

Utility of the Surgical Apgar Score

Validation in 4119 Patients

Scott E. Regenbogen, MD, MPH; Jesse M. Ehrenfeld, MD; Stuart R. Lipsitz, ScD;
Caprice C. Greenberg, MD, MPH; Matthew M. Hutter, MD, MPH; Atul A. Gawande, MD, MPH

Objectives: To confirm the utility of a 10-point Surgical Apgar Score to rate surgical outcomes in a large cohort of patients.

Design: Using electronic intraoperative records, we calculated Surgical Apgar Scores during a period of 2 years (July 1, 2003, through June 30, 2005).

Setting: Major academic medical center.

Patients: Systematic sample of 4119 general and vascular surgery patients enrolled in the National Surgical Quality Improvement Program surgical outcomes database at a major academic medical center.

Main Outcome Measures: Incidence of major postoperative complications and/or death within 30 days of surgery.

Results: Of 1441 patients with scores of 9 to 10, 72 (5.0%) developed major complications within 30 days,

including 2 deaths (0.1%). By comparison, among 128 patients with scores of 4 or less, 72 developed major complications (56.3%; relative risk, 11.3; 95% confidence interval, 8.6-14.8; $P < .001$), of whom 25 died (19.5%; relative risk, 140.7; 95% confidence interval, 33.7-587.4; $P < .001$). The 3-variable score achieves C statistics of 0.73 for major complications and 0.81 for deaths.

Conclusions: The Surgical Apgar Score provides a simple, immediate, objective means of measuring and communicating patient outcomes in surgery, using data routinely available in any setting. The score can be effective in identifying patients at higher- and lower-than-average likelihood of major complications and/or death after surgery and may be useful for evaluating interventions to prevent poor outcomes.

Arch Surg. 2009;144(1):30-36

Author Affiliations:

Department of Health Policy and Management, Harvard School of Public Health (Drs Regenbogen, Greenberg, and Gawande), Departments of Surgery (Drs Regenbogen and Hutter) and Anesthesia and Critical Care (Dr Ehrenfeld), Massachusetts General Hospital, and Center for Surgery and Public Health, Department of Surgery, Brigham and Women's Hospital (Drs Lipsitz, Greenberg, and Gawande), Boston, Massachusetts.

SURGICAL TEAMS LACK A ROUTINE, objective evaluation of patient condition after surgery to inform postoperative prognostication, guide clinical communication, and evaluate the efficacy of safety interventions in the operating room.¹ Instead, surgeons rely primarily on subjective assessment of available patient data.² Complex models, such as the Acute Physiology and Chronic Health Evaluation score³ and the Physiologic and Operative Severity Score for the Enumeration of Mortality,⁴ provide adequate predictions of a surgical patient's risk of complications. These scores have not come into standard use for surgical patients, however, because they are not easily calculated at the bedside, require numerous data elements that are not uniformly collected, and are often not well understood among the various members of a multidisciplinary care team.⁵ Efforts

to significantly reduce surgery's overall 3% major complication rate⁶ have been hampered in part because surgical departments in most hospitals have no easily applied tool for routine measurement and monitoring of surgical results.

See Invited Critique at end of article

We sought to develop a surgical outcome score that would be (1) simple for teams to collect immediately on completion of an operation for any patient in any setting, regardless of resource and technological capacity; (2) valid for predicting major postoperative complications and death; and (3) applicable throughout the fields of general and vascular surgery (at least). Our approach differs from that of risk-adjusted outcomes evaluations, such as the American College of Surgeons' National Surgical Quality Improvement Pro-

Table 1. The 10-Point Surgical Apgar Score^a

	Surgical Apgar Score, No. of Points				
	0	1	2	3	4
Estimated blood loss, mL	>1000	601-1000	101-600	≤100	
Lowest mean arterial pressure, mm Hg	<40	40-54	55-69	≥70	
Lowest heart rate/min	>85 ^b	76-85	66-75	56-65	≤55 ^b

^aThe 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 anesthesia record during the operation. The score is the sum of the points from each category.

^bOccurrence of pathologic bradyarrhythmia, including sinus arrest, atrioventricular block or dissociation, junctional or ventricular escape rhythms, and asystole, also receives 0 points for lowest heart rate.

gram (NSQIP).^{7,8} Rather than dissociating patient-related factors from those related to surgical performance, the Surgical Apgar Score takes a public health perspective on surgical results, seeking to promptly identify patients at highest risk and circumstances offering greatest opportunity for reducing complications and death, regardless of the prevailing cause. The Apgar score in obstetrics served a similar function in evaluating the condition of newborns and, as a result, became an indispensable clinical tool.⁹⁻¹⁵

We devised an Apgar score for surgery, a 10-point score to rate surgical outcomes at Brigham and Women's Hospital.¹⁵ The score is calculated from the estimated blood loss (EBL), lowest heart rate (HR), and lowest mean arterial pressure (MAP) during an operation. In a pilot study of 767 general and vascular surgery patients,¹⁵ the score was significantly associated with the occurrence of major complications or death within 30 days of surgery ($P < .001$, C statistic = 0.72). Poor-scoring patients (scores ≤ 4) were 16 times more likely to experience a major complication than were patients with the highest scores (9 or 10).

This preliminary study, however, was conducted in a single institution, with a limited sample size. To evaluate the broader applicability of the Surgical Apgar Score, we sought to evaluate its performance among a larger cohort of patients, from a different institution, and used electronic intraoperative data collection rather than the handwritten records from which it was derived. To evaluate its predictive ability, we compare its discrimination with that of the multivariate risk-adjustment models of the NSQIP, an established surgical risk-adjustment method, currently in use in selected centers.

METHODS

PATIENT SELECTION

The Massachusetts General Hospital (MGH) Department of Surgery maintains an outcomes database on a systematic sample of patients undergoing general and vascular surgical procedures as part of the private sector NSQIP. In this program,^{7,8} 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 ab-

dominal aortic aneurysm) regardless of anesthetic type, are eligible for inclusion. Children younger than 16 years and patients undergoing trauma surgery, transplant surgery, vascular access surgery, or endoscopic-only procedures are excluded. At the MGH, the first 40 consecutive patients undergoing operations that meet inclusion criteria in each 8-day cycle are enrolled. No more than 5 patients undergoing inguinal herniorrhaphies and 5 patients undergoing breast biopsies are enrolled per 8-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 (1) patients undergoing carotid endarterectomies performed concurrently with coronary artery bypass grafting because the score was not designed for application to patients receiving cardiopulmonary bypass and (2) patients receiving local anesthesia only because no electronic anesthesia record is generated for these procedures. The study protocol, including a waiver of informed consent for individual patients, was approved by the Human Research Committees of the MGH and the Harvard School of Public Health.

CALCULATION OF THE SCORE

We devised the Surgical Apgar Score by using multivariate logistic regression to screen a collection of intraoperative measures. We found that only 3 intraoperative variables remained independent predictors of 30-day major complications: the EBL, the lowest HR, and the lowest MAP during the operation. The score was thus developed using these 3 variables, and their β coefficients were used to weight the points allocated to each variable in a 10-point score. This procedure is described in detail elsewhere.¹⁵ **Table 1** gives the values used to calculate the 10-point score. The score for a patient with 50 mL of blood loss (3 points), a lowest MAP of 80 (3 points), and a lowest HR of 60 (3 points), for example, is 9. By contrast, a patient with more than a liter of blood loss (0 points), a MAP that decreased to 50 (1 point), and a lowest HR of 80 (1 point) receives a score of 2.

We used intraoperative data collected from handwritten anesthesia records to develop the score at Brigham and Women's Hospital.¹⁵ At the MGH, intraoperative records are maintained by an electronic Anesthesia Information Management System (Saturn; Dräger Medical, Telford, Pennsylvania) in a database that is accessible via the Structured Query Language. We developed a Structured Query Language query to examine the intraoperative physiologic data during the surgical period. Because electronic anesthesia data differ from handwritten records in a number of ways,^{16,17} particularly the tendency for inclusion of some artifactual or erroneous values (for example, false pressure readings when an arterial catheter is flushed), our data extraction algorithm excluded extraphysiologic values for HR (data points <20/min or >200/min) and MAP (data

points <25 mm Hg or >180 mm Hg) and then selected the median of remaining values in every 5-minute period for analysis. The lowest of these medians for each variable, along with the recorded EBL, was used to calculate the score.

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 individual factor values and the total score obtained by each method were compared by computing κ statistics for agreement, using Fleiss-Cohen weighting for ordered categorical data.¹⁸

PATIENT RISK FACTORS AND POSTOPERATIVE OUTCOMES

We collected all preoperative and postoperative patient variables from the NSQIP database. Some variables were aggregated by organ system. Pulmonary comorbidity was defined as preexisting chronic obstructive pulmonary disease, ventilator dependence, or pneumonia. Cardiovascular comorbidity was defined as prior myocardial infarction, angina, congestive heart failure, or coronary revascularization. Patients with a history of transient ischemic attack or stroke with or without residual neurologic deficit were pooled into a single group called "history of stroke or transient ischemic attack." On the basis of previous studies, American Society of Anesthesiologists' Physical Status Classification was dichotomized as 3 or greater and less than 3, and wound classification was dichotomized as clean and clean but contaminated operations vs contaminated and dirty operations.^{7,19} Laboratory data were categorized according to the fiscal year 2005 NSQIP models.²⁰ Procedural relative value units were calculated by linkage of *Current Procedural Terminology* codes to listings from the 2005 Medicare Physician Fee Schedule (Centers for Medicare and Medicaid Services). The magnitude of the surgical procedures was rated as either minor or intermediate (eg, breast, endocrine, groin and umbilical herniorrhaphy, appendectomy, laparoscopic cholecystectomy, perianal procedures, and skin or soft-tissue operations) or major or extensive (all other operations) as in previous studies of perioperative risk.^{21,22}

The primary end points were major complication and death within 30 days after surgery, as recorded in the NSQIP database. The following NSQIP-defined⁸ events were considered major complications: acute renal failure, bleeding that required a transfusion of 4 U or more of red blood cells within 72 hours after surgery, cardiac arrest requiring cardiopulmonary resuscitation, coma of 24 hours or longer, deep venous thrombosis, myocardial infarction, unplanned intubation, ventilator use for 48 hours or more, pneumonia, pulmonary embolism, stroke, wound disruption, deep or organ-space surgical site infection, sepsis, septic shock, systemic inflammatory response syndrome, 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.²³ "Other occurrences" that involved 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.

STATISTICAL ANALYSIS

All analyses were performed using the SAS statistical software, version 9.1 (SAS Institute Inc, Cary, North Carolina). We analyzed continuous variables using 2-sided *t* tests or, if skewed,

Wilcoxon rank sum tests. We analyzed categorical predictors using χ^2 tests. We performed univariate logistic regression to examine the relationship between major complication or death and the Surgical Apgar Score (treating the score as an ordered categorical variable) and calculated C statistics as a measure of model discrimination. We used χ^2 tests and the Cochran-Armitage χ^2 trend test²⁴ to evaluate the relationship between the score and the incidence of both outcomes.

For each outcome, we also compared the univariate logistic regression models with the score alone against the multivariate logistic regression models used for risk adjustment in the private-sector NSQIP for fiscal year 2005.²⁰ Only observations with complete data available were included in the NSQIP models. As measures of discrimination, we constructed receiver operating characteristic curves and calculated C statistics (equivalent to the area under the receiver operating characteristic curve) to compare the models.^{25,26}

RESULTS

DATA ACCRUAL AND VALIDATION

Of 4163 patients identified in the NSQIP database who met inclusion criteria, 4119 (98.9%) had complete electronic intraoperative records and constituted our final cohort. The automated data extraction algorithm achieved excellent agreement with manual record review, both for point values assigned to each variable ($\kappa=0.97$ for HR; $\kappa=0.75$ for MAP) and for the total score ($\kappa=0.94$).

PREDICTORS OF MAJOR COMPLICATIONS AND DEATH

Table 2 compares the demographic characteristics, baseline comorbidities, and laboratory data for patients with and without major complications. One or more major complications occurred within 30 days of surgery in 581 patients (14.1%), including 94 deaths (2.3%). All preoperative risk factors and laboratory values collected were significantly associated with the rate of major complications, with the exceptions of race and obesity.

The 3 variables that contributed to the Surgical Apgar Score were each significant univariate predictors of major complications, including death (**Table 3**). The mean lowest HRs were significantly lower (58 vs 63; $P < .001$) and the mean lowest MAPs were significantly higher (65 vs 61; $P < .001$) among patients with no complications compared with those with major complications. Likewise, median EBL was significantly lower in operations with no major complications than in those resulting in major complications (25 vs 200 mL; $P < .001$). The types of operations and their complication rates in the cohort are given in **Table 4**.

RELATIONSHIP OF SURGICAL APGAR SCORE AND OUTCOMES

With increasing scores, the incidence of major complications and death decreased monotonically ($P < .001$). In univariate logistic regression, the score demonstrated good discrimination, with a C statistic of 0.73 for major complications and 0.81 for death.²⁵

Table 2. Preoperative Risk Factors and 30-Day Outcomes^a

	No Major Complication (n=3538)	Major Complication and/or Death (n=581)	P Value ^b
Patient characteristics			
Age, mean (SD), y	56.6 (16.4)	64.2 (16.0)	<.001
Female	2124 (60.0)	285 (49.1)	<.001
Race other than white	311 (8.8)	51 (8.8)	.99
ASA class \geq 3	1019 (28.8)	371 (63.9)	<.001
Obese (BMI >35) ^c	613 (19.1)	75 (15.8)	.08
Underweight (BMI <18.5) ^c	137 (4.3)	26 (6.1)	<.001
Cardiovascular disease (MI, CHF, angina, coronary revascularization)	418 (11.8)	138 (23.8)	<.001
Hypertension	1415 (40.0)	314 (54.0)	<.001
Pulmonary disease (pneumonia, COPD, ventilator dependent)	145 (4.1)	116 (20.0)	<.001
Dyspnea	217 (6.1)	98 (16.9)	<.001
Diabetes mellitus	397 (11.2)	96 (16.5)	<.001
Renal failure	37 (1.0)	56 (9.6)	<.001
Sepsis	63 (1.8)	109 (18.8)	<.001
Open wound	156 (4.4)	99 (17.0)	<.001
Bleeding disorder or \geq 5 U of PRBCs transfused during admission	86 (2.4)	67 (11.5)	<.001
History of stroke or TIA	212 (6.0)	60 (10.3)	<.001
Current smoker	583 (16.5)	137 (23.6)	<.001
Disseminated cancer	147 (4.2)	46 (7.9)	<.001
Weight loss >10% in 6 mo	226 (6.4)	78 (13.4)	<.001
Oral or parenteral corticosteroid use	136 (3.8)	44 (7.6)	<.001
Ascites	36 (1.0)	46 (7.9)	<.001
Esophageal varices	8 (0.2)	5 (0.9)	.01
Rest pain or gangrene	74 (2.1)	40 (6.9)	<.001
Coma	0	1 (0.2)	.01
Do-not-resuscitate status	9 (0.3)	13 (2.2)	<.001
Laboratory data ^d			
WBC count >11 000/ μ L (3)	387 (10.9)	158 (27.2)	<.001
Hematocrit <38% (4)	1277 (36.1)	340 (58.4)	<.001
Platelet count <150 000/ μ L (4)	172 (5.1)	77 (13.5)	<.001
Platelet count >400 000/ μ L (4)	221 (6.5)	66 (11.5)	<.001
Sodium <135 mEq/L (14)	162 (5.4)	73 (13.2)	<.001
Sodium >145 mEq/L (14)	40 (1.3)	18 (3.3)	.001
SUN >40 mg/dL (8)	70 (2.2)	57 (10.1)	<.001
Creatinine >1.2 mg/dL (12)	423 (13.9)	160 (28.5)	<.001
Aspartate aminotransferase >40 U/L (33)	290 (12.6)	109 (22.9)	<.001
Albumin, mean (SD), g/dL (37)	3.9 (0.6)	3.2 (0.9)	<.001

Abbreviations: ASA, American Society of Anesthesiologists; BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; MI, myocardial infarction; PRBC, packed red blood cells; SUN, serum urea nitrogen; TIA, transient ischemic attack; WBC, white blood cell.

SI conversion factors: To convert albumin to grams per liter, multiply by 10; creatinine to micromoles per liter, by 88.4; hematocrit to a proportion of 1, by 0.01; platelet count to number of cells $\times 10^9$ per liter, by 1; sodium to millimoles per liter, by 1; and SUN to millimoles per liter, by 59.485.

^aResults are presented as number (percentage) of patients unless otherwise indicated. Preoperative patient characteristics and laboratory data were obtained from the Massachusetts General Hospital National Surgical Quality Improvement Program database for 4119 patients who underwent general or vascular surgery between July 1, 2003, and June 30, 2005.

^bThe P values were obtained from χ^2 tests, t tests, or Wilcoxon rank sum tests.

^cPatients with a BMI of 18.5 or greater and 35 or less served as the comparison group; data were missing for 438 patients (10.6%). $P < .001$ for all variables.

^dThe percentage of patients with missing laboratory data, ranging from 3% missing white blood cell count to 37% missing albumin levels, is shown in parentheses after each analyte.

The rates of major complications and death at each level of the Surgical Apgar Score are shown in **Figure 1**. Among 1441 patients with scores of 9 or 10, 72 (5.0%) developed major complications within 30 days, including 2 deaths (0.1%). By comparison, among 128 patients with scores of 4 or less, 72 (56.3%) developed major complications, of whom 25 died (19.5%). Compared with scores of 9 or 10, the relative risk of major complications for scores of 4 or less is 11.3 (95% confidence interval [CI], 8.6-14.8; $P < .001$), and the relative risk of death is 140.7 (95% CI, 33.7-587.4; $P < .001$). In every 2-point score category (as in

Figure 1), the incidence of both major complications and death was significantly greater than that of the next-highest category ($P < .001$), except for the comparisons between the 0- to 2-point and the 3- or 4-point groups ($P = .11$ for major complications and $P = .009$ for death), in which statistical power was limited by the low prevalence of these poorest scores.

Even after stratifying the patients by the magnitude of operation, the score remained a highly significant predictor of outcomes. Among major or extensive operations, patients with scores of 4 or less were 6.5 times more likely to have a major complication (95% CI, 4.7-8.9;

Table 3. Procedure Characteristics and 30-Day Outcomes^a

Operative Characteristic	No Major Complication (n=3538)	Major Complication and/or Death (n=581)
EBL, median (IQR), mL	25 (Minimum-150)	200 (50-750)
Lowest MAP, mean (SD), mm Hg	65 (11)	61 (10)
Lowest HR, mean (SD), beats/min	58 (11)	63 (16)
Emergency procedure, No. (%)	192 (5.4)	121 (20.8)
RBC transfusion, No. (%) ^b	252 (7.1)	212 (36.5)
Duration, median (IQR), min	109 (61-175)	164 (101-260)
Inpatient operation, No. (%)	2912 (82.3)	571 (98.3)
Work RVUs, mean (SD)	17.2 (9.9)	22.9 (12.7)

Abbreviations: EBL, estimated blood loss; HR, heart rate; IQR, interquartile range; MAP, mean arterial pressure; RBC, red blood cell; RVUs, relative value units.

^aProcedure characteristics were obtained from the Anesthesia Information Management System and the Massachusetts General Hospital National Surgical Quality Improvement Program database. The 3 variables that constitute the Surgical Apgar Score (EBL, lowest MAP, and lowest HR) were all significantly related to the incidence of major complications and/or death within 30 days of surgery.

^bFive patients had missing data.

Table 4. Types of Operations and Complication Rates^a

Type of Operation	No. of Operations	Major Complications and/or Deaths Rate, No. (%)	Deaths Alone Rate, No. (%)
Minor or intermediate	1801	87 (4.8) ^a	8 (0.4) ^b
Simple alimentary (eg, laparoscopic cholecystectomy, appendectomy, and perianal)	497	31 (6.2)	1 (0.2)
Breast	448	16 (3.6)	0
Thyroid, parathyroid, or parotid	443	11 (2.5)	3 (0.7)
Groin or umbilical hernia repair	228	7 (3.1)	0
Skin or soft tissue	147	18 (12.2)	3 (2.0)
Endovascular	38	4 (10.5)	1 (2.6)
Major or extensive	2318	494 (21.3) ^a	86 (3.7) ^b
Complex alimentary or retroperitoneal (eg, colectomy and gastrectomy)	770	182 (23.6)	40 (5.2)
Open vascular	699	153 (21.9)	31 (4.4)
Hepatobiliary or pancreas	410	114 (27.8)	14 (3.4)
Bariatric (open or laparoscopic)	327	33 (10.1)	0
Ventral or incisional hernia repair	112	12 (10.7)	1 (0.9)

^aThe frequency of different types of operations and their major complication and death rates among 4119 general and vascular surgery patients are given. Operations were classified as minor or intermediate or major or extensive according to the classification in Arvidsson et al.^{21,22}

^b $P < .001$ (χ^2 test).

$P < .001$) and 112.0 times more likely to die (95% CI, 15.3-819.7; $P < .001$) within 30 days. After minor or intermediate operations, patients with scores of 4 or less were 22.8 times more likely to experience a major complication (95% CI, 12.6-41.1; $P < .001$) and 81.4 times more likely to die (95% CI, 5.4-1219.5; $P < .001$).

COMPARISON OF MODELS

Receiver operating characteristic curves for the Surgical Apgar Score and for multivariate models based on the separate NSQIP risk adjustment models for morbidity and mortality are shown in **Figure 2**. Complete risk predictions could be generated, however, for only 2482 patients (60.3%) in the NSQIP morbidity model and 2370 (57.5%) patients in the mortality model because required information, most often laboratory data, was missing. Among the restricted set of patients for whom all required data were available, the NSQIP models provided better discrimination than did the score alone for both morbidity ($C = 0.81$ vs $C = 0.72$; $P < .001$) and mortality ($C = 0.93$ vs $C = 0.78$; $P < .001$).²⁶

COMMENT

A simple surgical score based on blood loss, lowest HR, and lowest MAP during an operation provides a meaningful estimate of patients' condition and risk after general and vascular surgery. The 10-point Surgical Apgar Score is predictive of both major complications and death in the immediate postoperative period and is valid across the diversity of general and vascular surgery. We have shown that it remains highly predictive in a different institution from where it was derived and remains robust to the known differences between handwritten and electronic intraoperative records.

The score successfully identifies not only the patients at highest risk of postoperative complications but also those at markedly lower-than-average risk. The 1441 patients with scores of 9 or 10 (35.0% of the sample) had only a 5.0% incidence of major complications and a 0.1% incidence of death. In contrast, most patients with scores of 4 or less had major complications and more than 1 in 5 died. Despite the relatively low prevalence of scores of 4 or less (3.1% overall), the consistent trend toward worse

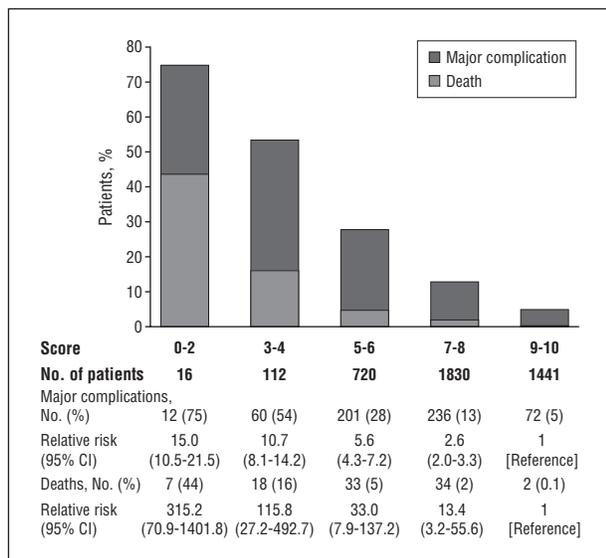


Figure 1. Thirty-day major complications and deaths among 4119 general and vascular surgery patients in relation to Surgical Apgar Score. Major complication and death rates are shown according to the 10-point Surgical Apgar Score from the operation. Patients with scores of 9 or 10 served as the reference group. Risk of major complications and death decreased significantly with increasing scores (Cochran-Armitage trend test, both $P < .001$). CI indicates confidence interval.

outcomes even at the extreme low end of the scale suggests that the score has good discriminative ability across the full point spectrum.²⁵

The Surgical Apgar Score could serve several important purposes. Like the Apgar score for newborns, its primary value may be to provide teams with immediate feedback on operative condition for every patient¹³—an objective metric to complement their “gut feelings”^{2,27} about an operation. Because the feedback is immediate, the score would assist surgical teams in distinguishing patients most in need of increased intensity of postoperative monitoring and care from those likely to have an uncomplicated course. As a quantitative adjunct to surgeons’ subjective impressions, the score may serve as a simple aid in communication among surgeons, postanesthesia care providers, surgical residents, and surgical ward staff regarding patients’ immediate postoperative status and thereby assist decision making about, for example, unplanned admission after outpatient surgery, admission to the intensive care unit, or frequency of postoperative examinations by physicians and nurses. Surgeons might also use the score to convey to patients and families an appraisal of condition and prognosis after surgery. Looking forward, the score might be used as a metric for quality monitoring and innovation, even in resource-poor settings. Routine surveillance and case review for patients with low scores (eg, a score of ≤ 4), even when no complications result, may also enable earlier identification of safety problems.²⁸

Like the obstetric Apgar score, however, our surgical score does not allow comparison of quality among institutions or physicians because its 3 variables are each influenced not only by the performance of medical teams but also by the patients’ prior condition and the magnitude of the operations they undergo. The NSQIP has developed a

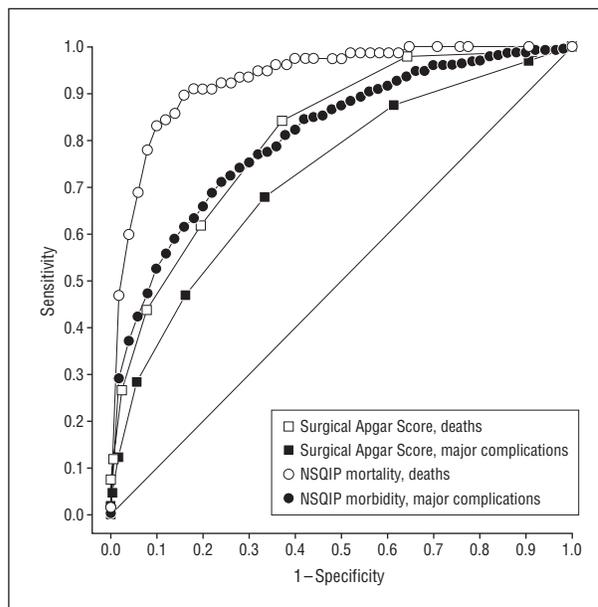


Figure 2. Receiver operating characteristic curves for the Surgical Apgar Score and the National Surgical Quality Improvement Program (NSQIP) morbidity and mortality models as predictors of major complications and death. The sensitivity and specificity of the Surgical Apgar Score were compared with the separate morbidity and mortality models from the NSQIP.¹⁹ The score achieved a C statistic of 0.73 for predicting major complications and 0.81 for predicting deaths. C statistics for the NSQIP model were significantly greater for both major complications ($C = 0.81$, $P < .001$) and deaths ($C = 0.93$, $P < .001$).

risk-adjustment algorithm for detailed modeling and case mix adjustment that serves these purposes.^{7,8} The Surgical Apgar Score is not intended to supplant these methods of institutional quality assessment because its motivation and intended uses are distinct. Nevertheless, we provide a comparison between this intraoperative score and the preoperative risk-adjustment models from the NSQIP in Figure 2 as a point of reference by which its discriminative ability may be appraised.

As a simple, objective measure, the Surgical Apgar Score offers an important addition to risk-adjustment strategies for institutional quality assessment. Because of the expense of data collection, comprehensive risk-adjusted 30-day outcome tracking is not yet achievable in most US hospitals, let alone hospitals worldwide. Complex, multivariate models are not commonly used in clinical settings because they are difficult for teams to interpret and communicate at the bedside and often require statistical imputation of key information because of missing data.^{28,29} The Surgical Apgar Score can be available in real time, immediately usable for clinical decision support, and easily and inexpensively collected in any hospital. It is these same characteristics that made the Apgar score such a powerful tool for broad safety improvement in obstetrics.^{13,14}

Nonetheless, our study has several limitations. First, the Surgical Apgar Score has been tested only in general and vascular surgery patients 16 years or older. Whether the score is effective in grading risk in other fields of surgery remains uncertain, and it has not been adapted for use in pediatric populations. Second, the score has not been evaluated beyond major

academic medical centers because of a lack of reliable and comprehensive outcomes assessment against which these measures could be validated. It is possible that, among other patient populations, some modifications to the score factors could be necessary. Third, blood loss estimation is inevitably imprecise. The broad categories used to calculate the score, however, are well within observers' range of precision in careful volumetric studies.^{30,31} Reliance on anesthesiologists' independent estimation further improves the reliability and insulates against surgeon bias.³⁰ The variables in our score are at least as reliably quantified as any in the Apgar score and potentially more so than some Apgar components (such as grading of newborn muscle tone and color).¹²

Our results, therefore, demonstrate that a simple clinimetric surgical outcome score can be derived from intraoperative data alone. This 10-point score based on the lowest HR, lowest MAP, and EBL discriminates well between groups of patients at higher- and lower-than-average risk of major complications and death within 30 days of surgery and holds promise as both a prognostic measure and a clinical decision support tool. Our hope is that this score will prove useful for routine care, quality improvement, and clinical research in surgery.

Accepted for Publication: November 8, 2008.

Correspondence: Scott E. Regenbogen, MD, MPH, Department of Health Policy and Management, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115 (sregenbogen@partners.org).

Author Contributions: Drs Regenbogen and Gawande had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Regenbogen, Ehrenfeld, Lipsitz, Greenberg, Hutter, and Gawande. *Acquisition of data:* Regenbogen, Ehrenfeld, and Hutter. *Analysis and interpretation of data:* Regenbogen, Ehrenfeld, Lipsitz, Greenberg, Hutter, and Gawande. *Drafting of the manuscript:* Regenbogen, Lipsitz, and Gawande. *Critical revision of the manuscript:* Regenbogen, Ehrenfeld, Lipsitz, Greenberg, Hutter, and Gawande. *Statistical analyses:* Regenbogen, Lipsitz, and Greenberg. *Obtained funding:* Regenbogen and Gawande. *Study supervision:* Gawande.

Financial Disclosure: None reported.

Funding/Support: This research was supported by a grant from the Risk Management Foundation of the Harvard Medical Institutions. Dr Regenbogen was also supported by Kirschstein National Research Service Award T32-HS000020 from the Agency for Healthcare Research and Quality.

Role of the Sponsors: The funding agencies were not involved in any aspect of the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript.

Additional Contributions: Lynn Devaney, RN, assisted with the MGH-NSQIP database and John Walsh, MD, assisted with the intraoperative anesthesia record.

1. Vincent C, Moorthy K, Sarker SK, Chang A, Darzi AW. Systems approaches to surgical quality and safety: from concept to measurement. *Ann Surg.* 2004; 239(4):475-482.
2. Hartley MN, Sagar PM. The surgeon's "gut feeling" as a predictor of post-operative outcome. *Ann R Coll Surg Engl.* 1994;76(6)(suppl):277-278.
3. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE—acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med.* 1981;9(8):591-597.
4. Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg.* 1991;78(3):355-360.
5. Whiteley MS, Prytherch DR, Higgins B, Weaver PC, Prout WG. An evaluation of the POSSUM surgical scoring system. *Br J Surg.* 1996;83(6):812-815.
6. Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery.* 1999;126(1):66-75.
7. Fink AS, Campbell DA Jr, Mentzer RM Jr, et al. The National Surgical Quality Improvement Program in non-veterans administration hospitals: initial demonstration of feasibility. *Ann Surg.* 2002;236(3):344-354.
8. Khuri SF, Daley J, Henderson W, et al. The National Veterans Administration Surgical Risk Study: risk adjustment for the comparative assessment of the quality of surgical care. *J Am Coll Surg.* 1995;180(5):519-531.
9. Apgar V. A proposal for a new method of evaluation of the newborn infant. *Curr Res Anesth Analg.* 1953;32(4):260-267.
10. Apgar V, Holaday DA, James LS, Weisbrod IM, Berrien C. Evaluation of the newborn infant; second report. *J Am Med Assoc.* 1958;168(15):1985-1988.
11. Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the assessment of newborn infants. *N Engl J Med.* 2001;344(7):467-471.
12. Finster M, Wood M. The Apgar score has survived the test of time. *Anesthesiology.* 2005;102(4):855-857.
13. Apgar V, James LS. Further observations on the newborn scoring system. *Am J Dis Child.* 1962;104:419-428.
14. Feinstein AR. Multi-item "instruments" vs Virginia Apgar's principles of clinimetrics. *Arch Intern Med.* 1999;159(2):125-128.
15. Gawande AA, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ. An Apgar score for surgery. *J Am Coll Surg.* 2007;204(2):201-208.
16. 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(3):612-616.
17. Cook RI, McDonald JS, Nunziata E. Differences between handwritten and automatic blood pressure records. *Anesthesiology.* 1989;71(3):385-390.
18. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas.* 1973;33(3):613-619.
19. Pasternak LR. Risk assessment in ambulatory surgery: challenges and new trends. *Can J Anesthesiol.* 2004;51(6):R1-R5.
20. American College of Surgeons' NSQIP Advisory Committee. *Semi-annual Report.* Chicago, IL: ACS NSQIP Advisory Committee; 2006.
21. Arvidsson S, Ouchterlony J, Nilsson S, Sjostedt L, Svardsudd K. The Gothenburg study of perioperative risk, I: preoperative findings, postoperative complications. *Acta Anaesthesiol Scand.* 1994;38(7):679-690.
22. Arvidsson S, Ouchterlony J, Sjostedt L, Svardsudd K. Predicting postoperative adverse events: clinical efficiency of four general classification systems: the project perioperative risk. *Acta Anaesthesiol Scand.* 1996;40(7):783-791.
23. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg.* 2004;240(2):205-213.
24. Armitage P. Test for linear trends in proportions and frequencies. *Biometrics.* 1955;11(3):375-386.
25. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143(1):29-36.
26. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* 1983;148(3):839-843.
27. Markus PM, Martell J, Leister I, Horstmann O, Brinker J, Becker H. Predicting post-operative morbidity by clinical assessment. *Br J Surg.* 2005;92(1):101-106.
28. Jones HJ, de Cossart L. Risk scoring in surgical patients. *Br J Surg.* 1999;86(2):149-157.
29. Sutton R, Bann S, Brooks M, Sarin S. The Surgical Risk Scale as an improved tool for risk-adjusted analysis in comparative surgical audit. *Br J Surg.* 2002; 89(6):763-768.
30. Delikan AE. Comparison of subjective estimates by surgeons and anaesthetists of operative blood loss. *Br Med J.* 1972;2(5814):619-621.
31. Gardiner AJ, Dudley HA. The measurement of blood loss at operation. *Br J Anaesth.* 1962;34:653-656.