Does the Surgical Apgar Score Measure Intraoperative Performance?

Scott E. Regenbogen, MD, MPH1,2, R. Todd Lancaster, MD1,2, Stuart R. Lipsitz, ScD3, Caprice C. Greenberg, MD, MPH3, Matthew M. Hutter, MD, MPH2, and Atul A. Gawande, MD, MPH1,3

1 Department of Health Policy and Management, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, USA
2 Department of Surgery, Massachusetts General Hospital, 55 Fruit Street, Boston, Massachusetts 02114, USA
3 Center for Surgery and Public Health, Brigham and Women’s Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA

Abstract

Objective—To evaluate whether Surgical Apgar Scores measure the relationship between intraoperative care and surgical outcomes.

Summary Background Data—With preoperative risk-adjustment now well-developed, the role of intraoperative performance in surgical outcomes may be considered. We previously derived and validated a ten-point Surgical Apgar Score—based on intraoperative blood loss, heart rate, and blood pressure—that effectively predicts major postoperative complications within 30 days of general and vascular surgery. This study evaluates whether the predictive value of this score comes solely from patients’ preoperative risk, or also measures care in the operating room.

Methods—Among a systematic sample of 4,119 general and vascular surgery patients at a major academic hospital, we constructed a detailed risk-prediction model including 27 patient-comorbidity and procedure-complexity variables, and computed patients’ propensity to suffer a major postoperative complication. We evaluated the prognostic value of patients’ Surgical Apgar Scores before and after adjustment for this preoperative risk.

Results—After risk-adjustment, the Surgical Apgar Score remained strongly correlated with postoperative outcomes (p<0.0001). Odds of major complications among average-scoring patients (scores 7–8) were equivalent to preoperative predictions (likelihood ratio (LR) 1.05, 95%CI 0.78–1.41), significantly decreased for those who achieved the best scores of 9–10 (LR 0.52, 95%CI 0.35–0.78), and were significantly poorer for those with low scores—LRs 1.60 (1.12–2.28) for scores 5–6, and 2.80 (1.50–5.21) for scores 0–4.

Conclusions—Even after accounting for fixed preoperative risk—due to patients’ acute condition, comorbidities and/or operative complexity—the Surgical Apgar Score appears to detect differences in intraoperative management that reduce odds of major complications by half, or increase them by nearly three-fold.
Introduction

Hospitals and surgical teams strive to provide a consistently low occurrence of major complications for patients undergoing any given operation. Marked variability in outcomes is inevitable, if only because of differences in patients’ preoperative risks. However, the degree to which intraoperative performance further contributes to variation in patients’ risk of complications remains unclear.1

Prevailing techniques of surgical quality assessment, such as the American College of Surgeons’ National Surgical Quality Improvement Program (NSQIP),2–4 evaluate surgical performance indirectly, using multivariable adjustment for preoperative risk, and attributing disparities between observed and expected complication rates to the care provided. In the operating room, surgeons have relied principally on “gut-feeling” clinical assessments of the operative course to inform postoperative prognostication, and guide clinical care.5 Most believe that intraoperative management contributes importantly to overall outcomes, but quantitative metrics of operative care have not been available.1 Among intraoperative factors, alterations of patient condition, including hypotension,6–23 hypertension,12, 15–18, 22, 24 hypothermia,25–27 bradycardia,20, 22 tachycardia,11, 12, 20, 22, 24, 28–30 and blood loss31–35 have been independently linked with adverse outcomes. And some risk prediction methods have integrated intraoperative variables,32, 36–38 yet no consensus has been reached on how to directly evaluate performance and safety in the operating room.39

To provide surgeons with a simple, objective, and direct rating, we previously developed and validated a ten-point Surgical Apgar Score.40 In deriving the score, we screened more than two dozen parameters collected in the operating room, and found that just three intraoperative variables remained independently predictive of major postoperative complications and death—the lowest heart rate, lowest mean arterial pressure, and estimated blood loss. A score built from these three predictors has proved strongly predictive of the risk of major postoperative complications and death in general and vascular surgery.40 Yet, it remains simple enough for teams to collect immediately upon completion of an operation for patients in any setting, regardless of resource and technological capacity.

Like the obstetrical Apgar score,41–44 however, it provides a measure only of the relative success of care. It cannot by itself assess the quality of care, as its three variables are influenced not only by the performance of medical teams, but also by the patients’ prior condition and the magnitude of the operations they undergo.18, 22, 45 For the score to be a clinically useful predictor of postoperative complications, it should inform operative teams about their contribution to surgical outcomes, even after accounting for fixed preoperative risk—an insight not previously available. In this study, we therefore evaluated the predictive ability of the score after application of a validated risk-adjustment method, incorporating both patient- and procedure-related risk characteristics.

Methods

Patient cohort

The Massachusetts General Hospital (MGH) Department of Surgery maintains an outcomes database on a systematic sample of patients undergoing general and vascular surgical procedures, for submission to the NSQIP. In this program,2, 3 trained nurse-reviewers retrospectively collect 49 preoperative, 17 intraoperative, and 33 outcome variables on surgical patients, for the monitoring of risk-adjusted outcomes. Patients undergoing general or vascular surgery with general, epidural, or spinal anesthesia, or specified operations (carotid endarterectomy, inguinal herniorrhaphy, thyroidectomy, parathyroidectomy, breast biopsy, and endovascular repair of abdominal aortic aneurysm) regardless of anesthetic type, are
eligible for inclusion. Children under age 16 and patients undergoing trauma surgery, transplant surgery, vascular access surgery, or endoscopic-only procedures are excluded. At MGH, at least forty consecutive operations meeting inclusion criteria in each eight-day cycle are enrolled. No more than five inguinal herniorrhaphies and five breast biopsies are enrolled per eight-day cycle to ensure diversity of operations in the case mix.

We evaluated all patients in the MGH-NSQIP database who underwent surgery between July 1, 2003, and June 30, 2005, and for whom complete 30-day follow-up was obtained. We excluded (i) carotid endarterectomies performed concurrently with coronary artery bypass grafting, because the score was not designed for application to patients on cardiopulmonary bypass; and (ii) operations performed with local anesthesia only, because no electronic anesthesia record is generated for these procedures.

The study protocol, including a waiver of informed consent from individual patients, was approved by the Human Subjects Research Committees of Massachusetts General Hospital and the Harvard School of Public Health.

Preoperative risk factors and postoperative outcomes

We collected all preoperative patient variables from the NSQIP database. All variables were either treated as dichotomous or categorized according the FY2005 NSQIP models.46 Missing laboratory values were imputed with the overall sample median (because patients for whom preoperative laboratory data were not obtained were typically low-risk). Procedural work Relative Value Units were calculated by linkage of Current Procedural Terminology codes with listings from the 2005 Medicare Physician Fee Schedule (Centers for Medicare and Medicaid Services).

The primary endpoint was the occurrence of any major complication within 30 days after surgery, as recorded in the NSQIP database. The following NSQIP-defined3 events were considered major complications: acute renal failure, bleeding requiring ≥4 units of red cell transfusion within 72 hours after surgery, cardiac arrest requiring CPR, coma for ≥24 hours, deep venous thrombosis, myocardial infarction, unplanned intubation, ventilator use for ≥48 hours, pneumonia, pulmonary embolism, stroke, wound disruption, deep or organ-space surgical site infection, sepsis, septic shock, systemic inflammatory response syndrome (SIRS), and vascular graft failure. All deaths were assumed to include a major complication. Superficial surgical site infection and urinary tract infection were not considered major complications. Patients having complications categorized in the database as “other occurrence” were reviewed individually and severity of the occurrence was evaluated according to the Clavien classification.47 “Other occurrences” involving complications of Clavien Class III and greater (those that require surgical, endoscopic or radiologic intervention or intensive care admission, or are life-threatening) were considered major complications, in accordance with our previous methods.40

Preoperative risk stratification

To estimate each patient’s preoperative likelihood of complications, we performed multivariable logistic regression using the variables included in the FY2005 NSQIP morbidity risk-adjustment model46 as predictors, and the occurrence of any major complications as the outcome. We derived de novo regression coefficients from our dataset and computed the predicted likelihood of major complication from these regression parameters for each operation. These preoperative likelihoods were then stratified by quintiles for tabulation.48
Calculation of the intraoperative score

As described previously, we originally devised the Surgical Apgar Score by using multivariable logistic regression to screen a collection of intraoperative measures. We found that only three intraoperative parameters remained independent predictors of 30-day major complications: the estimated blood loss (EBL), the lowest heart rate (HR), and the lowest mean arterial pressure (MAP) during the operation. The score was thus developed using these three variables, and their beta coefficients were used to weight the points allocated to each variable in a ten-point score (Table 1).

In this study, we extracted intraoperative hemodynamic data from the electronic Anesthesia Information Management System (Saturn, Dräger Medical, Telford, PA) database, using a Structured Query Language algorithm to filter out artifactual readings, using criteria developed through comparisons of electronic and hand-written intraoperative records. For data quality assurance, we manually reviewed the printed electronic anesthesia record for 50 operations, and compared the results with those of the electronic data acquisition algorithm for these cases. The parameter values, as well as the total Score obtained, by each method were compared by computing kappa statistics for agreement, using Fleiss-Cohen weighting for ordered categorical data.

Statistical analysis

All analyses were performed using the SAS 9.1 statistical software package (SAS Institute, Cary, N.C., 2003). We evaluated relationships between patient and procedure characteristics and levels of the Surgical Apgar Score using Spearman correlation coefficients for continuous variables, and Cochran-Armitage chi-square trend tests for categorical predictors. Preoperative risk stratification was compared with the Surgical Apgar Score using Spearman correlation coefficients and kappa statistics, with Fleiss-Cohen weighting. We calculated c-statistics for model discrimination (equivalent to the area under the ROC curve) and compared models with the Hanley-McNeil z-test. We used the Hosmer-Lemeshow goodness-of-fit test to assess calibration.

Within quintiles of preoperative risk, the relationship between Surgical Apgar Score groups and postoperative occurrences was evaluated with the Cochran-Armitage chi-square trend test. Controlling for preoperative risk predictions as a linear variable, we performed logistic regression with the Surgical Apgar Score as a categorical predictor and the incidence of major complications as the outcome, to compute adjusted effect sizes for each level of the Surgical Apgar Score. Adjusted likelihood ratios (LRs) were computed as the proportional change in odds of complications, comparing prior odds (preoperative prediction) with postoperative odds. We obtained confidence intervals for the LRs using the Bonferroni inequality.

To evaluate the robustness of the risk-adjusted relationship between Surgical Apgar Scores and postoperative complications, we conducted sensitivity analyses, modeling the preoperative risk-adjustment parameters in a variety of ways. For simplicity, the primary analyses we present are stratification by quintiles and logistic regression with the preoperative predictions treated linearly.

Results

Data accrual and validation

Of 4,163 NSQIP cases that met inclusion criteria, 4,119 (98.9%) had complete electronic intraoperative records and comprised our final cohort. The automated data extraction algorithm achieved excellent agreement with manual record review, both for point values assigned to each variable (κ = 0.97 for HR; κ = 0.75 for MAP), and for the total Score (κ = 0.94).
Baseline patient and procedure characteristics

In univariate analyses, most demographic characteristics and preoperative risk factors varied significantly between levels of the Surgical Apgar Score (see Table 2). With decreasing scores, patients were increasingly older (p=0.009), and more likely to be male (p=0.0004) and of non-white race (p=0.04). Patients assigned higher American Society of Anesthesiologists’ (ASA) Physical Status Classification had significantly lower Surgical Apgar Scores (Spearman’s r = −0.24, p<0.0001). Two-thirds (85 of 128) of patients with scores ≤4 were ASA Class 3 or 4, whereas three quarters (1091 of 1441) of patients with scores of 9 or 10 were ASA Class 1 or 2 (p<0.0001).

Low-scoring patients were significantly more likely to be underweight (p=0.01), but not more likely to be obese (p=0.12). Among the 26 other preoperative comorbidity conditions, 22 of them were increasingly prevalent as patients’ scores decreased. Only hypertension (p=0.06), coma (p=0.89), Do Not Resuscitate status (p=0.15), and alcohol use (p=0.53) were not significantly correlated with Surgical Apgar Scores. Abnormalities in all 12 preoperative laboratory measures were also increasingly more common as patients’ scores decreased (all p<0.01). Operations with lower scores had increasing complexity (as measured by Work RVUs) and were more likely to be emergencies (both p<0.0001).

Surgical Apgar Scores were also predictive of postoperative outcomes. The incidence of major postoperative complications increased monotonically from 5% among patients with scores of 9–10, to 56% of those with scores ≤4 (p<0.0001). Patients with low scores were more likely to suffer multiple complications (p<0.0001), and had significantly longer median length of stay (p<0.0001). Among patients who experienced a complication, the likelihood of dying from that complication was nearly 20-fold greater for patients with scores 0–2 than for those with scores of 9–10 (p<0.0001).

Preoperative risk-adjustment

Logistic regression, using the 27 preoperative NSQIP variables as predictors and the incidence of major postoperative complications as the outcome, generated a multivariable preoperative risk prediction model with a c-index of 0.820 (equivalent to that of the 34,000 patient FY2005 NSQIP cohort; p=0.23). The Hosmer-Lemeshow chi-square statistic demonstrated adequate model calibration (p=0.49).

Forty percent of patients were missing at least one of the laboratory measures required for the model, ranging from 3.5% missing white blood cell count to 37% missing albumin. In sensitivity analyses, results from imputation with the sample median were not meaningfully different from those of multiple imputation, so median imputation was used for simplicity.

Patients were stratified into preoperative risk quintiles, based on their predicted likelihoods of major complication according to this model. Quintile 1 included patients with preoperative risk ≤3.8%; Quintile 2, 3.8–6.5%; Quintile 3, 6.5–10.6%; Quintile 4, 10.6–19.2%; Quintile 5, ≥19.2%. In logistic regression, discrimination by quintiles (c=0.795) was not significantly different from that of the saturated risk prediction model (p=0.12).

Risk-adjusted analysis of the Surgical Apgar Score

Patients’ preoperative risk predictions and Surgical Apgar Scores were negatively correlated (r = −0.42, p<0.0001), confirming that the elements of the score are associated with preoperative risk factors. Accordingly, there was fair agreement between a patient’s preoperative risk quintile and level of the score, with a weighted kappa of 0.24 (95% confidence interval 0.22–0.26).
The stratified data in Table 3 demonstrate this relationship. Of the 128 patients with scores ≤4, 112 (88%) came from the highest two risk quintiles—patients with preoperative likelihood of complication greater than 10.6%. In contrast, patients from the two lowest risk quintiles—with estimated preoperative risks less than 6.5%—comprised 59% of patients with scores of 9–10.

Despite concordance between preoperative factors and intraoperative metrics, after accounting for preoperative risk, the Surgical Apgar Score remained a significant predictor of postoperative complications. Within each quintile, patients with scores of 7–8 experienced complication rates similar to the expected mean rate for their stratum. Patients with scores of 9–10 had consistently lower incidence, and patients with scores <7 had consistently greater incidence of major postoperative complications than was expected preoperatively. Among the three highest risk strata, Surgical Apgar Scores remained significantly predictive of postoperative outcomes (each p<0.001). In the two lowest quintiles, however, we had limited power to detect a significant effect, because of the rarity of both complications and low scores among these low-risk patients. Nevertheless, a trend toward significant relationship was found (both p<0.10).

Adjusted and unadjusted likelihood ratios (LRs), representing the proportional change in odds of a major complication comparing preoperative expectations with postoperative predictions, are shown in Table 4. Unadjusted LRs compare posterior odds of a complication for patients within each score category against the average preoperative odds for the entire cohort. The adjusted LRs estimate the degree to which the Surgical Apgar Score alters any given patient’s odds of complication, after accounting for the patient’s fixed preoperative odds as measured by NSQIP risk factors.

Patients with scores of 7–8 had postoperative outcomes no different from preoperative predictions (LR 1.05, 95% confidence interval 0.78–1.41). Those with scores of 9–10, however, had significantly lower odds of complication than would be expected based on preoperative risk (LR 0.52, 95% confidence interval 0.35–0.78), and those with scores less ≤6 had significantly increased posterior odds, with LR of 1.60 (95% confidence interval 1.12–2.28) for scores 5–6, 2.79 (1.47–5.31) for scores 3–4. Confidence intervals around the LR for scores 0–2 (2.87, 95% confidence interval 0.49–16.76) crossed one due to small sample size.

**Sensitivity analyses**

We tested eight additional methods of modeling patients’ preoperative risk: (i) adding the Surgical Apgar Score to the 27-variable NSQIP morbidity model; (ii) modeling predictions as a linear plus quadratic and/or (iii) cubic term; (iv) stratifying by deciles; (v) stratifying by quintiles or (vi) deciles of patients with complications only; (vii) one-to-one matching of patients with and without complications by their preoperative risk score; and (viii) construction of our own risk-adjustment model, using all available patient-comorbidity and procedure-complexity variables available to us, from either the NSQIP record, or other data collection, and including all significant interaction terms. Regardless of the risk-adjustment technique, the Surgical Apgar Score remained a strong predictor of postoperative outcomes (p values all <0.0001), and all point estimates for the odds ratios at each score level remained within the 95% confidence bounds of the primary analysis.

**Conclusion**

We find that even after detailed adjustment for comorbidity and procedure-specific risk factors, the amount of blood loss, lowest heart rate and lowest blood pressure were still important predictors of the risk of a major complication. The Surgical Apgar Score, therefore, conveyed useful prognostic information, either in isolation or in combination with assessments of the risks that patients brought to the operating room. It also may provide an immediate assessment.
of how well or poorly the operation has gone for a patient. In this cohort, surgical teams could cut a patient’s risk-adjusted odds of major complications nearly in half with a score of 9–10, or conversely, nearly triple the risk-adjusted odds with scores ≤4.

This finding, that intraoperative blood loss, heart rate, and blood pressure are critical predictors of postoperative risk, is consistent with a variety of previous observations. Hemodynamic stability,6–23, 29, 30 and the amount of blood loss31–35 during surgery have long been recognized as important independent factors in patient outcomes. What had not been recognized were the collective importance of these variables, and their potential contribution to an easily-implemented intraoperative performance metric.40

As an adjunct to surgeons’ subjective impressions of the operation,5 the score may thus aid decision-making about unplanned admission after outpatient surgery, admission to the intensive care unit, or frequency of postoperative examinations by physicians and nurses, with the goal of preventing poor outcomes among low-scoring patients. More broadly, the Surgical Apgar Score provides a novel metric for evaluating the efficacy of safety interventions in the operating room—a much-needed tool for surgical safety initiatives, because more than two-thirds of surgical adverse events involve complications in the operating room.58–60

In this study, we use likelihood ratios to quantify the degree to which a patient’s Surgical Apgar Score alters his or her likelihood of complications. The likelihood ratio describes the proportional change in odds of complication, comparing information available before the operation with what is available after the operation. The unadjusted likelihood ratios associated with levels of the score give a measure of the change in postoperative odds of complications, compared with the baseline rate of complications for the sample as a whole. This type of comparison could be used as part of a broad-based public health audit for surgical safety. Targeting low scores would allow surgeons and administrators to focus on patients coming out of surgery who are at highest risk of major complications or death. Routine surveillance and case-review for patients with low surgical scores (e.g., a score of 4 or less), even when no complications result, may enable early identification of latent safety problems. The score could also provide a target for surgical teams and researchers aiming to improve outcomes, and a measure for quality monitoring and improvement programs, even in resource-poor settings. The ultimate goal would be to encourage development and implementation of practices that reduce the proportion of patients with low scores and increase the proportion with the highest scores.

The risk-adjusted likelihood ratios provide different, yet complementary information, allowing individual surgeons to objectively discern whether, and how much, their operation increased or decreased a patient’s predicted risk of major complications. As Table 4 indicates, an operation with a score of 7 or 8 has not altered the expected risk; one with a score of 9 or 10 has reduced it by half; one with a score of 5–6 has increased the odds of complications by approximately 60%; and one with a score of 4 or less has increased it by almost 200%.

Even in hospitals such as ours, where preoperative risk information is avidly collected for outcomes monitoring in the NSQIP, detailed preoperative risk predictions are rarely, if ever, available at the time of surgery. Missing data (especially laboratory data) and the computational complexity of multivariable prediction models preclude their routine use.37, 45, 61 The majority of surgeons in our institution and elsewhere depend instead on their subjective impressions,5, 62 and rate patients in broader categories, similar to the risk quintiles we analyzed. For these surgeons, the data in Table 3 provide a means for using the Surgical Apgar Score as additional prognostic information, beyond their preoperative expectations.

There are, however, other possible explanations for these findings. Because this was an observational study, it is possible that alterations in intraoperative hemodynamics simply
represent residual confounding—still unmeasured aspects of patient- or procedure-related risk. The sensitivity analyses argue against this being the only explanation, however. Relationships were similar across nine different modeling techniques, and effect sizes for the score were not attenuated by the inclusion of additional predictors to the model. Another explanation could be that the stress of surgery unmasks debility or risk not otherwise measured among preoperative variables. As seen in Table 2, the Surgical Apgar Score is closely associated with many of the same variables that predict postoperative complications. Still, the critical variables for the Surgical Apgar Score are measures that have been consistently recognized as important independent contributors to surgical morbidity. It is reasonable to believe, therefore, that interventions that produce measurable improvement in Surgical Apgar Scores will also improve intraoperative safety and reduce postoperative complications.

Other important limitations remain. The score has not been evaluated beyond major academic medical centers, adult patients, or general and vascular surgery, due to a lack of reliable and comprehensive outcomes assessment against which the measures could be validated in these areas. Whether the score will be effective at grading risk in trauma, pediatric surgery or other surgical specialties remains uncertain.

In summary, we have found that a simple clinimetric surgical outcome score can provide both clinical surgeons and surgical safety researchers with useful and important information. The Surgical Apgar Score integrates components of patient susceptibility, procedure complexity and operative performance, providing a measure of immediate postoperative condition and prognostication, beyond standard risk-adjustment. As a decision-support tool, the score can inform postoperative prognostication, communication and triage, regardless of the sophistication of preoperative risk stratification available. And as a simple intraoperative outcome measure and safety improvement metric, it may prove useful as an indicator of surgical performance.

Acknowledgements

We are indebted to Dr. John Walsh for assistance with the intraoperative anesthesia record, Dr. Jesse Ehrenfeld, for design and implementation of the electronic intraoperative data query methods, and to Ms. Lynn Devaney for assistance with the MGH-NSQIP database.

References

31. Gatch WD, Little WD. Amount of blood lost during some of the more common operations. JAMA 1924;83:1075–1076.


## Table 1

### The Ten-Point Surgical Apgar Score

The Surgical Apgar Score is calculated at the end of any general or vascular surgery operation, from the estimated blood loss, lowest mean arterial pressure and lowest heart rate entered in the anaesthesia record during the operation. The score is the sum of the points from each category.

<table>
<thead>
<tr>
<th></th>
<th>0 points</th>
<th>1 point</th>
<th>2 points</th>
<th>3 points</th>
<th>4 points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated blood loss (mL)</td>
<td>&gt;1000</td>
<td>601–1000</td>
<td>101–600</td>
<td>≤100</td>
<td></td>
</tr>
<tr>
<td>Lowest mean arterial pressure (mm Hg)</td>
<td>&lt;40</td>
<td>40–54</td>
<td>55–69</td>
<td>≥70</td>
<td></td>
</tr>
<tr>
<td>Lowest heart rate (beats per min)</td>
<td>&gt;85*</td>
<td>76–85</td>
<td>66–75</td>
<td>56–65</td>
<td>≤55*</td>
</tr>
</tbody>
</table>

*Occurrence of pathologic bradyarrhythmia, including sinus arrest, atrioventricular block or dissociation, junctional or ventricular escape rhythms, and asystole also receive 0 pts for lowest heart rate.*
## Table 2

Characteristics of patients, procedures, and outcomes, by Surgical Apgar Score

<table>
<thead>
<tr>
<th>Score</th>
<th>0–2 N=14</th>
<th>3–4 N=112</th>
<th>5–6 N=720</th>
<th>7–8 N=1830</th>
<th>9–10 N=1441</th>
<th>Total N=4119</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years) (mean ± sd)</td>
<td>62 ± 13</td>
<td>60 ± 13</td>
<td>59 ± 17</td>
<td>57 ± 15</td>
<td>58 ± 16</td>
<td>58 ± 17</td>
<td>0.01</td>
</tr>
<tr>
<td>Male sex</td>
<td>11 (69)</td>
<td>54 (47)</td>
<td>343 (48)</td>
<td>662 (38)</td>
<td>599 (42)</td>
<td>1710 (47)</td>
<td>0.0004</td>
</tr>
<tr>
<td>Non-white race</td>
<td>2 (13)</td>
<td>12 (11)</td>
<td>69 (10)</td>
<td>171 (9)</td>
<td>108 (7)</td>
<td>362 (9)</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>ASA Class</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>1 (0.9)</td>
<td>27 (4)</td>
<td>181 (10)</td>
<td>216 (15)</td>
<td>425 (10)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>3 (19)</td>
<td>39 (35)</td>
<td>334 (46)</td>
<td>1053 (58)</td>
<td>875 (61)</td>
<td>2304 (56)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3</td>
<td>44 (45)</td>
<td>50 (45)</td>
<td>304 (42)</td>
<td>558 (30)</td>
<td>334 (23)</td>
<td>1253 (30)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4</td>
<td>6 (38)</td>
<td>22 (20)</td>
<td>55 (8)</td>
<td>38 (2)</td>
<td>16 (1)</td>
<td>137 (3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Body mass index (mean ± sd)</td>
<td>60 ± 18</td>
<td>58 ± 17</td>
<td>59 ± 16</td>
<td>57 ± 15</td>
<td>58 ± 16</td>
<td>58 ± 17</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Obese (BMI&gt;35)</strong></td>
<td>4 (6)</td>
<td>11 (10)</td>
<td>123 (17)</td>
<td>350 (19)</td>
<td>293 (14)</td>
<td>643 (15)</td>
<td>0.79</td>
</tr>
</tbody>
</table>

**Laboratory data**

<table>
<thead>
<tr>
<th>Value</th>
<th>0–2 N=14</th>
<th>3–4 N=112</th>
<th>5–6 N=720</th>
<th>7–8 N=1830</th>
<th>9–10 N=1441</th>
<th>Total N=4119</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemoglobin &lt;13g/dL</td>
<td>3 (3)</td>
<td>13 (11)</td>
<td>82 (11)</td>
<td>126 (7)</td>
<td>70 (5)</td>
<td>295 (7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>WBC count &gt;11,000/mm³</td>
<td>9 (6)</td>
<td>37 (33)</td>
<td>142 (20)</td>
<td>230 (13)</td>
<td>177 (9)</td>
<td>545 (13)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Prothrombin time &gt;13.67 seconds</td>
<td>12 (75)</td>
<td>42 (38)</td>
<td>120 (17)</td>
<td>190 (10)</td>
<td>106 (7)</td>
<td>470 (11)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sodium &lt;135 mEq/L</td>
<td>5 (3)</td>
<td>22 (20)</td>
<td>70 (10)</td>
<td>96 (5)</td>
<td>42 (3)</td>
<td>235 (6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Blood urea nitrogen &gt;40 mg/dL</td>
<td>5 (31)</td>
<td>15 (13)</td>
<td>82 (11)</td>
<td>126 (7)</td>
<td>70 (5)</td>
<td>295 (7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Creatinine &gt;1.2 mg/dl</td>
<td>6 (38)</td>
<td>32 (29)</td>
<td>127 (18)</td>
<td>229 (13)</td>
<td>189 (13)</td>
<td>583 (14)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Score</td>
<td>0-2 N=16</td>
<td>3-4 N=112</td>
<td>5-6 N=720</td>
<td>7-8 N=1830</td>
<td>9-10 N=1441</td>
<td>Total N=4119</td>
<td>p value</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td>Albumin g/dL (mean ± sd)</td>
<td>2.7 ± 0.9</td>
<td>2.9 ± 0.9</td>
<td>3.4 ± 0.8</td>
<td>3.8 ± 0.7</td>
<td>4.0 ± 0.5</td>
<td>3.7 ± 0.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>SGOT &gt;40 units/L</td>
<td>6 (38)</td>
<td>22 (20)</td>
<td>107 (15)</td>
<td>173 (9)</td>
<td>91 (6)</td>
<td>399 (10)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Bilirubin &gt;1 g/dL</td>
<td>7 (44)</td>
<td>27 (24)</td>
<td>82 (11)</td>
<td>117 (6)</td>
<td>79 (5)</td>
<td>301 (7)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Alkaline phosphatase &gt;125 units/L</td>
<td>9 (56)</td>
<td>26 (23)</td>
<td>114 (16)</td>
<td>164 (9)</td>
<td>79 (5)</td>
<td>392 (10)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Procedure Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emergency operation</td>
<td>10 (63)</td>
<td>29 (26)</td>
<td>89 (12)</td>
<td>128 (7)</td>
<td>57 (4)</td>
<td>313 (8)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Work Relative Value Units (mean ± sd)</td>
<td>22 ± 19</td>
<td>25 ± 12</td>
<td>24 ± 13</td>
<td>18 ± 10</td>
<td>14 ± 7</td>
<td>18 ± 11</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Postoperative Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major complication(s)</td>
<td>12 (75)</td>
<td>60 (54)</td>
<td>201 (28)</td>
<td>236 (13)</td>
<td>72 (5)</td>
<td>581 (14)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Number of major complications</td>
<td>1</td>
<td>2 (13)</td>
<td>21 (19)</td>
<td>83 (12)</td>
<td>132 (7)</td>
<td>49 (3)</td>
<td>287 (7)</td>
</tr>
<tr>
<td>2-3</td>
<td>8 (50)</td>
<td>20 (18)</td>
<td>75 (10)</td>
<td>76 (4)</td>
<td>18 (1)</td>
<td>197 (5)</td>
<td></td>
</tr>
<tr>
<td>&gt;3</td>
<td>2 (13)</td>
<td>19 (17)</td>
<td>43 (6)</td>
<td>28 (2)</td>
<td>5 (0.3)</td>
<td>97 (2)</td>
<td></td>
</tr>
<tr>
<td>Deaths</td>
<td>7 (44)</td>
<td>18 (16)</td>
<td>33 (5)</td>
<td>34 (2)</td>
<td>2 (0.1)</td>
<td>94 (2)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Mortality rate among patients with major complications</td>
<td>58%</td>
<td>30%</td>
<td>16%</td>
<td>14%</td>
<td>3%</td>
<td>16%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Length of stay (days; median, IQR)</td>
<td>15 (5-25)</td>
<td>8 (6-17)</td>
<td>6 (3-9)</td>
<td>3 (1-5)</td>
<td>1 (0-2)</td>
<td>2 (1-5)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* Hypothesis testing employed Cochrane-Armitage chi-square trend tests for categorical variables, and Spearman correlation coefficients for continuous variables, except for length of stay, for which we used the Kruskal-Wallis test.
Table 3

Postoperative outcomes, by preoperative risk stratum and Surgical Apgar Score

Using a detailed preoperative risk prediction model, patients were stratified by quintiles of likelihood of a major postoperative complication. The predicted and observed incidence of major postoperative complications are presented.

<table>
<thead>
<tr>
<th>Preoperative Risk Predictions</th>
<th>0–2</th>
<th>3–4</th>
<th>5–6</th>
<th>7–8</th>
<th>9–10</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quintile</td>
<td>Mean Risk</td>
<td>Range</td>
<td>N</td>
<td>% Major Complication</td>
<td>N</td>
<td>% Major Complication</td>
</tr>
<tr>
<td>1</td>
<td>1.9%</td>
<td>0.5–3.8%</td>
<td>0</td>
<td>—</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>5.2%</td>
<td>3.8–6.5%</td>
<td>0</td>
<td>—</td>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>8.4%</td>
<td>6.5–10.6%</td>
<td>0</td>
<td>—</td>
<td>14</td>
<td>28.6%</td>
</tr>
<tr>
<td>4</td>
<td>14.1%</td>
<td>10.6–19.2%</td>
<td>2</td>
<td>100%</td>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td>5</td>
<td>41.0%</td>
<td>19.2–99.7%</td>
<td>2</td>
<td>71.4%</td>
<td>7</td>
<td>94.6%</td>
</tr>
<tr>
<td>Total</td>
<td>14.1%</td>
<td>0.5–99.7%</td>
<td>16</td>
<td>75.0%</td>
<td>11</td>
<td>53.6%</td>
</tr>
</tbody>
</table>

*p value

Cochrane-Armitage chi-square trend test

Ann Surg. Author manuscript; available in PMC 2008 October 7.
Table 4

Complication rates and adjusted and unadjusted likelihood ratios for levels of the Surgical Apgar Score

<table>
<thead>
<tr>
<th>Model</th>
<th>SURGICAL APGAR SCORE</th>
<th>Complication Rate</th>
<th>Unadjusted Likelihood Ratio (95% CI)</th>
<th>Adjusted Likelihood Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-2</td>
<td>16</td>
<td>11</td>
<td>0.32 (0.27–0.42)</td>
<td>0.52 (0.35–0.78)</td>
</tr>
<tr>
<td>3-4</td>
<td>112</td>
<td>72</td>
<td>2.36 (2.15–2.84)</td>
<td>1.60 (1.12–2.28)</td>
</tr>
<tr>
<td>5-6</td>
<td>720</td>
<td>271</td>
<td>0.90 (0.83–1.05)</td>
<td>1.06 (0.78–1.41)</td>
</tr>
<tr>
<td>7-8</td>
<td>1830</td>
<td>189</td>
<td>18.27 (5.33–107.40)</td>
<td>2.87 (0.49–16.76)</td>
</tr>
<tr>
<td>9-10</td>
<td>1441</td>
<td>411</td>
<td>7.03 (5.01–10.97)</td>
<td>2.79 (1.47–5.31)</td>
</tr>
<tr>
<td>Total</td>
<td>4119</td>
<td></td>
<td>4.9%</td>
<td>14.1%</td>
</tr>
</tbody>
</table>