Scoring Systems for Outcome Prediction in a Cardiac Surgical Intensive Care Unit: A Comparative Study

Themistocles Exarchopoulos, RN, MSc, Efstratia Charitidou, MSc, Panagiotis Dedelias, MD, Christos Charitos, MD, Christina Routsi, MD

Background: Most scoring systems used to predict clinical outcome in critical care were not designed for application in cardiac surgery patients.

Objectives: To compare the predictive ability of the most widely used scoring systems (Acute Physiology and Chronic Health Evaluation [APACHE] II, Simplified Acute Physiology Score [SAPS] II, and Sequential Organ Failure Assessment [SOFA]) and of 2 specialized systems (European System for Cardiac Operative Risk Evaluation [EuroSCORE] II and the cardiac surgery score [CASUS]) for clinical outcome in patients after cardiac surgery.

Methods: Consecutive patients admitted to a cardiac surgical intensive care unit (CSICU) were prospectively studied. Data on the preoperative condition, intraoperative parameters, and postoperative course were collected. EuroSCORE II, CASUS, and scores from 3 general severity-scoring systems (APACHE II, SAPS II, and SOFA) were calculated on the first postoperative day. Clinical outcome was defined as 30-day mortality and in-hospital morbidity.

Results: A total of 150 patients were included. Thirty-day mortality was 6%. CASUS was superior in outcome prediction, both in relation to discrimination (area under curve, 0.89) and calibration (Brier score = 0.043, \( \chi^2 = 2.2, P = .89 \)), followed by EuroSCORE II for 30-day mortality (area under curve, 0.87) and SOFA for morbidity (Spearman \( \rho = 0.37 \) and 0.35 for the CSICU length of stay and duration of mechanical ventilation, respectively; Wilcoxon \( W = 3675, P = .03 \) for probability of readmission to CSICU).

Conclusions: CASUS can be recommended as the most reliable and beneficial option for benchmarking and risk stratification in cardiac surgery patients. (American Journal of Critical Care. 2015;24:327-335)
Scoring systems were introduced in both research and clinical practice of intensive care units (ICUs) in order to provide a reliable tool for objectively assessing severity of acute illness in critically ill patients, for better profiling of patients’ risk of mortality, and for improved group stratification and analysis. Most of the scoring systems that are currently in use were initially validated in general ICU populations and not in special units or subgroups of patients. Specifically, cardiac surgery patients have been excluded from the development studies of general predictive scoring systems. Cardiac surgical ICUs (CSICUs) serve a specific population of patients. The acute pathophysiological consequences of cardiopulmonary bypass (CPB), although transient, influence the values of the variables used by the general scoring systems during the early postoperative course. Moreover, several pathophysiological changes may be obscured by the influence of system support devices, such as intra-aortic balloon pumps, ventricular assist devices, hemofiltration, and mechanical ventilation.

Nevertheless, several of the general scoring systems have been subsequently validated and are currently in use in CSICUs, owing to the lack of a suitable and qualified system for this specific population. The most established are the Acute Physiology and Chronic Health Evaluation (APACHE) II, the Simplified Acute Physiology Score (SAPS) II, and the Sequential Organ Failure Assessment (SOFA), with a good reported performance. Originally published in 1999, the European System for Cardiac Operative Risk Evaluation (EuroSCORE), a model for predicting operative mortality following open heart surgery, is based on a large database. It recently has been updated as EuroSCORE II, in order to enhance the performance of the model so as to maintain and optimize its usefulness in contemporary cardiac surgery practice. However, EuroSCORE is limited to preoperative variables and does not take into account intraoperative or postoperative circumstances. Finally, in 2005, the cardiac surgery score (CASUS), specifically for cardiac surgery patients, was developed in order to fill this gap. CASUS has been constructed and validated in a large prospective study in Germany; however, it is not yet widely used. Possible different populations of patients, surgical procedures, and postoperative handling may influence the predictive ability of a scoring system.

We hypothesized that CASUS would perform better than the other scoring systems (even outside of Germany, where it had proven successful), since it was designed to specifically target the peculiarities of cardiac surgery patients. To this end, we conducted the present study to assess and compare the capacity and reliability of CASUS and the recently updated EuroSCORE II as well as the general scoring systems APACHE II, SAPS II, and SOFA for predicting the clinical outcome of patients undergoing open heart surgery in our hospital.

Methods
Study Design and Setting
We conducted a single-center prospective cohort study. All consecutive admissions in the CSICU of Evangelismos Hospital, an 1100-bed, tertiary-care institution, between February 15, 2012, and May 15, 2012, were included in the study. In this 8-bed CSICU, the mean number of new admissions was approximately 3 patients per day for each year.

Patients
Our sample comprised all adult patients who underwent open heart surgery during the study period. Patients were transferred from the operating room directly to the CSICU, where they remained for postoperative monitoring for at least 24 hours; then they were stepped down to the general care area in order to be discharged from the hospital. More specific exclusion criteria were not applied.

Data on preoperative condition, intraoperative parameters, and postoperative course were collected daily from each patient. The postoperative scores

About the Authors
Themistocles Exarchopoulos is a critical care nurse at Mitera Hospital, Marousi, Greece. Efstratia Charitidou is a statistician and PhD student at the National Technical University of Athens, Athens, Greece. Panagiotis Dedelias and Christos Charitos are cardiac surgeons at Evangelismos Hospital, Athens, Greece. Christina Routsi is an associate professor at the Medical School of the University of Athens, Athens, Greece.

Corresponding author: Themistocles Exarchopoulos, RN, MSc. 31-33 Ethnikis Antistaseos Str, Kessariani, Athens 16121, Greece (e-mail: th.exarcho@gmail.com).
were calculated by using the most abnormal value for each variable during the first postoperative 24 hours in the CSICU. EuroSCORE II was based on the logistic version, whereas the 4 postoperative scoring systems were calculated through additive versions. No data were missing.

Clinical outcome was defined as postoperative morbidity and mortality. Morbidity included the following variables: duration of mechanical ventilation, length of stay in the CSICU and in the hospital’s general care area, readmission to the CSICU, and reintubation. Furthermore, we recorded preoperative and intraoperative parameters to examine the potential statistical correlation with the outcome. Mortality was defined as death within 30 days after surgery (either during hospitalization or after discharge).

For patients readmitted to the CSICU postoperatively, we took into account only the length of the initial stay in the CSICU. In cases of reintubation, we evaluated not only the duration of mechanical ventilation, but also the initial duration of mechanical ventilation, in accordance with the rest of the sample.

The study protocol was approved by the scientific council and the bioethics committee of the Evangelismos Hospital (Approval No. 208/01-08-2011). Since it was an observational study and medical confidentiality and personal data were preserved, the requirement for informed consent from each patient or relatives was waived.

**Statistical Analyses**

Statistical analyses were performed by using R Statistical Software, version 2.15.1 (The R Foundation for Statistical Computing). The level of statistical significance was set at 5% (P < .05). Multiple Kolmogorov-Smirnov tests were applied and indicated that all quantitative variables of the study are not normally distributed (P < .001); therefore, we mainly used nonparametric statistical procedures. Qualitative factors are presented by their absolute (N) and relative (%) frequencies. Quantitative characteristics are presented by their mean and standard deviation (SD) when symmetric and by their median and interquartile range (IQR) when severely skewed. In case of quantitative characteristics, the Wilcoxon test was used; otherwise the P value is associated with the Fisher exact test.

The nonparametric Spearman correlation coefficient was used to measure the correlation between quantitative variables. The nonparametric Wilcoxon test was implemented to investigate a potential statistical association between a quantitative variable and a dichotomous factor. The Fisher exact test was used between qualitative factors.

The discriminating capacity of all scoring systems over 30-day mortality prediction was measured via the receiver operating characteristic (ROC) curves, which illustrate the sensitivity (true-positive cases) against 1 minus the specificity (false-positive cases). In order to examine the potential association of each scoring system with 30-day mortality, univariate logistic regression models were fitted. Various aspects of performance and goodness of fit were assessed by several indices, including the Hosmer-Lemeshow test statistic, the Nagelkerke pseudo-$R^2$, the overall correct classification (OCC) index, and the Brier score. The Brier score and the Hosmer-Lemeshow statistic both constitute calibration measures, while the Brier score also measures the discriminative ability of each scoring system; when a system is both correctly calibrated and has good discriminative ability, the value of the Brier score approaches zero. The closer to 100% the pseudo-$R^2$ is, the larger is the contribution of the relevant scoring system to the prediction of the clinical outcome. The OCC index is the ratio of correctly predicted cases (both survivals and deaths) to the total number of patients.

**Results**

**Patients**

The basic characteristics of the study patients are shown in Table 1. They were classified in chronological order and appear in groups of preoperative, intraoperative, and postoperative variables. We recorded basic demographic and clinical parameters along with the values of the scoring systems under study for each patient. P values corresponding to potential statistically significant associations of the sample characteristics with the 30-day mortality are shown.

A total of 150 patients were admitted to the CSICU during the study period. The majority of patients were male (87%). The mean age was 64 (SD, 11; range, 23-84) years. Of these patients, 11% were admitted to the CSICU as emergency cases and 3% had undergone primarily open heart surgery (reoperation). Four patients with chronic renal failure were treated with hemodialysis. The 30-day mortality among the study population was 6%. The median (IQR) length of stay in the CSICU was 46 (27.7) hours, ranging from 24 to 924 hours, and in the hospital was 8 (3.7) days, ranging from 1 to 68 days. The duration of mechanical ventilation was from 4 up to 372 hours (median [IQR], 13 [12] hours).

The duration of the surgical procedure was from 90 up to 560 minutes (mean [SD], 260 [82] min). The majority of patients (n = 129) underwent surgery...
on CPB (median [IQR]: CPB time, 131 [56] minutes; aortic cross-clamp time, 82 [46] minutes; circulatory arrest time, 33 [19.7] minutes). Nine patients of this group underwent an on-pump beating-heart technique. The remaining 21 patients underwent surgery without CPB and with the beating-heart technique.

There were 10 cases of reintubation (7%) due to weaning failure and 9 cases of readmission to the CSICU (6%) caused by various reasons such as cardiac arrest, hemodynamic or respiratory instability, renal dysfunction, and neurological impairment. The types of surgical procedures that were performed on the sample are shown in Table 2.

**Performance of the Scoring Systems**

Figure 1 illustrates the smooth ROC curves concerning the evaluated scoring systems as to 30-day mortality. All curves demonstrate very good discrimination between survivors and nonsurvivors. The area under the ROC curve (AUC) for CASUS was 0.89, greater than those of the other scoring systems; nonetheless, the AUC was high in every case. De Long tests have been conducted in order to compare the differences between AUC pairs (all P values > .05). A summary of the main results is given in Table 3, including details of the ROC analysis and the logistic regression models. Particularly, all scoring systems are significantly associated with the 30-day mortality. CASUS is accompanied by the highest value of Nagelkerke pseudo-$R^2$, whereas EuroSCORE II has the lowest value. OCC is maximized under the SAPS II, but all OCC values are substantially high (> 90%). The value of the Brier score is very close to zero for all scoring systems.
with the lowest values pertaining to CASUS and SAPS II. Finally, the Hosmer-Lemeshow test revealed quite high P values, implying good calibration, with best results (highest P value) concerning CASUS and SAPS II. Therefore, the results from both calibration procedures implemented here seem to coincide.

The correlations among the scoring systems are pictured in the Spearman graphical correlation matrix of Figure 2, where the more intense the positive correlation between 2 variables, the more intense is the shade of grey for the corresponding cell. It is obvious that all the variables are positively correlated. All correlations are statistically significant (P < .01). The highest correlation was found between SOFA and CASUS.

Correlations with Morbidity
As for predictors of in-hospital morbidity, mortality correlated significantly with the duration of mechanical ventilation, reintubation, and CSICU LOS (Table 1). The nonparametric Wilcoxon test was used to identify potential statistical associations with quantitative characteristics. Only CASUS (W = 331.5, P = .005) and APACHE II score (W = 404.0, P = .02) are significantly differentiated by the reintubated group of patients, whereas SOFA score (W = 367.5, P = .03) and CASUS (W = 320.5, P = .01) are significantly differentiated in case of readmission to the CSICU. SAPS II was not correlated with any of the postoperative morbidity factors.

Because most cardiac surgery patients do not experience significant postoperative morbidity, the predictive value of a severity of illness scoring system is inherently limited. Restricting use of this type of predictive system to a population of cardiac surgery patients with complicated postoperative course may add predictive power to the system. EuroSCORE II may be considered a good summary of preoperative and intraoperative risk factors, and SOFA, with its derived variables, may be considered a reliable descriptor of postoperative complications and severity of illness, namely the degree and the progression of postoperative organ dysfunction and failure. We aimed to examine the performance of combined EuroSCORE II with SOFA compared with CASUS. We have modeled the performance of EuroSCORE II and SOFA. The new model (OCC = 88.6%, Brier score = 0.04, R² = 33.9%, Hosmer-Lemeshow P = .79) does not appear to be overall better than CASUS.

Discussion
In this study, we evaluated and compared the accuracy of 5 scoring systems to predict morbidity and mortality in patients undergoing cardiac surgery. All systems showed an acceptable performance; however, CASUS appeared to be superior to the others for predicting mortality in these patients. This superiority was reflected by an AUIC of 0.89. Because CASUS was designed to target cardiac surgery patients specifically, this finding seems plausible and is not unexpected. Despite the fact that several studies have been focused on use of severity scores to predict mortality after cardiac surgery, only a limited number of studies have compared the predictive ability of all the available scoring systems commonly used in both general ICUs and CSICUs.

<table>
<thead>
<tr>
<th>Operation</th>
<th>No. (deaths)</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronary artery bypass graft</td>
<td>82 (3)</td>
<td>54.7</td>
</tr>
<tr>
<td>Isolated valve surgery (aortic valve replacement, mitral valve replacement, tricuspid valve repair)</td>
<td>36 (1)</td>
<td>24.0</td>
</tr>
<tr>
<td>Combined coronary artery bypass graft and valve surgery</td>
<td>16 (0)</td>
<td>10.7</td>
</tr>
<tr>
<td>Ascending aorta and aortic arch surgery</td>
<td>7 (0)</td>
<td>4.7</td>
</tr>
<tr>
<td>Combined ascending aorta and valve surgery—Bentall procedure</td>
<td>6 (2)</td>
<td>4.0</td>
</tr>
<tr>
<td>Combined ascending aorta and coronary surgery</td>
<td>1 (1)</td>
<td>0.7</td>
</tr>
<tr>
<td>Postinfarction ventricular septal rupture closure</td>
<td>2 (2)</td>
<td>1.3</td>
</tr>
<tr>
<td>Total</td>
<td>150 (9)</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 1 Smooth receiver operating characteristic curves for the scoring systems under study: cardiac surgery score (CASUS), Sequential Organ Failure Assessment (SOFA), Acute Physiology and Chronic Health Evaluation (APACHE) II, Simplified Acute Physiology Score (SAPS) II, European System for Cardiac Operative Risk Evaluation (EuroSCORE) II.
Our findings are in accordance with results of a recent study in cardiac surgery patients where CASUS was better for discriminating between survivors and nonsurvivors than were other outcome prediction systems, with an AUC of more than 0.90, followed by SOFA, whereas SAPS II and APACHE II did not perform well in terms of calibration and discrimination statistics.

Apart from the systems specialized for cardiac surgery patients, scores from the general prognostic scoring systems have also shown a significant correlation with 30-day mortality and morbidity. Although these scoring systems were not designed for the specific features of open heart surgery, and they had excluded these patients from their initial development studies, the main and most commonly used scoring systems in the field of intensive care, such as the APACHE II, SAPS II, and SOFA, have been evaluated in this specific subset of patients.

In our sample, SOFA exhibited the least satisfactory performance in relation to the prediction of 30-day mortality, something not unforeseen, because the model was originally designed over time to describe the changing severity in the process of organ failure in ICU patients with sepsis; thus, it was designated more as an indicator of morbidity. In 2003, Ceriani et al examined the use of SOFA scores to predict postoperative morbidity in cardiac surgery patients. That study showed that despite the peculiarities of cardiac surgery, mainly because of the use of CPB, which is an important confounder in evaluating the severity of organ dysfunction, the model was able to describe reliably the evolving course of organ failure after open heart surgery and without any specific adjustments and modifications. Similarly, SOFA score was associated with morbidity in the early postoperative period after cardiac surgery in another study. This conclusion was confirmed.

### Table 3
Summary of test results for SOFA, CASUS, SAPS II, APACHE II, and EuroSCORE II

<table>
<thead>
<tr>
<th></th>
<th>Preoperative day EuroSCORE II</th>
<th>Day 1 in cardiac surgical intensive care unit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SOFA</td>
<td>CASUS</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>1.17</td>
<td>1.95</td>
</tr>
<tr>
<td></td>
<td>1.06-1.28</td>
<td>1.34-2.84</td>
</tr>
<tr>
<td>Overall correct classification, %</td>
<td>95.3</td>
<td>94.7</td>
</tr>
<tr>
<td>Nagelkerke pseudo-$R^2$, %</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.048</td>
<td>0.048</td>
</tr>
<tr>
<td>Hosmer-Lemeshow</td>
<td>8.3</td>
<td>2.9</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>$df$</td>
<td>.42</td>
<td>.57</td>
</tr>
<tr>
<td>$P$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Receiver operating characteristic analysis

<table>
<thead>
<tr>
<th></th>
<th>Area under curve</th>
<th>95% CI</th>
<th>0.76</th>
<th>0.89</th>
<th>0.80</th>
<th>0.82</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best cutoff</td>
<td>4.73</td>
<td>7.5</td>
<td>8.5</td>
<td>43.5</td>
<td>22.5</td>
<td></td>
</tr>
<tr>
<td>Sensitivity, %</td>
<td>88.9</td>
<td>100.0</td>
<td>88.9</td>
<td>55.5</td>
<td>66.6</td>
<td></td>
</tr>
<tr>
<td>Specificity, %</td>
<td>82.3</td>
<td>47.5</td>
<td>73.8</td>
<td>93.6</td>
<td>89.3</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: APACHE, Acute Physiology and Chronic Health Evaluation; CASUS, cardiac surgery score; EuroSCORE, European System for Cardiac Operative Risk Evaluation; SAPS, Simplified Acute Physiology Score; SOFA, Sequential Organ Failure Assessment.

Figure 2 Spearman's graphical correlation matrix among the scoring systems: cardiac surgery score (CASUS), Sequential Organ Failure Assessment (SOFA), Acute Physiology and Chronic Health Evaluation (APACHE) II, Simplified Acute Physiology Score (SAPS) II, European System for Cardiac Operative Risk Evaluation (EuroSCORE) II.
in our study, as in 3 of the 4 variables of in-hospital morbidity (CSICU LOS, mechanical ventilation time, and CSICU readmission) SOFA excelled, compared with the other scoring systems except for CASUS, displaying more than adequate results at predicting prognosis. An exception was observed in the case of potential reintubation, where SOFA score was barely significant whereas APACHE II score was significantly associated with the possibility of a reintubation. APACHE II and SAPS II scores were the postoperative scores that were least correlated with the morbidity determinants.

In the past decade, EuroSCORE has been used to predict both in-hospital mortality and morbidity in numerous cardiac surgery centers worldwide. In a large study including 6222 cardiac surgery patients and comparing 19 scoring systems, the discriminatory power for 30-day mortality was highest for EuroSCORE, almost similarly to the Cleveland Clinic score, developed in the United States. All these systems, initially developed to estimate intraoperative and perioperative death, are based mainly on preoperative general risk factors.

As shown elsewhere, EuroSCORE overestimated mortality, having been already characterized as an outdated scoring system. EuroSCORE II, recently created as a necessary feature for prognostic description of the cardiac surgery mortality risk, has shown remarkable level of discrimination with an AUC of 0.81 in predicting clinical outcome in the development study. Similarly, in the present study, EuroSCORE II exhibited an excellent discrimination concerning 30-day mortality, with an AUC of 0.87. However, contrary to previous findings, it did not appear to be equally reliable for predicting in-hospital morbidity. One possible explanation is that EuroSCORE II exclusively consists of preoperative and intraoperative parameters, whereas postoperative events may have influenced the duration of mechanical ventilation and LOS in the CSICU in our study.

Nonetheless, the performance of the general scoring systems in accordance to discrimination and calibration statistics could not prevail over a specialized cardiac surgery scoring system that took into account the special circumstances encountered in the ICU after cardiac surgery, such as CASUS. Thus the primary research hypothesis for our study seems to be confirmed. CASUS clearly demonstrates superiority over other scoring systems for predicting outcomes, both in previously published studies and in our study.

Our study does have certain limitations. First, the relatively limited size of our sample, due to lack of financial and human resources as well as the single-center character, might limit the generalizability of the findings. Second, scores for the examined systems were calculated only for patients' first 24 hours in the CSICU. Daily calculation of CASUS and of scores from other systems would allow mortality to be predicted more precisely for every additional day of hospitalization in the CSICU, since logistic versions are enhanced with coefficients weighted for the length of stay in the ICU. The fact that the overall length of stay in the CSICU in our case appeared limited, with a median value of 46 hours, prevented us from continuing to collect clinical data and from calculating scores beyond the first 24 hours.

On the other hand, the prospective design of the present study minimizes potential sources of bias and confounding, as well as incomplete data, which are common problems in retrospective design, thus giving the study remarkable strength. Additionally, selection bias was prevented in other ways; for example, our sample consisted of a rather homogeneous population, such as cardiac surgery patients, and mirrors the daily practice in our CSICU, since all the admissions were consecutive. Finally, although several studies comparing the performance of different severity scores in CSICU patients have been published, the present study is among the few studies that tested the CASUS along with EuroSCORE II and the general prognostic scoring systems.

Conclusions

All the examined scoring systems were highly correlated with the clinical outcome of cardiac surgery patients and feature remarkable statistics concerning the predictive power. CASUS transcends and emerges as the most reliable and beneficial option for benchmarking and risk stratification in cardiac surgery patients. EuroSCORE II and SOFA follow in performance at predicting 30-day mortality and in-hospital morbidity, respectively.

ACKNOWLEDGMENTS

We warmly thank the medical and nursing staff of the cardiac surgical intensive care unit at Evangelismos Hospital, Athens, Greece for their generous contributions in clinical data collection for the study.

FINANCIAL DISCLOSURES

None reported.

eLetters

Now that you’ve read the article, create or contribute to an online discussion on this topic. Visit www.ajcconline.org and click “Submit a response” in either the full-text or PDF view of the article.
REFERENCES


To purchase electronic or print reprints, contact American Association of Critical-Care Nurses, 101 Columbia, Aliso Viejo, CA 92656. Phone, (800) 899-1712 or (949) 362-2050 (ext 532); fax, (949) 362-2049; e-mail, reprints@aacn.org.