Acute Physiology and Chronic Health Evaluation (APACHE) III Outcome Prediction After Major Vascular Surgery

Mark T. Keegan, MB, MRCPI,* Francis X. Whalen, MD,* Daniel R. Brown, MD, PhD, FCCM,* Tuhin K. Roy, MD, PhD,* and Bekele Afessa, MD†

Objective: To investigate the performance of the Acute Physiology and Chronic Health Evaluation (APACHE) III scoring system in patients admitted to the intensive care unit (ICU) after major vascular surgery.

Design: Retrospective cohort study.

Setting: A tertiary referral center.

Participants: Three thousand one hundred forty-eight patients who underwent major vascular surgery between October 1994 and March 2006.

Interventions: None.

Measurements and Main Results: Data were abstracted from an institutional APACHE III database. Standardized mortality ratios (SMRs) (with 95% confidence intervals) were calculated. The area under the receiver operating characteristic curve (AUC) and Hosmer-Lemeshow C statistic were used to assess discrimination and calibration, respectively. The mean age of 3,148 patients studied was 70.5 years (±standard deviation 9.6). The mean Acute Physiology Score and the APACHE III score on the day of ICU admission were 31.0 (±17.5) and 45.1 (±18.8), respectively. The mean predicted ICU and hospital mortality rates were 3.2% (±7.8%) and 5.0% (±9.5%), respectively. The median (and interquartile range) ICU and hospital lengths of stay were 4.3 (3.6-5.1) and 14 days (11.9-16.8 days), respectively. The observed ICU mortality rate was 2.4% (75/3,148 patients) and hospital mortality rate was 3.7% (116/3,148). The ICU and hospital SMRs were 0.74 (0.58-0.91) and 0.74 (0.61-0.88), respectively. The AUC of APACHE III–derived prediction of hospital mortality was 0.840 (95% confidence interval, 0.799-0.880), indicating excellent discrimination. The Hosmer-Lemeshow C statistic was 28.492, with a p value <0.01, indicating poor calibration.

Conclusions: The APACHE III scoring system discriminates well between survivors and nonsurvivors after major vascular surgery, but calibration of the model is poor.

© 2008 Elsevier Inc. All rights reserved.

KEY WORDS: vascular surgery, abdominal aortic aneurysm, intensive care unit, Acute Physiology and Chronic Health Evaluation scoring systems, prognosis

METHODS

After institutional review board approval and the granting of a waiver of informed consent, a retrospective study of prospectively collected data was performed. Adult patients who underwent major vascular surgery at the authors’ center between October 1994 and March 2006 were identified and studied. Major vascular surgery was defined as elective and emergent abdominal, descending thoracic and thoracoabdominal aneurysm repair, surgical management of aortic dissections, and aortoiliac and aortofemoral bypass procedures. Patients requiring full cardiopulmonary bypass were excluded as were those in whom authorization for research-related medical record review was unavailable.

Vascular surgical patients who require ICU admission at the authors’ tertiary referral center are managed in a single 20-bed ICU by an anesthesiologist-intensivist–led multidisciplinary critical care team. The critical care service works in conjunction with the vascular surgical services. Since 1994, the APACHE III scoring system has been used in a prospective fashion for all patients in this ICU (except for the small percentage of patients who spend fewer than 4 hours in the ICU). Although a vascular surgical progressive care unit was opened in 2001, patients after major vascular surgery are routinely admitted to the ICU postoperatively. During the period of the study, the vast majority of aortic aneurysm repairs were performed via the open, rather than the endovascular, technique. Aneurysm repair via the endovascular route was introduced at the authors’ institution in late 1999. Although patients after such surgery are now commonly admitted to the progressive care unit rather than the ICU, during the initial years of endovascular surgery, patients after endovascular aortic aneurysm repairs were admitted to the ICU via the vascular surgical intensive care unit rather than the ICU, during the initial years of endovascular surgery.
anastomotic repairs, these patients were primarily admitted to the ICU in the postoperative period. Thus, the authors collected data on the overwhelming majority of major vascular procedures during the study period.

For each of the identified patients, the institutional APACHE III database was searched using the software provided by Cerner Corporation (Kansas City, MO), and APACHE III data for the first ICU day were abstracted. Only data for the first ICU admission of an index hospitalization were collected. Data collected included demographics, the acute physiology score and APACHE III scores, and predictions of ICU and hospital mortality and lengths of stay. Patients' ICU and hospital discharge status (survivor vs nonsurvivor) and discharge location were recorded. Laboratory values used for APACHE III scoring were entered automatically using software that interfaces with the laboratory. Bedside nurses abstracted vital signs, urine output, and Glasgow coma scores according to a formalized protocol, and these were entered into the computer by trained specialists. Experienced clinical ICU nurses performed an audit of the collected data for missing and discrepant admission and physiologic and outcome values. To successfully pass the audit, criteria were set including at least 90% agreement on admission variables overall; 100% agreement on admission and discharge dates; a minimum of 80% agreement on admission diagnosis, admission and discharge times, chronic health items, readmission status, surgical status, and active treatment status; and at least 85% agreement on overall physiology variables. Calculations were performed at the Cerner Corporation after the online transfer of entered data. The authors have previously described the use of APACHE III at their institution.

SPSS 11.5 (SPSS Inc, Chicago, IL) and MedCalc 9.0 (MedCalc Software, Mariakerke, Belgium) were used to perform the data analyses. Descriptive data are summarized as mean (standard deviation), median (interquartile range), or percentage. A Student t test and rank-sum tests were used to compare continuous variables, and a chi-square analysis was used to compare categoric variables. Patients with missing data were excluded from the analyses involving the missing data. Two-tailed statistical tests were used, and tests were considered statistically significant if p was less than 0.05. The 95% confidence intervals (CIs) were calculated for each of the standardized mortality ratios and ICU and hospital length of stay ratios.

Discrimination describes the accuracy of a given prediction. Calculation of the area under the receiver operating characteristic curve (AUC) was used to assess discrimination of the predictive model. An AUC of greater than 0.9 was considered to be outstanding, greater than 0.8 to 0.9 excellent, 0.7 to 0.8 acceptable, and less than 0.7 was considered poor. Calibration describes how the prognostic instrument performs over a wide range of predicted mortalities. A predictive instrument is highly calibrated if it is accurate at predicted mortalities of 20%, 50%, and 90%. Calibration is measured by the Hosmer-Lemeshow statistic. For calculation of the C statistic, observations are sorted into their expected probability, and 10 groups of equal size are formed. A chi-square test (with 8 degrees of freedom) is performed by using observed and expected numbers of deaths in each decile. The value of the chi-square test is the C statistic. A high C statistic is associated with a small p value. This implies significant differences between observed and predicted mortality and a lack of fit of the model. Thus, for the Hosmer-Lemeshow statistic, a small p value (<0.05) indicates poor calibration.

**RESULTS**

There were 3,148 patients who underwent major vascular surgical procedures requiring ICU admission during the period of the study. The characteristics of the patients overall, along with comparison of survivors and nonsurvivors, are described in Table 1. The distribution of the APACHE III scores in the present cohort is shown in Figure 1. The majority of the patients (1,912 patients, 60.7% of the total) underwent abdominal aortic aneurysm repair (AAA). The repair of thoracoabdominal aneurysms accounted for 252 (8%) of the procedures. Five hundred fifty-six patients (17.7%) and 428 patients (13.6%) underwent aortofemoral and aortoiliac bypass, respectively. Of the total number of surgical procedures, elective and emergency surgery accounted for 87.4% and 12.6%, respectively. The median (and interquartile range) ICU and hospital lengths of stay were 4.3 days (3.6-5.1) and 14 days (11.9 to 16.8), respectively.

The observed ICU mortality rate was 2.4% (75/3,148 patients), and the hospital mortality rate was 3.7% (116/3,148 patients admitted to the ICU). Forty-one deaths occurred in patients after discharge from the ICU. These included patients in whom support was withdrawn on the surgical floor or non-ICU ventilator-dependency unit, in whom a decision had been made not to readmit to the ICU, or in whom a sudden decompensation (eg, cardiac arrest) occurred that led to the patient’s death before transfer back to the ICU. The standardized ICU and hospital mortality ratios with their 95% CI were 0.74 (0.58-0.91) and 0.74 (0.61-0.88), respectively.

The AUC of APACHE III–derived prediction of hospital mortality was 0.840 (95% CI, 0.799-0.880). The receiver operator characteristic curve is shown in Figure 2. The Hosmer-Lemeshow C statistic was 28.492, with a p value <0.01. The observed and APACHE III–predicted numbers of hospital survivors and nonsurvivors according to deciles of risk are given in Table 2. For the APACHE III score, the AUC was 0.821 (95% CI, 0.781-0.862). The Hosmer-Lemeshow C statistic was 6.511, with a p value of 0.590.

<table>
<thead>
<tr>
<th>Overall (n = 3,148)</th>
<th>Survivor (n = 3,032)</th>
<th>Nonsurvivor (n = 116)</th>
<th>p Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>2395</td>
<td>2308</td>
<td>87</td>
</tr>
<tr>
<td>Female</td>
<td>753</td>
<td>724</td>
<td>29</td>
</tr>
<tr>
<td>Mean (±SD) age (y)</td>
<td>70.5 (±9.6)</td>
<td>70.3 (±9.6)</td>
<td>76.5 (±8.2)</td>
</tr>
<tr>
<td>Mean Acute Physiology Score</td>
<td>31.0 (±17.5)</td>
<td>30.0 (±15.7)</td>
<td>58.7 (±32.8)</td>
</tr>
<tr>
<td>Mean APACHE III Score</td>
<td>45.1 (±18.8)</td>
<td>43.9 (±16.9)</td>
<td>75.6 (±33.1)</td>
</tr>
<tr>
<td>Predicted ICU mortality (%)</td>
<td>3.2 (±7.8)</td>
<td>2.6 (±5.4)</td>
<td>19.5 (±26.1)</td>
</tr>
<tr>
<td>Predicted hospital mortality</td>
<td>5.0 (±9.5)</td>
<td>4.3 (±7.2)</td>
<td>24.6 (±26.6)</td>
</tr>
<tr>
<td>Elective surgery</td>
<td>2750</td>
<td>2678</td>
<td>72</td>
</tr>
</tbody>
</table>

*p Values are for survivor versus nonsurvivor.
DISCUSSION

The present data show that the APACHE III prognostic scoring system has excellent discriminatory capacity in patients after major vascular surgery. However, the calibration of the model that included the APACHE III–derived prediction of hospital death was poor. Thus, while the scoring system can, on the first ICU day, distinguish survivors to hospital discharge from nonsurvivors in the cohort as a whole, this accuracy is not consistent across all deciles of risk. This problem has been shown by a number of other investigators and is an example of one of the reasons for the need to update prognostic scoring systems.16

Multiple methods have been developed in an attempt to allow clinicians to prognosticate on patients undergoing major vascular surgical procedures.17-26 These scoring systems tend to include only parameters related to pre-existing comorbid conditions of the patients rather than combining premorbid conditions with objective physiologic parameters as used in the APACHE III. A single scoring system has not gained acceptance. Nesi et al24 evaluated five risk scoring systems in 268 patients undergoing elective infrarenal AAA. Although they described the scores as predicting “with reasonable accuracy” the risk of in-hospital death, each AUC was less than 0.8. There are limited data on the performance of APACHE in vascular surgical patients. In their review of 413 patients operated on for ruptured AAA at Mayo Clinic over an 18-year period between 1980 and 1998, Noel et al27 included information on the APACHE II (not III) scores of the patients. As in the present study, the APACHE scores were higher in nonsurvivors when compared with survivors. The authors reported an odds ratio of death of 1.05 for each unit increase in the APACHE II score. Lazarides et al28 studied the APACHE II prognostic

Table 2. The Observed and APACHE III–Derived Predicted Numbers of Hospital Survivors and Nonsurvivors According to Deciles of Risk

<table>
<thead>
<tr>
<th>Decile of Risk</th>
<th>Number of Nonsurvivors</th>
<th>Number of Survivors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Predicted</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>5.88</td>
</tr>
<tr>
<td></td>
<td>312</td>
<td>309.12</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>6.01</td>
</tr>
<tr>
<td></td>
<td>314</td>
<td>308.99</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>6.16</td>
</tr>
<tr>
<td></td>
<td>316</td>
<td>309.84</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>6.29</td>
</tr>
<tr>
<td></td>
<td>313</td>
<td>308.71</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>6.49</td>
</tr>
<tr>
<td></td>
<td>311</td>
<td>308.51</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>6.73</td>
</tr>
<tr>
<td></td>
<td>306</td>
<td>308.21</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>7.24</td>
</tr>
<tr>
<td></td>
<td>308</td>
<td>307.76</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>8.06</td>
</tr>
<tr>
<td></td>
<td>310</td>
<td>306.94</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>10.26</td>
</tr>
<tr>
<td></td>
<td>297</td>
<td>304.75</td>
</tr>
<tr>
<td>10</td>
<td>67</td>
<td>52.83</td>
</tr>
<tr>
<td></td>
<td>245</td>
<td>259.17</td>
</tr>
</tbody>
</table>

NOTE. The first decile is the decile of lowest risk.
system in 40 consecutive patients with ruptured infrarenal aortic aneurysms. APACHE II was “a good predictor of outcome,” although the AUC was not quoted. In a study of 88 patients with ruptured AAAs, APACHE II had limited ability to predict outcome.29 Wolters et al30 examined the predictive ability of APACHE II, POSSUM, and SAPS II on 107 patients undergoing aortic aneurysm surgery. Hadjianastassiou et al13 developed the APACHE-AAA prognostic scoring system in a cohort of approximately 1,900 patients who underwent open AAA repair in the United Kingdom. This scoring system was based on APACHE II (not III), although the authors used variables ascertained in the operating room or at ICU admission in contrast to using the worst values in the first 24 hours in the ICU as originally described for the APACHE II and III. The AUC of the APACHE-AAA (0.845, 95% CI, 0.821-0.868) was marginally better than the AUC of APACHE II prediction but not statistically better than the discriminatory ability of APACHE III in the present cohort.

Outcomes after abdominal aortic surgery are related to the organizational characteristics of ICUs and institutional experience.32-36 It is imperative that any assessment of the quality of care delivered in the ICU involves a consideration of the severity of patient illness using a reliable measure of illness severity.37 The Joint Committee has proposed the severity-adjusted mortality rate as a specific measure that should be recorded.38 Standardized mortality ratios may be calculated as the ratio of the observed mortality to the mortality predicted by a scoring system.39 The adjustment for the severity of illness allows for the monitoring of an ICU’s performance over time and for comparison of ICUs in the same or different hospitals (“benchmarking”). The prediction of outcome for individual patients is theoretically possible, although it is problematic and controversial.16,40

The “holy grail” of scoring systems used in major vascular surgery is a system that would allow the clinician to determine with certainty whether a given patient will survive operative intervention based on preoperative parameters. The ideal system would be simple, require collection of a small number of variables not prone to observer bias, and be amenable to bedside use. Such a system would be especially useful in the setting of ruptured aneurysm because it would guide practitioners toward operative or palliative intervention in a given patient. APACHE III does not meet the criteria required to be an ideal scoring system for use in vascular surgery. APACHE III is an ICU-specific scoring system. Admission to the ICU typically, although not always, occurs after the surgical procedure has been performed. Thus, the scoring system is not helpful in deciding who should be operated on. Furthermore, because the patient has already been admitted to the ICU before the score is calculated, it is not useful for prospectively determining whether a specific patient’s ICU admission is warranted.

In addition to determining who should be operated on, an ideal scoring system for vascular surgery would allow postoperative re-evaluation of patients in whom surgery was performed. Laukontaues41 investigated the predictors of survival after 48 hours of intensive care after the rupture of an AAA. The Sequential Organ Failure Assessment score, a measure of organ dysfunction, was an independent predictor of death. Kniemeyer et al42 showed that the use of a scoring system after 48 hours was appropriate to determine the need for continued ICU care. Although such decision-making may aid in the preservation of limited health care resources, APACHE III predicts mortality based on parameters recorded on the first ICU day, and predictions based on subsequent ICU days have not been validated. Furthermore, prediction of the outcome of individual patients is fraught with difficulty. For example, of 17 patients in the present cohort predicted to have a hospital mortality of greater than 70% on the first ICU day, 6 survived to hospital discharge (data not shown). These results are consistent with the results of the study of the APACHE II by Lazarides et al28 that showed that in patients with ruptured infrarenal AAAs the accuracy of outcome prediction in any given individual patient was limited. Berge et al32 showed that even predictions of mortality of greater than 95% may be erroneous. Although APACHE III is not an ideal scoring system for outcome prediction in vascular surgery, it may be useful in other ways. For example, the practice of admitting patients to the ICU after elective infrarenal AAA repair has been questioned. The introduction of endovascular repair has especially prompted this debate.44 The use of prognostic scoring systems may serve to identify a cohort of patients who, by virtue of their low predicted mortality, might be cared for in a progressive care unit (also known as an intermediate care unit or stepdown unit) rather than in an ICU.45-47

The present study is limited in that it reflects the experience of a single medical center, and the results may not be applicable to other centers. However, the validation of APACHE in a group of patients at a single center may more accurately reflect the performance of the model without the confounding influence of standards of care at different institutions. It is interesting that the model that includes the APACHE III score shows good calibration, but the calibration of the model that includes the APACHE III–derived prediction of mortality is poor. The relatively low number of deaths may have influenced the calibration analysis, perhaps artificially improving the performance of the model that uses the APACHE III score as shown by Kramer and Zimmerman.46

Future investigations could involve customization of the APACHE III model in an effort to improve on either the calibration or discrimination of the model (or both). The addition of variables identified as relevant by some of the other vascular surgery–predictive models would be one method of such a customization. Also, analysis of the performance of APACHE IV (with or without customization) in this patient population would be of considerable interest.

CONCLUSION

The APACHE III scoring system discriminates well between survivors and nonsurvivors in patients admitted to the ICU after major vascular surgery, but calibration of the model is poor. Although APACHE III is not useful for the prediction of individual patient outcome or for the identification of specific patients in whom surgery is appropriate, it may have utility for risk stratification and benchmarking in vascular surgery when used for analysis of cohorts of patients.
REFERENCES


