Application of a population-based severity scoring system to individual patients results in frequent misclassification. Crit Care. 2005;9:R522–9.
14. Thomas NK. Resident burnout. JAMA. 2004;292:2880–9.
15. Fahrenkopf AM, Sectish TC, Barger LK, Sharek PJ, Lewin D, Chiang VW, Edwards S, Wiedermann BL, Landrigan CP. Rates of medication errors among depressed and burnt out residents: prospective cohort study. BMJ. 2008;336(7642):488–91.
16. Embriaco N, Papazian L, Kentish-Barnes N, Pochard F, Azoulay E. Burnout syndrome among critical care healthcare workers. Curr Opin Crit Care. 2007;13:482–8.
17. Holdgate A, Ching N, Angonese L. Variability in agreement between physicians and nurses when measuring the Glasgow Coma Scale in the emergency department limits its clinical usefulness. Emerg Med Australas. 2006;18:379–84.
18. Gill MR, Reiley DG, Green SM. Interrater reliability of Glasgow Coma Scale scores in the emergency department. Ann Emerg Med. 2004;43:215–23.
19. Polderman KH, Jorna EM, Girbes AR. Inter-observer variability in APACHE II scoring: effect of strict guidelines and training. Intensive Care Med. 2001;27:1365–9.
20. Arts DG, Bosman RJ, de Jonge E, Joore JC, de Keizer NF. Training in data definitions improves quality of intensive care data. Crit Care. 2003;7:179–84.
21. Gill M, Martens K, Lynch EL, Salih A, Green SM. Interrater reliability of 3 simplified neurologic scales applied to adults presenting to the emergency department with altered levels of consciousness. Ann Emerg Med. 2007;49(4):403–7.
22. Romm FJ, Putnam SM. The validity of the medical record. Med Care. 1981;19(3):310–5.