Neuroanesthesia and Intensive Care

Limited ability of SOFA and MOD scores to discriminate outcome: a prospective evaluation in 1,436 patients

[Limites de la capacité discriminatoire des scores de SOFA et de DMV : une évaluation prospective chez 1 436 patients]

David A. Zygun MD FRCPC,*†§ Kevin B. Laupland MD FRCPC,*† Gordon H. Fick PhD,‡
James Dean Sandham MD FRCP FACP,*† Christopher J. Doig MD MSc FRCPC*†‡

Purpose: The multiple organ dysfunction (MOD) score and sequential organ failure assessment (SOFA) score are measures of organ dysfunction and have been validated based on the association of these scores with mortality. We sought to compare the performance of the SOFA and MOD scores in a large cohort of consecutive multisystem intensive care unit (ICU) patients.

Methods: Prospective automated daily measurements of MOD and SOFA scores were performed in 1,436 patients admitted to a multisystem ICU in the Calgary Health Region over a one-year period. Logistic regression modeling techniques were used to describe the association of SOFA and MODS with mortality. Receiver operator characteristic (ROC) curves were used to assess the model's discriminatory ability.

Results: For ICU and hospital mortality, there was very little practical difference between the SOFA and MOD scores in their ability to discriminate outcome as determined by the area under the ROC. However, compared to previous literature, the discriminatory ability of both scores in this population was weak. As well, the calibration of the models was poor for both scores. The SOFA cardiovascular component score performed better than the MOD cardiovascular component score in the discrimination of both ICU and hospital mortality.

Conclusions: SOFA and MOD scores had only a modest ability to discriminate between survivors and non-survivors. These results question the appropriateness of using organ dysfunction scores as a 'surrogate' for mortality in clinical trials and suggest further work is necessary to better understand the temporal relationship and course of organ failure with mortality.

Objectif : *Le score de défaillance multiviscérale (DMV) et le score de SOFA (Sequential Organ Failure Assessment) mesurent la dysfonction organique et ont été validés sur leur association avec la mortalité. Nous avons comparé leur performance chez des patients successivement admis à l'unité des soins intensifs (USI) pour une atteinte multiviscérale.*

Méthode : *Les mesures quotidiennes, prospectives et automatisés des scores de DMV et de SOFA ont été faites chez 1 436 patients admis à une USI multiviscéraux dans la région de Calgary pendant une année. Une modélisation de régression logistique a servi à décrire l'association des scores de SOFA et de DMV à la mortalité. La capacité discriminatoire du modèle a été évaluée par les courbes ROC (Receiver Operator Characteristic).*

Résultats : *Concernant la mortalité à l'USI et à l'hôpital, les scores de SOFA et de DMV présentaient une très petite différence pratique de capacité à distinguer les résultats comme l'a montré l'aire sous la courbe ROC. Comparée aux données des publications antérieures, la*

From the Departments of Critical Care Medicine,* Medicine,† Community Health Sciences,‡ and Clinical Neurosciences,§ University of Calgary, Calgary, Alberta Canada.

Address correspondence to: Dr. Christopher J. Doig, Departments of Medicine and Community Health Sciences, Faculty of Medicine, University of Calgary, Department of Critical Care, Calgary Health Region, Room EG23G, Foothills Medical Centre, 1403 - 29th Street, NW, Calgary, Alberta T2N 2T9, Canada. Phone: 403-944-1691; Fax: 403-283-9994; E-mail: cdoig@ucalgary.ca

Support: Dr. Zygun was supported by the Meredith Graduate Masters Scholarship funded by the Workers Compensation Board of Alberta. Dr. Laupland was supported by the 2000 Bayer Healthcare/Canadian Institutes of Health Research/Canadian Infectious Diseases Society Research Fellowship and a clinical fellowship award from the Alberta Heritage Foundation for Medical Research (AHFMR). Funding for this work was supported in part by grants from the Alberta Heritage Foundation for Medical Research Health Research Fund, and the Calgary Regional Health Authority Special Competition Fund.

Accepted for publication January 27, 2004.

Revision accepted December 7, 2004.

capacité discriminatoire des deux scores était faible pour la population évaluée. Aussi, le calibrage des modèles était pauvre pour les deux scores. Le score de la composante cardiovasculaire du SOFA a présenté une meilleure performance que celui de la DMV quant à la détermination de la mortalité à l'USI et à l'hôpital.

Conclusion : *Les scores de SOFA et de DMV n'ont qu'une faible capacité à distinguer les patients qui vont survivre ou non. Cela remet en question la pertinence d'utiliser des scores de dysfonction organique comme «substitut» à la mortalité dans les essais cliniques et incite à chercher à découvrir la relation temporelle entre l'évolution de la défaillance organique et la mortalité.*

MULTIPLE organ dysfunction (MOD) syndrome-also variably known as multi-system organ failure or multi-organ failure syndrome-is a clinical pattern of progressive and sequential organ dysfunction which is common in intensive care unit (ICU) patients, and is often a precursor of death in critically ill patients.[1–4] To describe the severity of organ dysfunction, and the natural history, the sequential organ failure assessment (SOFA) score, and the MOD score have been reported, and are now commonly incorporated into clinical trials in critical care.[5–8] Both systems possess, as originally suggested by Marshall *et al.*, ideal important characteristics for systems to describe organ dysfunction: the variables should be objective, simple and easily available but reliable.[6] They should be specific for the organ system being considered and of a continuous nature. In addition, they should also be independent of patient type and therapeutic intervention.

Although SOFA and MOD evaluate the same six organ systems, there are practical differences which may affect the operating characteristics of each system. The MOD score was developed based on a critical appraisal of the literature, and then 'cut-off' scores for each organ system were derived and validated based on probability of subsequent mortality in a sample of 692 patients from one Canadian ICU.[6] The SOFA score was developed by a consensus conference at a meeting of the European Society of Intensive Care Medicine (ESICM). The SOFA score has been validated against mortality in multiple subsequent studies.[9–12]

The other significant difference is in the timing in calculating the score in each system. SOFA is calculated based on the most abnormal value in a 24-hr period, whereas the MOD score is calculated using physiologic values measured at the same point in time every day (first morning values) to avoid capturing momentary physiologic changes unrelated to changes in the patient's underlying physiologic status.

Given that these two scores are now being incorporated into clinical trials in critical care,[7,8] despite the caveats that each was developed using disparate techniques and each uses a different technique to assess the severity of physiologic derangement, the objective of our study was to compare the correlation of the MOD and SOFA scores with mortality in a large heterogeneous population of adult multisystem ICU patients.

Methods

The study was approved by the local Institutional Review Board. The Calgary Health Region has three adult multisystem ICUs, and serves a referral population of approximately 1.2 million persons. The study sample involved an inception cohort that included consecutive critically ill patients admitted in any one of these three ICUs from May 1, 2000 to April 30, 2001. Patients admitted to a regional coronary care unit, or admitted to a postoperative cardiovascular surgery unit were excluded. Consistent with previous recommendations, patients who survived ICU with a length of stay < 48 hr following uncomplicated elective surgery were also excluded. Only a patient's first admission to the ICU during that hospital admission was considered.

The SOFA and MOD scores were collected following the recommendations in the original publications.[5,6] An electronic patient information system [Quantitative Sentinel (QS), GE-Marquette Medical Systems Inc., Chicago, IL, USA] interfaced to all bedside devices collected physiologic data, and these data were validated (accepted by the system) by nursing or respiratory therapy staff on at least an hourly basis by examining the representativeness and sensibility of the data. An HL-7 interface with the regional laboratory information system (Cerner PathNet Classic, version 306, Kansas City, MO, USA) was utilized to collect all laboratory data. Two programs were developed in visual basic (Microsoft VBL, Microsoft Corporation, Seattle, WA, USA) to examine all physiologic and laboratory values in each 24-hr period, measured daily from 0000 to 2359 hours. For the SOFA score, one visual basic program determined the most abnormal value for each variable. The program then calculated the appropriate SOFA value (range 0–4), which was then exported to a local longitudinal ICU database known as TRACER (Microsoft Access, Microsoft Corporation, Seattle, WA, USA). Missing SOFA values were replaced between a preceding and subsequent SOFA value with the lower of the two scores. In the absence of a preceding or subsequent SOFA value, the score was calculated at zero. In the second visual basic program, the least abnormal value at 0700 ± 2 hr

was used to calculate the appropriate MOD score. Imputation of missing values was conducted in a similar manner as to the SOFA score. The calculation of each component system value and the total values for both SOFA and MOD scores were checked manually (C.J.D.) for their accuracy by comparing to the laboratory or physiologic data recorded in the QS system over a one-month period (683 patient days) prior to the start of the study; no errors were found in the calculation of either score. Clinical data were collected on all patients at the time of entry to the ICU. Outcome data were collected at the time of ICU discharge, and hospital discharge.

Statistical analysis

All data were analyzed using STATA-7 and STATA-8 (Stata Corporation, College Station, TX, USA). Our primary objective was to describe the relationship of SOFA and MOD scores with outcome. Consistent with previous publications we chose admission, patient mean, delta and maximum values to describe the association of SOFA and MOD scores with outcome. Maximum values were defined as the sum of the most abnormal component scores during the patient's ICU stay. Delta scores, a measure of the degree of organ dysfunction acquired during ICU stay, were defined as the difference between the maximum score and admission score. In addition, we also compared individual components scores. Organ failure was defined by a component score ≥ 3. Logistic regression modelling techniques were used to describe the relative strength of the relationship of SOFA and MOD scores with mortality. Model calibration was examined using the Hosmer-Lemeshow goodness of fit test.[13] Area under the receiver operator characteristic (AuROC) curves were used to assess model discriminatory ability. A P value of < 0.05 was considered significant.

TABLE I Patient demographics

Number of patients	1436
Number of observations	8588
Age (mean ± SD)	58 ± 19
Age range	14–100
Gender (male/female)	825/596
Diagnostic category	
Cardiovascular	12.9%
Gastrointestinal	7.7%
Metabolic/endocrine/renal	2.0%
Poisoning	1.8%
Neurological	8.3%
Shock	13.7%
Trauma	11.5%
Respiratory	31.3%
Other/missing	10.9%
Admission Apache (mean ± SD)	25 ± 9
Postoperative	26.0%
Infection at admission	20.5%
ICU length of stay, days [median (IQR)]	4.1 (2.3–8.0)
Hospital length of stay, days [median (IQR)]	15 (7–32)
ICU mortality	27.0%
Hospital mortality	35.1%

ICU = intensive care unit; IQR = interquartile range.

TABLE II SOFA and MOD scores

SOFA	*Mean (± SD)*
Admission	6.2 (3.5)
Maximum	10.0 (4.2)
Delta	2.6 (2.7)
Mean	6.1 (3.2)
MOD	
Admission	4.6 (2.8)
Maximum	8.4 (4.1)
Delta	2.7 (2.8)
Mean	4.6 (2.6)

SOFA = sequential organ failure assessment; MOD = multiple organ dysfunction.

Results

The demographics of the population are presented in Table I. Eight thousand five hundred and eighty-eight observations were collected on 1,436 patients during the study period. Mean admission, maximum, delta and patient mean SOFA and MOD scores are presented in Table II. The daily mean values of these scores were parallel over time. The time course of organ dysfunction as measured by mean SOFA and MOD scores by survival status is presented in Figure 1.

Association with and ability to discriminate outcome

Logistic regression models were created to describe the association of both hospital and ICU mortality with SOFA and MOD scores and the results are presented in Table III. For ICU mortality, both scoring systems only displayed a modest ability to discriminate outcome and there was little practical difference between these two scoring systems as determined by the AuROC. Mean SOFA and MOD values had the best discriminatory ability while the delta MOD and SOFA values had the worst (Figure 2). However, these models had poor calibration (Hosmer-Lemeshow goodness of fit test, $P < 0.01$). Models incorrectly classified patients 20 to 27% of the time (classified as positive if predicted probability of mortality ≥ 0.5). For hospital mortality, discrimination and/or calibration was also generally poor for each of

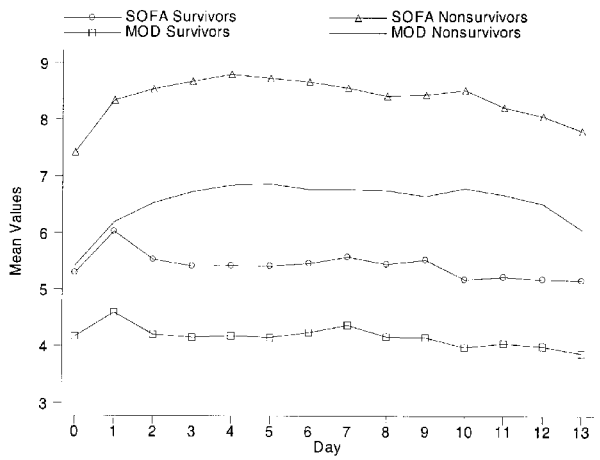

FIGURE 1 Mean daily sequential organ failure assessment (SOFA) and multiple organ dysfunction (MOD) scores by survival status.

the models with little difference between the scoring systems (Figure 2). Models incorrectly classified patients 25 to 35% of the time.

Component scores and outcome

In a similar fashion, we also investigated the performance of each patient's maximum SOFA and MOD component scores. For ICU mortality, SOFA cardiovascular component score had a significantly better ability to discriminate outcome than the MOD cardiovascular score [AuROC curve (95% confidence interval (CI)) of 0.74 (0.71, 0.77) *vs* 0.64 (0.60, 0.67)]. Other organ system component scores were not able to discriminate outcome as well as the cardiovascular component scores and there was little difference between SOFA and MOD. Except for the hepatic component scores, calibration was poor for all component score models. For hospital mortality, the SOFA cardiovascular component score again performed significantly better than the MOD cardiovascular component score [AuROC (95% CI) 0.71 (0.68, 0.74) *vs* 0.63 (0.59, 0.66)]. Other SOFA and MOD component scores did not differ in their ability to discriminate hospital mortality. The discrimination of all models was poor to modest at best (AuROC curves were all less than 0.63). Calibration was poor for all systems except the hepatic system.

To investigate the independent association of each component score with hospital mortality, multivariable logistic regression models were created for MOD and SOFA component scores. Backwards step-wise elimination was used to produce the most parsimonious model. For both the SOFA and MOD scores, neurological, cardiovascular and renal component scores remain significantly and independently associated with hospital mortality while scores for the hepatic, coagulation and respiratory fell out of both models.

Discussion

In this prospective, population-based study, SOFA and MOD scores were found to be associated with the ICU and hospital mortality, congruent with previous literature. However, the discriminatory ability of the maximum and admission scores in our population was considerably lower than previously described. Moreno *et al.* on behalf of the working group on sepsis related problems of the ESICM reported on the discriminatory ability of the maximum and admission SOFA scores to predict ICU mortality.[11] They found the AuROC to be 0.85 for the maximum SOFA score and 0.77 for the admission SOFA score compared with 0.69 and 0.67, respectively, in our study. Although these studies were similar in size and both excluded routine postoperative patients staying less than 48 hr, several important differences exist. The Moreno study involved 40 ICUs from 16 countries and collected data over one month in 1995. Our study was population-based as we included all critically ill patients admitted to adult multidisciplinary ICUs (both academic and community based) in a large well-defined region. Thus, as sampling was not performed, selection bias was minimized. Further, the duration of our study prevents over or under representation of certain diagnoses because of seasonal variation.

Marshall *et al.*, in their initial description of the MOD score, reported excellent discrimination (AuROC of 0.93) for the maximum MOD score.[6] We found an AuROC of only 0.64. It should be noted that patients in Marshall's study were all surgical patients while our study included surgical and medical patients. In addition, it is unclear whether Marshall enrolled all consecutive patients. The admission APACHE II score in our patients was significantly higher than those in Marshall's study (25 *vs* 13) with a proportionate greater ICU mortality (27% *vs* 9%). Marshall's results are based on the validation sample which consisted of 356 patients compared to our study of 1,400 patients. Given the marked differences in patient demographics, it is difficult to definitely resolve the disparity of the results between these studies. Further, as the MOD score was developed over a decade ago, it is possible the change in discrimination ability observed in this study is related to changes in patient care over these years. Although a more recent study by Cook and colleagues[14] examined the relation

TABLE III Prediction of mortality: logistic regression models

Variable	*OR (95% CI)*	*P value*	*AuROC (95% CI)*	*Correctly* classified*	*Goodness of fit χ^2 statistic*	*Goodness of fit P value*
ICU mortality						
Admission SOFA	1.20 (1.16, 1.25)	< 0.001	0.67 (0.63, 0.70)	74%	11.66	0.17
Admission MOD	1.18 (1.13, 1.24)	< 0.001	0.63 (0.59, 0.67)	74%	14.29	0.05
Maximum SOFA	1.18 (1.14, 1.22)	< 0.001	0.69 (0.66, 0.73)	74%	7.30	0.50
Maximum MOD	1.14 (1.10, 1.18)	< 0.001	0.64 (0.60, 0.68)	74%	9.09	0.33
Mean SOFA	1.47 (1.40, 1.55)	< 0.001	0.79 (0.75, 0.82)	80%	28.92	< 0.01
Mean MOD	1.47 (1.39, 1.56)	< 0.001	0.75 (0.71, 0.78)	78%	42.72	< 0.01
Delta SOFA	1.08 (1.03, 1.13)	< 0.001	0.51 (0.47, 0.55)	73%	98.01	< 0.01
Delta MOD	1.09 (1.04, 1.14)	< 0.001	0.52 (0.48, 0.56)	73%	70.52	< 0.01
Hospital mortality						
Admission SOFA	1.20 (1.15, 1.25)	< 0.001	0.67 (0.63, 0.71)	68%	8.8	0.36
Admission MOD	1.18 (1.13, 1.23)	< 0.001	0.62 (0.58, 0.66)	67%	10.28	0.17
Maximum SOFA	1.19 (1.15, 1.23)	< 0.001	0.70 (0.66, 0.73)	69%	9.2	0.33
Maximum MOD	1.15 (1.11, 1.19)	< 0.001	0.65 (0.61, 0.68)	68%	8.07	0.43
Mean SOFA	1.44 (1.37, 1.52)	< 0.001	0.77 (0.74, 0.80)	75%	22.66	< 0.01
Mean MOD	1.47 (1.39, 1.56)	< 0.001	0.74 (0.70, 0.77)	72%	46.13	< 0.01
Delta SOFA	1.09 (1.05, 1.13)	< 0.001	0.54 (0.50, 0.57)	65%	53.48	< 0.01
Delta MOD	1.11 (1.06, 1.16)	< 0.001	0.55 (0.51, 0.59)	65%	31.25	< 0.01

OR = odds ratio; 95% CI = 95% confidence interval; AuROC = area under the receiver operating characteristic curve; ICU = intensive care unit; SOFA = sequential organ failure assessment; MOD = multiple organ dysfunction; *Classified as + if predicted mortality ≥ 0.5.

between the six components of the MOD score with time to death in a selected sample of 1,200 ICU patients, comparisons between Cook's data and ours are not valid for two important reasons. Firstly, the Cook study did not collect the MOD score as originally described but opted for a "simpler" cardiovascular component score. Neither the rationale for this change nor the ability of this unique cardiovascular score to discriminate outcome was reported. In addition, the data were derived from patients entered into a large randomized controlled trial.[15] In this trial, only 1,200 of 4,232 patients with ICU length of stay greater than 48 hr were included. The use of this highly selected sample makes generalization of their results problematic.

Two other studies have directly compared SOFA and MOD scores. Pettilä and colleagues recently compared the ability of SOFA, MOD, logistic organ dysfunction score (LODS) and APACHE III to predict hospital mortality.[16] In this single-centre study, SOFA, MOD and LOD scores were calculated retrospectively for day one, three, five, and seven. However, MOD score was calculated for each day by using the worst single value of that day. This is contrary to previously described methods that used physiologic values measured at the same point in time every day (first morning values). Thus, this study does not use the MOD score as described by Marshall *et al.*[6] Further, because this study did not collect SOFA and MOD scores for all days of ICU admission, it must be interpreted with caution. The authors found an AuROC curve of 0.776 for admission SOFA score and 0.695 for admission MOD score. The AuROC curve for maximum SOFA score was 0.816 and 0.817 for the maximum MOD score. Given the lack of daily collection and the variant method of collection of the MOD score, it is difficult to comment on the comparability of these results to previous studies. Although the AuROC curve is lower than previously reported, this may be secondary to methodological differences.

The second study was published in 2002 by Peres Bota and colleagues.[17] In this single-centre study, they collected SOFA and MOD scores every 48 hr until discharge. They found both scores to have good discrimination in outcome prediction (maximum MOD AuROC 0.900, maximum SOFA AuROC 0.898) and concluded they are reliable outcome measures. Several important differences between our study and the study by Peres Bota warrant discussion. First, our study calculated prospectively for each day of ICU stay in a manner consistent with previously reported methods. In addition, we used a population-based cohort thereby strengthening the generalizability of our results. Another major difference in the collection of these scores was our use of an automated collection system. Automated sampling has been shown to influence the

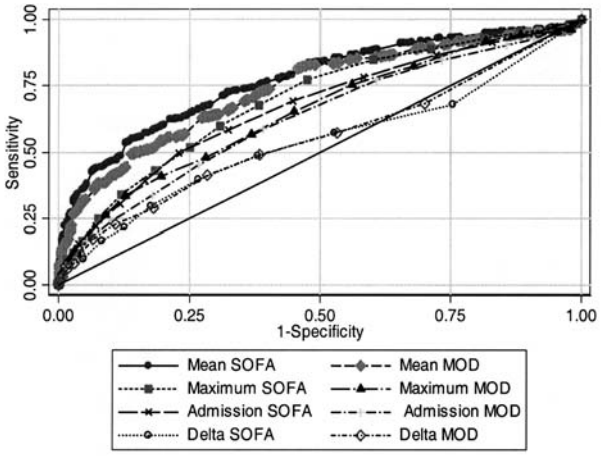

FIGURE 2 Area under the receiver operator characteristic curves: ability to discriminate hospital mortality.

overall score when compared to manual sampling.[18,19] Bosman *et al.* suggest that sampling of score data is different due to a higher frequency of data gathering during automatic data transfer, an absence of errors and the non-existing pre-selection of data by care personnel.[19] The increased density of screening possible with automated collection makes an occurrence of extremely high or low values much more likely and increases the volume of data taken into account for score calculation. Hammond *et al.* found that almost 20% of all health care personnel note vital data incorrectly.[20] Two studies have suggested that there is a tendency for nurses to favour normal readings and ignore extreme values resulting in a reduced variability in manually collected values.[21,22] Because physiological values in our studies were validated by nursing or respiratory staff, our data are likely subject to similar bias but we would expect less so than previous studies as data were not generated de novo by the nurses.

Both systems had only modest ability to discriminate and most models suffered from poor calibration. Discrimination evaluates the ability of a model to distinguish dying patients from those who survive. This is commonly assessed using the AuROC curve. A value of > 0.70 for the AuROC curve is considered satisfactory for discrimination,[23] but too low to be useful for mortality prediction in individual patients.[24] Calibration evaluates the degree of correspondence between the predicted probabilities of mortality and the observed mortality. If calibration is poor, customization of the model can be attempted but is not always successful.[25,26] Further, the practicality of such an approach is questionable.

A practical disadvantage to the MOD score is the calculation of the pressure-adjusted heart rate. Due to the absence of a central venous monitor, this value cannot be calculated in a significant proportion of ICU patients. In the original study describing the MOD score, one half of the patients could not have a cardiovascular component calculated. Unlike the MOD score, the SOFA score has been described in both surgical and medical patients. The two scores are also different with respect to the timing of measurement. The SOFA score requires a review of data over a 24-hr period to identify the most abnormal value. In contrast, the MOD score's use of measurements at one particular time avoids capturing momentary physiological changes unrelated to patient condition, and is perhaps practically easier. However, the ideal choice of data collection time and its effect on validity has not yet been investigated.

Conclusions

Given the emerging use of these scores as surrogate outcomes in clinical trials, the results of this study are timely. In our multicentre, population-based study, these scores had only a modest ability to discriminate between survivors and non-survivors and they incorrectly classified survival status over one quarter of the time. These scores were originally created to follow organ dysfunction over time. We feel the use of these scores as surrogate outcomes is subservient to their use to understand the temporal sequence of MOD and the factors that affect this temporal sequence. The disparate results on the relationship between organ dysfunction and outcome in our study and previously published work suggest that the replacement of mortality as a primary outcome in interventional studies with an organ dysfunction score requires further study prior to adoption.[27]

Acknowledgements

We thank the members of the Department of Critical Care Medicine at the University of Calgary and the Calgary Health Region for their support.

References

1 *Baue AE.* Multiple, progressive, or sequential systems failure. A syndrome of the 1970s (Editorial). Arch Surg 1975; 110: 779–81.
2 *Beal AL, Cerra FB.* Multiple organ failure syndrome in the 1990s. Systemic inflammatory response and organ dysfunction. JAMA 1994; 271: 226–33.
3 *Baue AE.* Multiple organ failure, multiple organ dys-

function syndrome, and systemic inflammatory response syndrome. Why no magic bullets? Arch Surg 1997; 132: 703–7.

4 *Deitch EA.* Multiple organ failure. Pathophysiology and potential future therapy. Ann Surg 1992; 216: 117–34.

5 *Vincent JL, Moreno R, Takala J, 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.

6 *Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ.* Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. Crit Care Med 1995; 23: 1638–52.

7 *Bernard GR, Vincent JL, Laterre PF, et al.* Efficacy and safety of recombinant human activated protein C for severe sepsis. N Engl J Med 2001; 344: 699–709.

8 *Rivers E, Nguyen B, Havstad S, et al.* Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001; 345: 1368–77.

9 *Antonelli M, Moreno R, Vincent JL, et al.* Application of SOFA score to trauma patients. Sequential Organ Failure Assessment. Intensive Care Med 1999; 25: 389–94.

10 *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.

11 *Moreno R, Vincent JL, Matos 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: 686–96.

12 *Metnitz PG, Lang T, Valentin A, Steltzer H, Krenn CG, Le Gall JR.* Evaluation of the logistic organ dysfunction system for the assessment of organ dysfunction and mortality in critically ill patients. Intensive Care Med 2001; 27: 992–8.

13 *Hosmer D, Lemeshow S.* A goodness -of-fit test for the multiple logistic regression model. Communications in Statistics 1980; A10: 1043–69.

14 *Cook R, Cook D, Tilley J, Lee K, Marshall J; Canadian Critical Care Trials Group.* Multiple organ dysfunction: baseline and serial component scores. Crit Care Med 2001; 29: 2046–50.

15 *Cook D, Guyatt G, Marshall J, et al.* A comparison of sucralfate and ranitidine for the prevention of upper gastrointestinal bleeding in patients requiring mechanical ventilation. Canadian Critical Care Trials Group. N Engl J Med 1998; 338: 791–7.

16 *Pettila V, Pettila M, Sarna S, Voutilainen P, Takkunen O.* Comparison of multiple organ dysfunction scores in the prediction of hospital mortality in the critically ill. Crit Care Med 2002; 30: 1705–11.

17 *Peres Bota D, Melot C, Lopes Ferreira F, Nguyen Ba V, 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.

18 *Suistomaa M, Kari A, Ruokonen E, Takala J.* Sampling rate causes bias in APACHE II and SAPS II scores. Intensive Care Med 2000; 26: 1773–8.

19 *Bosman RJ, Oudemane van Straaten HM, Zandstra DF.* The use of intensive care information systems alters outcome prediction. Intensive Care Med 1998; 24: 953–8.

20 *Hammond J, Ward CG, Johnson M, Varas R, Marcial E.* The computerized burn unit: experience with a patient data management system. Int J Clin Monit Comput 1989; 6: 87–9.

21 *Reich DL, Wood RK Jr, Mattar R, et al.* Arterial blood pressure and heart rate discrepancies between handwritten and computerized anesthesia records. Anesth Analg 2000; 91: 612–6.

22 *Taylor DE, Whamond JS.* Reliability of human and machine measurements in patient monitoring. Eur J Intensive Care Med 1975; 1: 53–9.

23 *Lemeshow S, Le Gall JR.* Modeling the severity of illness of ICU patients. A systems update. JAMA 1994; 272: 1049–55.

24 *Randolph AG, Guyatt GH, Richardson WS.* Prognosis in the intensive care unit: finding accurate and useful estimates for counseling patients. Crit Care Med 1998; 26: 767–72.

25 *Markgraf R, Deutschinoff G, Pientka L, Scholten T, Lorenz C.* Performance of the score systems acute physiology and chronic health evaluation II and III at an interdisciplinary intensive care unit, after customization. Crit Care 2001; 5: 31–6.

26 *Moreno R, Apolone G.* Impact of different customization strategies in the performance of a general severity score. Crit Care Med 1997; 25: 2001–8.

27 *Nathens AB, Marshall JC.* Sepsis, SIRS, and MODS: what's in a name? World J Surg 1996; 20: 386–91.