An Apgar Score for Surgery

Atul A Gawande, MD, MPH, FACS, Mary R Kwaan, MD, MPH, Scott E Regenbogen, MD,
Stuart A Lipsitz, SCD, Michael J Zinner, MD, FACS

BACKGROUND: Surgical teams have not had a routine, reliable measure of patient condition at the end of an operation. We aimed to develop an Apgar score for the field of surgery, an outcomes score that teams could calculate at the end of any general or vascular surgical procedure to accurately grade a patient's condition and chances of major complications or death.

STUDY DESIGN: We derived our surgical score in a retrospective analysis of data from medical records and the National Surgical Quality Improvement Program for 303 randomly selected patients undergoing colectomy at Brigham and Women's Hospital, Boston. The primary outcomes measure was incidence of major complication or death within 30 days of operation. We validated the score in two prospective, randomly selected cohorts: 102 colectomy patients and 767 patients undergoing general or vascular operations at the same institution.

RESULTS: A 10-point score based on a patient's estimated amount of blood loss, lowest heart rate, and lowest mean arterial pressure during general or vascular operations was significantly associated with major complications or death within 30 days ($p < 0.0001$; c-index = 0.72). Of 767 general and vascular surgery patients, 29 (3.8%) had a surgical score ≤ 4. Major complications or death occurred in 17 of these 29 patients (58.6%) within 30 days. By comparison, among 220 patients with scores of 9 or 10, only 8 (3.6%) experienced major complications or died (relative risk 16.1; 95% CI, 7.6–34.0; $p < 0.0001$).

CONCLUSIONS: A simple score based on blood loss, heart rate, and blood pressure can be useful in rating the condition of patients after general or vascular operations. (J Am Coll Surg 2007;204:201–208. © 2007 by the American College of Surgeons)

In 1953, Virginia Apgar introduced a 10-point scoring system for evaluation of the condition of newborns, which revolutionized obstetric care.[1] Until that point, obstetricians had only their subjective impressions of the immediate outcomes of childbirth. The Apgar score proved simple to use; effective at providing clinicians with clear, graded feedback on how delivery had gone for the child; and predictive of 28-day survival.[2,3] As a result, it enabled more consistent identification of newborns at high risk for death, prompted development of better methods to treat them, and provided a clear measure for testing clinical innovations. The Apgar score became an indispensable tool in achieving the remarkable safety of modern child delivery.

Similar to obstetrics in 1953, surgery today is without a routine and reliable gauge of overall patient condition after surgical procedures to guide clinical practice. Surgical teams rely mainly on subjective assessment of the patient and delayed feedback from 30-day outcomes. The APACHE score[4] and the Physiologic and Operative Severity Score (POSSUM)[5] for the Enumeration of Mortality and Morbidity have both been proposed as clinical measures of patient condition. These scores are not easily calculated at the bedside, require numerous data elements, and rely on laboratory data that are not uniformly collected. As a result, neither has come into standard use for surgical patients.

Our goal for this study was to develop an outcomes score that surgical teams could routinely and easily calculate to grade the condition of patients at the end of any general or vascular surgery procedure.

Competing Interests Declared: None.

Supported by a grant from the Risk Management Foundation of the Harvard Medical Institutions. Drs Kwaan and Regenbogen were also supported by a postdoctoral fellowship from the Agency for Healthcare Research and Quality.

Received July 7, 2006; Revised November 14, 2006; Accepted November 15, 2006.

From the Department of Health Policy and Management, Harvard School of Public Health, Boston, MA (Gawande, Kwaan, Regenbogen); Center for Surgery and Public Health, Department of Surgery, Brigham and Women's Hospital, Boston, MA (Gawande, Kwaan, Lipsitz, Zinner); Department of Surgery, Massachusetts General Hospital, Boston, MA (Regenbogen).

Correspondence address: Atul A Gawande, MD, MPH, Department of Surgery, Brigham and Women's Hospital, 75 Francis St, Boston, MA 02115. email: agawande@partners.org

Published by Elsevier Inc.

ISSN 1072-7515/07/$32.00
doi:10.1016/j.jamcollsurg.2006.11.011

Abbreviations and Acronyms

BWH	= Brigham and Women's Hospital
EBL	= estimated blood loss
MAP	= mean arterial pressure
NSQIP	= National Surgical Quality Improvement Program

METHODS

Overview

We developed and tested our score using medical records data from patients who underwent operations at the Brigham and Women's Hospital, Boston (BWH), between February 1, 2002, and December 31, 2004. We did so with the approval of the Human Subjects Committee at BWH and the Harvard School of Public Health. The study was done in three steps. First, we derived our score using preoperative, intraoperative, and outcomes data from a cohort of patients who underwent open colectomy, which we selected because it is a common procedure known to have a high rate of complications.[6] Then, we examined the ability of our score to predict outcomes in a different cohort of patients who underwent open colectomy. Finally, we examined the ability of our score to predict outcomes for a larger cohort of patients who underwent any general or vascular surgical procedure.

Patient selection and data collection

The BWH Department of Surgery maintains an outcomes database on a random selection of patients undergoing general and vascular surgical procedures as part of the American College of Surgeons' National Surgical Quality Improvement Program (NSQIP). Under this program,[7,8] trained nurse reviewers retrospectively collect 49 preoperative, 17 intraoperative, and 33 outcomes variables on surgical patients for monitoring outcomes. Patients undergoing major general or vascular surgery—defined by general, epidural, or spinal anesthesia, or specified cases (carotid endarterectomy, inguinal herniorrhaphy, thyroidectomy, parathyroidectomy, breast biopsy, and endovascular repair of abdominal aortic aneurysm) regardless of anesthetic type—are eligible for inclusion, except children under age 16 and patients undergoing trauma operation, transplantation operation, vascular access operation, or endoscopic-only procedures. Patient selection is random. At BWH, we enroll the first 40 consecutive operations that meet inclusion criteria in each 8-day cycle. To avoid dominating the sample, no more than 5 inguinal herniorraphies and 5 breast biopsies are enrolled per 8-day cycle.

For inclusion in our initial patient cohort (cohort 1), we selected all patients in the BWH-NSQIP database who underwent open colectomy as their primary procedure between February 1, 2002, and March 31, 2004, using the following CPT codes for partial or total colectomy: 44140, 44141, 44143 to 44147, 44150 to 44153, 44155, 44156, and 44160. For our second cohort (cohort 2), we selected all patients in the database who underwent open colectomy as their primary procedure between April 1, 2004, and December 31, 2004. For our final cohort (cohort 3), we selected all patients in the database who underwent any general or vascular surgery procedure between September 1, 2004, and December 31, 2004.

We included all variables in the BWH-NSQIP database. Some adjustments were made to these variables to create preoperative comorbidity categories or to transform variables that were not normally distributed. Pulmonary comorbidity was defined as preexisting COPD, ventilator dependence, or pneumonia. Cardiovascular comorbidity was defined as earlier MI, stroke, congestive heart failure, peripheral vascular disease, or coronary revascularization. American Society of Anesthesiologists (ASA) physical classification status was dichotomized as ≥ 3 compared with < 3, in accordance with recommendations for preoperative surgical risk assessment.[9] Wound classification was dichotomized, in accordance with other studies using the NSQIP;[8] clean and clean/contaminated operations constituted one group, and contaminated and dirty operations constituted the other.

We also collected 28 variables from the intraoperative anesthesia records for each patient in our study, including operative duration (incision-to-skin closure time); initial, final, highest, and lowest heart rate and blood pressure during the procedure; initial, final, and lowest temperature and oxygen saturation; volume of urine output; estimated blood loss (EBL), volume of IV fluids and blood products administered; use of pressure support medication; and anesthetic type. Because their data were markedly right-skewed, volume of IV fluid (mL) and operative duration (minutes) were transformed with the natural logarithm, and EBL was transformed with logarithm base 10. Lowest oxygen saturation was dichotomized into $> 95\%$ and $\leq 95\%$.

Primary outcomes for our study were death or major complication within 30 days after operation. The following events were defined as major complications: acute renal failure, bleeding requiring $\geq$ 4 U red cell transfusion within 72 hours after operation, cardiac arrest requiring CPR, coma for 24 hours or longer, deep venous thrombosis, septic shock, MI, unplanned intubation, ventilator use for 48 hours or longer, pneumonia, pulmonary embolism, stroke, wound disruption, deep or organ-space surgical site infection, sepsis, systemic inflammatory response syndrome, and vascular graft failure, according to NSQIP's established definitions.[7] All deaths were considered major complications. (Superficial surgical site infection and urinary tract infection were not major complications.) Patients having complications categorized in the database as "other occurrence" were reviewed individually to determine the severity of the complication. Anastomotic leak, cystic duct leak after cholecystectomy, pericardial effusion requiring drainage, and gastric outlet obstruction requiring reoperation were identified on review of these individual occurrences and classified as major complications.

Statistical analysis and development of the score

All analyses were performed using the SAS 9.1 statistical software package (SAS Institute). We performed a univariate analysis examining the relationship between each preoperative and intraoperative variable in the database and the outcomes of major complication or death. Variables with $p < 0.20$ and $< 10\%$ missing data were candidates for entry in a multivariable model. We chose among colinear variables based on p value and potential for use in an outcomes score. We entered the intraoperative variables alone and also with preoperative variables in a multivariable logistic regression model using a stepwise selection procedure ($\alpha = 0.10$ for entry, $\alpha = 0.05$ for inclusion) for outcomes of interest. Variables that independently predicted major complication or death became our candidates for the surgical score. We compared alternate combinations of variables for the surgical score, striving for simplicity of the score, safety in clinical use, and predictive power.

After choosing the final variables for inclusion, we used the magnitude of the β-coefficients from the regression equation to weight the points allocated to each variable appropriately in a 10-point score. We chose the size of the intervals and the cut-offs for each point level so that a 1-point increase in the score for each variable would produce an equivalent increase in the odds of a complication and so that the distribution of observed values was encompassed. The values for each cut-off were then rounded to clinically relevant values amenable to recall by surgeons. We performed logistic regression analysis to examine the relationship between major complications/death and the surgical score (treating surgical score as an ordered categorical variable), and the individual components of the score, in each of the three patient cohorts, and calculated c-statistics as a measure of model discrimination. We performed Pearson's validation goodness-of-fit test[10] to compare model calibration between the two colectomy cohorts. In the final cohort, we compared differences in complication rates by individual scores using chi-square tests.

RESULTS

We identified 311 patients in the BWH-NSQIP database to form cohort 1 (for derivation of our score), 103 patients to form cohort 2 (for validation in colectomy patients), and 775 patients for cohort 3 (for validation in patients undergoing general or vascular surgery). Of these, anesthesia records were available for review in 303 patients for cohort 1 (97.4%), 102 patients for cohort 2 (99.0%), and in 767 patients for cohort 3 (99.0%). Characteristics of these patients are listed in Table 1. Major complications occurred within 30 days in 66 patients (22%) in cohort 1 (including 9 deaths; 3%), 19 patients (19%) in cohort 2 (4 deaths; 4%), and 70 patients (9%) in cohort 3 (11 deaths; 1%).

Derivation of the surgical score

In cohort 1, 12 preoperative and 9 intraoperative variables were associated with major complication or death within 30 days, with a p value < 0.20 in univariate analysis and $< 10\%$ missing data (Table 2). In multivariable logistic regression with eight of the nine intraoperative variables (red cell transfusion was tightly colinear with log EBL and excluded), we found that lowest heart rate, log EBL, and lowest mean arterial pressure (MAP) were each independent predictors of outcomes. These 3 intraoperative variables constitute model 1 (see Table 3). Using both intraoperative and preoperative variables, we found that lowest heart rate, log EBL, pulmonary comorbidity, and age were independent predictors of outcomes (model 2). These two models had similar ability to discriminate among patients with and without major complications or death. (The c-index was 0.72 and 0.73, respectively, indicating good discrimination.)

Table 1. Patient Characteristics, Procedures, and Outcomes

	Colectomy patients		General and vascular surgery patients
	Cohort 1* (n = 303)	Cohort 2† (n = 102)	Cohort 3‡ (n = 767)
Preoperative characteristics			
Age (y)§	60.5 ± 15.1	63.6 ± 15.4	55.3 ± 15.8¶
Female‖	171 (56)	49 (48)	501 (65)¶
Non-Caucasian race‖	42 (14)	11 (11)	#
Body mass index§	27.0 ± 6.1	25.7 ± 6.8	28.8 ± 9.9¶
Cardiovascular disease (MI, CHF, PVD, stroke, prior revascularization)‖	34 (11)	12 (12)	76 (10)
Pulmonary disease (pneumonia, COPD, ventilator dependent)‖	16 (5)	1 (1)	42 (5)
Diabetes mellitus‖	37 (12)	16 (16)	98 (13)
Preoperative sepsis‖	9 (3)	4 (4)	45 (6)
Wound classification, contaminated or dirty	55 (18)	#	#
Procedure for malignancy	153 (51)	#	#
Bleeding disorder/transfusion > 4 U RBC preop‖	12 (4)	3 (3)	33 (4)
Emergency procedure‖	41 (14)	4 (4)¶	45 (6)¶
ASA class ≥ 3‖	121 (40)	40 (39)	264 (34)
Intraoperative characteristics			
Operative duration (min)§	160 ± 95	146 ± 75	112 ± 80¶
Initial heart rate (beats/min)§	80 ± 17	80 ± 16	79 ± 15
Lowest heart rate (beats/min)§	62 ± 13	62 ± 13	63 ± 12
Highest heart rate (beats/min)§	94 ± 17	92 ± 16	92 ± 16
Final heart rate (beats/min)§	80 ± 16	77 ± 14	79 ± 15
Highest mean arterial pressure (mmHg)§	95 ± 14	99 ± 15	100 ± 16¶
Lowest mean arterial pressure (mmHg)§	58 ± 8	61 ± 9¶	63 ± 10¶
Estimated blood loss (mL), median (IQR)¶**	200 (100−500)	250 (100−450)	Minimal (min−150)
Intraoperative RBC transfusion‖	64 (21)	13 (13)	71 (9)¶
Epidural anesthesia	209 (69)	#	#
Outcomes			
No major complications‖	235 (78)	83 (81)	697 (91)¶
Major complications‖	68 (22)	19 (19)	70 (9)¶
Death‖	9 (3)	4 (4)	11 (1)
Types of operations			
Complex abdominal (eg, bariatric, colectomy, pancreatectomy, splenectomy)	303 (100)	102 (100)	287 (37)
Breast/skin/soft tissue	NA	NA	174 (23)
Simple abdominal (laparoscopic cholecystectomy, appendectomy, gastrostomy)	NA	NA	101 (13)
Hernia repair	NA	NA	78 (10)
Vascular/endovascular	NA	NA	71 (9)
Thyroid/parathyroid	NA	NA	56 (7)

Data are given as the number (%) of patients, or as mean ± SD, except where noted.
*February 2003−March 2004.
†April 2004−December 2004.
‡September 2004−December 2004.
§*t*-test.
‖Chi-square test.
¶$p < 0.05$, cohort 1 was comparison group.
#Data not available for this cohort.
**Wilcoxon rank-sum test.
ASA, American Society of Anesthesiologists; CHF, congestive heart failure; IQR, interquartile range; NA, not applicable; PVD, peripheral vascular disease.

Table 2. Characteristics Associated with Major Complications/Death for 303 Colectomy Patients (Cohort 1), Univariate Analysis

	Odds ratio	95% CI	p Value
Preoperative characteristics			
Age (y)	1.02	1.00−1.04	0.07
Female	0.66	0.38−1.13	0.13
Non-Caucasian race	1.91	0.94−3.87	0.07
Body mass index	0.97	0.92−1.02	0.17
Cardiovascular disease	2.79	1.32−5.87	0.007
Pulmonary disease	6.58	2.30−18.84	0.0005
Preoperative sepsis	13.25	2.69−65.42	0.002
Wound classification, contaminated or dirty	3.21	1.72−6.00	0.0003
Procedure for malignancy	0.53	0.30−0.92	0.02
Bleeding disorder/transfusion > 4 U RBC preop	3.66	1.14−11.75	0.03
Emergency procedure	4.80	2.41−9.57	< 0.0001
ASA class ≥ 3	3.74	2.12−6.59	< 0.0001
Intraoperative characteristics			
Initial heart rate (beats/min)	1.02	1.01−1.04	0.006
Lowest heart rate (beats/min)	1.06	1.03−1.08	< 0.0001
Highest heart rate (beats/min)	1.03	1.01−1.05	0.0002
Final heart rate (beats/min)	1.03	1.02−1.05	0.0002
Lowest mean arterial pressure (mmHg)	0.98	0.95−1.01	0.14
Lowest O_2 saturation < 95%	1.97	0.90−4.32	0.09
Estimated blood loss ($\log_{10}$ mL)	2.13	1.30−3.49	0.003
Intraoperative RBC transfusion	3.87	2.13−7.05	< 0.0001
Epidural anesthesia	0.43	0.24−0.76	0.004

Because of its simplicity, because having heart rate and not blood pressure as a component of a performance measure could lead surgical teams to manage heart rate without sufficient regard to blood pressure, and because the discriminative ability of the two models was equivalent, we chose to develop our score based on model 1, with 3 intraoperative variables—lowest heart rate, log EBL, and lowest MAP—as shown in Table 4. The score for a patient with 50 mL blood loss (3 points), a lowest MAP of 80 (3 points), and a lowest heart rate of 60 (3 points), for example, would have been 9. By contrast, a patient with > 1 L blood loss (0 points), a MAP that dropped to 50 (1 point), and a lowest heart rate of 80 (1 point) would receive a score of 2.

Relationship between surgical score and surgical outcomes

We found that as the score increased, outcomes improved monotonically for colectomy patients in both cohort 1 and cohort 2 (Table 5). The surgical score was a significant predictor of major complication or death for both cohorts ($p < 0.0001$). The c-statistic for the score in univariate logistic regression was 0.69 for cohort 1 and 0.67 for cohort 2. Pearson's goodness-of-fit test showed no significant difference between cohorts in the relationship between score and outcomes ($p = 0.57$).

Table 3. Characteristics Associated with Major Complications and Death for 303 Colectomy Patients (Cohort 1), Multivariable Analysis

	Odds ratio	95% CI	p Value
Model 1: Intraoperative variables only (c-statistic 0.72)			
Lowest heart rate	1.06	1.03−1.08	< 0.0001
Log estimated blood loss	1.82	1.08−3.07	0.02
Lowest mean arterial pressure	0.96	0.93−0.99	0.02
Model 2: Intraoperative and preoperative variables (c-statistic 0.73)			
Lowest heart rate	1.06	1.03−1.08	< 0.0001
Log estimated blood loss	2.07	1.16−3.71	0.01
Pulmonary comorbidity (COPD, ventilator dependence, pneumonia)	5.73	1.52−21.58	0.01
Age (y)	1.03	1.01−1.06	0.009

We next examined the relationship between surgical score and outcomes for cohort 3, our 767 patients undergoing general or vascular operations. Of these patients, 11 (1.4%) died, major complications developed within 30 days in 70 (9.1%), 35 (4.6%) had minor complications, and 662 (86.3%) had no complications. Mean surgical score was 7.55 (± 1.49 SD). The occurrence of major complications or death was significantly associated with surgical score in univariate logistic regression ($p < 0.0001$). The c-statistic was 0.72, indicating good discrimination. The score had only slightly less discrimination in predicting 30-day outcomes than a multivariable logistic regression model using the 3 components of the score individually: log EBL, lowest heart rate, and lowest MAP (c-index = 0.75, $p < 0.0001$).

Differences in outcomes between patients with different scores were also statistically significant (see Table 6). Among the 29 (3.8%) patients with a surgical score ≤ 4 after general or vascular surgery, major complications or death occurred within 30 days in 17 (58.6%) patients. In contrast, among the 220 patients (28.7%) with a surgical score of 9 or 10, only 8 (3.6%) suffered a major complication or death within 30 days. This difference corresponds to

Table 4. A 10-Point Surgical Outcomes Score*

	0 points	1 point	2 points	3 points	4 points
Estimated blood loss (mL)	> 1,000	601−1,000	101−600	≤ 100	—
Lowest mean arterial pressure (mmHg)	< 40	40−54	55−69	≥ 70	—
Lowest heart rate (beats/min)	> 85	76−85	66−75	56−65	≤ 55†

Surgical score = sum of the points for each category in the course of a procedure.
*Based on model 1 from cohort 1.
†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.

a relative risk for major complication among low-scoring operations of 16.1 (95% CI, 7.7−34.0, $p < 0.0001$), compared with the highest-scoring operations (Fig. 1).

The surgical score was also highly predictive of death in the general and vascular surgery cohort ($p < 0.0001$ in univariate logistic regression, c-statistic = 0.92; Fig. 1). Death occurred for 0 of 220 patients with scores of 9 to 10; 1 of 395 (0.3%) with scores 7 to 8; 6 of 123 (4.9%) with scores 5 to 6; and 4 of 29 (13.8%) with scores 0 to 4.

DISCUSSION

We were able to derive a simple surgical score based on routinely available intraoperative data that accurately rates the condition of patients after general or vascular surgery. Like Virginia Apgar's score for newborns, its primary value is in providing surgical teams with immediate, graded feedback on how an operation went for a patient. Until now they have had only their subjective impressions from the data available.

This surgical score could serve several important purposes. It would allow surgeons to consistently identify patients coming out of operations who are at highest risk of major complications or death, and test standards and innovations to improve our ability to save such patients. It would provide information to patients and their families on patients' relative conditions after surgical procedures. Routine surveillance and case review for patients with low surgical scores (eg, a score ≤ 4), even when no complications result, can enable earlier identification of safety problems. The score would 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 might be to encourage development and implementation of practices that reduce the number of patients with low scores.

Like the Apgar score, our surgical score does not readily allow comparison of quality between institutions or practitioners, as its three variables are each influenced not only by the performance of medical teams, but also the patient's earlier condition. For such comparisons, we must still rely on tracking risk-adjusted, 30-day outcomes with more complex collection and modeling of patient data, such as in the Veterans' Affairs National Surgical Quality Improvement Program.[7] For the same reason, the score is also not a suitable measure to guide physician payments. This is a score that can accurately grade a patient's condition after operation only, which is always a function of both how fit the patient is coming to the operation and how the team performs.

Our finding that the critical variables for the score are measures of blood loss, heart rate, and blood pressure is consistent with previous findings. Stability of patient vital signs during the surgical procedure[11-13] and the

Table 5. Thirty-Day Surgical Outcomes for Two Colectomy Patient Cohorts, in Relation to the Surgical Score

	Cohort 1* (n = 303)			Cohort 2† (n = 102)		
		Major complication/death			Major complication/death	
Surgical score	n	n	%	n	n	%
0−2	5	5	100	0	—	—
3−4	24	10	42	8	5	63
5−6	107	30	28	25	6	24
7−8	143	22	15	58	7	12
9−10	24	1	4	11	1	9

Pearson's goodness-of-fit test showed no significant difference between cohorts 1 and 2 in the relationship between score and outcomes (p = 0.57).
*February 2002 to March 2004. Cohort 1: c-statistic = 0.69.
†April 2004 to December 2004. Cohort 2: c-statistic = 0.67.

Table 6. Thirty-day Outcomes for 767 Patients Undergoing General or Vascular Surgery, in Relation to Surgical Scores

Surgical score	n	Major complication/death n	Major complication/death %	Relative risk (95% CI)	p Value*
0−2	4	3	75	20.6 (8.5−50.0)	< 0.0001
3−4	25	14	56	15.4 (7.2−33.1)	< 0.0001
5−6	123	20	16	4.5 (2.0−9.8)	< 0.0001
7−8	395	25	6	1.7 (0.8−3.8)	0.16
9−10	220	8	4	1	—

c-statistic = 0.72.
*Chi-square test. Patients with scores of 9 or 10 served as the reference group.

amount of blood loss[14-17] have long been recognized as important in patient outcomes. The score is also consistent with findings that outcomes can be improved with more appropriate use of β-blockers in patients undergoing major operations.[18,19] What had not been recognized was the collective importance of these variables.

Together, these three predictors allow teams to successfully identify not only the patients at highest risk of postoperative complications, but also those at markedly lower risk than average. The 220 patients with scores of 9 or 10 in cohort 3 (29% of the sample) had a < 4% incidence of major complications, and no deaths. In contrast, those with scores of ≤ 4 had a > 50% risk of major complications, including a 14% mortality rate. Despite the relatively low prevalence of scores ≤ 4 (4% of the cohort), the c-statistic of 0.72 suggests that the score has good overall discriminative ability.

There remain several limitations to our findings. First, this score was tested only at a single, large, teaching hospital. Although we expect the score to be generalizable, validity in other settings has not yet been established. Second, although there is a strong association between surgical score and risk of major complications, the confidence intervals around the risk estimates for any individual score remain wide. Additional validation with a larger cohort would be necessary to define the precise risk associated with a particular score. Third, the surgical score was tested only in general and vascular surgery patients aged 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.

Finally, the score might have sources of considerable measurement variability. We took data for the score from written anesthesia records, in which anesthesiologists attempt to record vital sign information every 5 minutes. These are inevitably incomplete records; for any given patient, data can be missing or smoothed to disregard perceived artifacts in readings.[20] The score might also not translate easily to settings with electronic anesthesia records, which record data continuously or every few seconds.[21] Scoring using electronic anesthesia records would have to disregard artifacts and momentary fluctuations in heart rate and blood pressure recordings (we are currently developing an algorithm to accomplish this).

Blood loss estimation can be similarly imprecise, but the broad categories used to calculate the score (< 100 mL, 100 to 600 mL, 601 to 1,000 mL, > 1,000 mL) are well within observers' range of precision in careful volu-

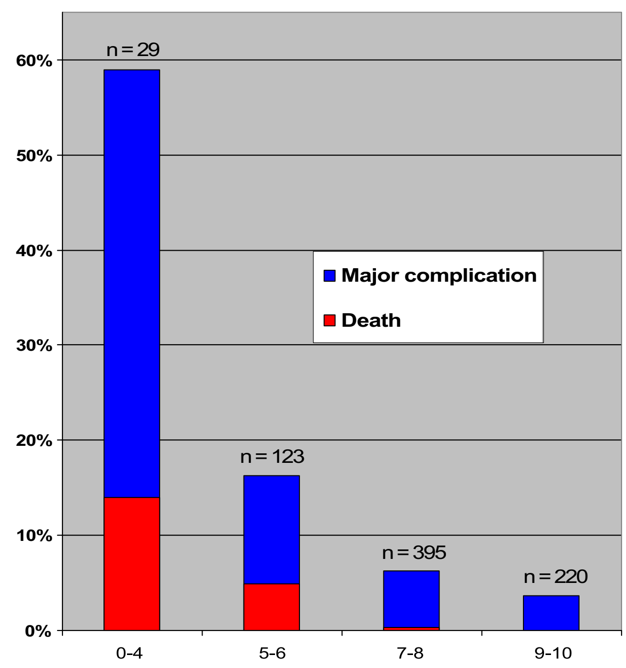

Figure 1. Thirty-day mortality and major complications for 767 patients undergoing general or vascular surgery, in relation to surgical scores. *p < 0.0001 for an association between surgical score and major complications, and for an association between surgical score and death.

metric studies.[22,23] Relying on anesthesiologists' independent estimation improves the reliability and insulates against surgeon bias.[22] Nonetheless, some inaccuracy in blood loss estimation can remain.

Still, the variables in the surgical 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).[24] The surgical score is just as easily calculated and as predictive of later patient outcomes. Our results demonstrate that a simple clinimetric surgical outcomes score can be derived from intraoperative data alone. This 10-point score based on the lowest heart rate, lowest MAP, and EBL discriminates well between groups of patients at high and low risk of major complications and death within 30 days of operation. Whether this surgical score will prove to be as useful as Virginia Apgar's obstetrical score for routine care, quality improvement, and research remains to be seen. It is certainly our hope.

Author Contributions

Study conception and design: Gawande, Kwaan, Regenbogen, Lipsitz, Zinner

Acquisition of data: Gawande, Kwaan, Regenbogen

Analysis and interpretation of data: Gawande, Kwaan, Regenbogen, Lipsitz, Zinner

Drafting of manuscript: Gawande, Kwaan, Regenbogen, Lipsitz, Zinner

Acknowledgment: We are indebted to Ms Tess Panizales and Ms Jill Steinberg for assistance with data collection and the BWH-NSQIP database and to Dr John Orav for his statistical advice. Dr Gawande had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

REFERENCES

1. Apgar V. A proposal for a new method of evaluation of the newborn infant. Curr Res Anesth Analg 1953;32:260–267.
2. Apgar V, Holaday DA, James LS, Weisbrot IM. Evaluation of the newborn infant—second report. JAMA 1958;168:1985–1988.
3. Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the assessment of newborn infants. N Engl J Med 2001;344:467–471.
4. Knaus WA, Zimmerman JE, Wagner DP, et al. APACHE—acute physiology and chronic health evaluation: a physiologically based classification system. Crit Care Med 1981;9:591–597.
5. Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. Br J Surg 1991;78:356–360.
6. Gawande AA, Thomas EJ, Zinner MJ. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. Surgery 1999;126:66–75.
7. Khuri SF, Daley J, Henderson WG, et al. The National VA Surgical Risk Study. Risk adjustment for the comparative assessment of the quality of surgical care. J Am Coll Surg 1995;180:519–531.
8. Fink AS, Campbell DA, Mentzer RM, et al. The National Surgical Quality Improvement Program in non-Veterans Administration hospitals: initial demonstration of feasibility. Ann Surg 2002;236:344–354.
9. Pasternak LR. Preanesthesia evaluation of the surgical patient. ASA Refresher Courses in Anesthesiology 1996;24:205–219.
10. Terrin N, Schmid CH, Griffith JL, et al. External validity of predictive models: a comparison of logistic regression, classification trees, and neural networks. J Clin Epidemiol 2003;56:721–729.
11. Hartmann B, Junger A, Röhrig R, et al. Intra-operative tachycardia and perioperative outcome. Langenbecks Arch Surg 2003;388:255–260.
12. Reich DL, Bennett-Guerrero E, Bodian CA, et al. Intraoperative tachycardia and hypertension are independently associated with adverse outcome in noncardiac surgery of long duration. Anesth Analg 2002;95:273–277.
13. Mong TG, Saini V, Weldon BC, Sigl JC. Anesthetic management and one-year mortality after noncardiac surgery. Anesth Analg 2005;100:4–10.
14. Gatch WD, Little WD. Amount of blood lost during some of the more common operations. JAMA 1924;83:1075–1076.
15. Holland AJ, Bell R, Ibach EG, et al. Prognostic factors in elective aortic reconstructive surgery. Aust NZ J Surg 1998;68:16–20.
16. Wolters U, Wolf T, Stützer H, Schröder T. ASA classification and perioperative variables as predictors of postoperative outcome. Brit J Anesthesia 1996;77:217–222.
17. Pasternak LR. Risk assessment in ambulatory surgery: challenges and new trends. Can J Anesthesia 2004;51:R1–R5.
18. Mangano DT, Layug EL, Wallace A, Tateo I. Effect of atenolol on mortality and cardiovascular morbidity after noncardiac surgery. N Engl J Med 1996;335:1713–1720.
19. Auerbach A, Goldman L. β-blockers and reduction of cardiac events in noncardiac surgery: scientific review. JAMA 2002;287:1435–1444.
20. Reich DL, Wood RK, Mattar R, et al. Arterial blood pressure and heart rate discrepancies between handwritten and computerized anesthesia records. Anesth Analg 2000;91:612–616.
21. Cook RI, McDonald JS, Nunziata E. Differences between handwritten and automatic blood pressure records. Anesthesiology 1989;71:385–390.
22. Delilkan AE. Comparison of subjective estimates by surgeons and anaesthetists of operative blood loss. Br Med J 1972;2:619–621.
23. Gardiner AJS, Dudley HAF. The measurement of blood loss at operation. Br J Anesth 1962;34:653–656.
24. Finster M, Wood M. The Apgar score has survived the test of time. Anesthesiology 2005;102:855–857.