

Expansion of the Surgical Apgar Score across All Surgical Subspecialties as a Means to Predict Postoperative Mortality

Paul Q. Reynolds, M.D.,* Neal W. Sanders, Ph.D.,† Jonathan S. Schildcrout, Ph.D.,‡ Nathaniel D. Mercaldo, M.S.,§ Paul J. St. Jacques, M.D.||

ABSTRACT

Background: A surgical scoring system, akin to the obstetrician's Apgar score, has been developed to assess postoperative risk. To date, evaluation of this scoring system has been limited to general and vascular services. The authors attempt to externally validate and expand the Surgical Apgar Score across a wide breadth of surgical subspecialties.

Methods: Intraoperative data for 123,864 procedures including all surgical subspecialties were collected and associated with Surgical Apgar Scores (created by the summation of point values associated with the lowest mean arterial pressure, lowest heart rate, and estimated blood loss). Patients' death records were matched to the corresponding score, and logistic regression models were created in which mortality within 7, 30, and 90 days was regressed on the Apgar score.

Results: Lower Surgical Apgar Scores were associated with an increased risk of death. The magnitude of this association varied by subspecialty. Some subspecialties exhibited higher odds ratios, suggesting that the score is not as useful for them. For most of the subspecialties the association between the Apgar score and mortality decreased as the time since surgery increased, suggesting that predictive ability ceases to be helpful over time. After

What We Already Know about This Topic

- The Surgical Apgar Score, a simple 0–10 score based on blood loss, lowest blood pressure, and lowest heart rate during surgery, predicts mortality in general and vascular surgery patients.

What This Article Tells Us That Is New

- In a review of more than 120,000 patients, the Surgical Apgar Score correlated with risk of death across many subspecialties, although the strength of the correlation varied.

adjusting for the patient's American Society of Anesthesiologists classification, Apgar scores remained associated with death among most of the subspecialties.

Conclusion: A previously published methodology for calculating risk among general and vascular surgical patients can be applied across many surgical services to provide an objective means of predicting and communicating patient outcomes in surgery as well as planning potential interventions.

CLINICIANS have a need for predictive tools to assess perioperative risk. Several algorithms have been used or developed for risk stratification such as the American Society of Anesthesiologists Physical Status Classification System (ASA classification), the Physiologic and Operative Severity Score for enUmeration of Mortality and morbidity (POSSUM), the Acute Physiology and Chronic Health Evaluation (APACHE), and the Simplified Acute Physiology Score (SAPS). However, each of these systems has limitations and restricted uses. The ASA classification was initially intended as a means to stratify a patient's systemic illness but not

* Resident, Department of Anesthesiology, Vanderbilt University School of Medicine, Nashville, Tennessee. † Assistant, Center for Perioperative Research in Quality, Vanderbilt University School of Medicine. ‡ Assistant Professor, Department of Biostatistics, Vanderbilt University School of Medicine. § Biostatistician III, Department of Biostatistics, Vanderbilt University School of Medicine. || Associate Professor, Department of Anesthesiology, Vanderbilt University School of Medicine.

Received from the Department of Anesthesiology, Vanderbilt University School of Medicine, Nashville, Tennessee. Submitted for publication September 21, 2010. Accepted for publication February 1, 2011. Support was provided solely from institutional and/or departmental sources. Dr. St. Jacques is a minority shareholder in Acuitec, LLC, Birmingham, Alabama. Acuitec licenses the perioperative information management system used and referred to in this study. Previously presented, in part, as a poster abstract at the American Society of Anesthesiologists Annual Meeting, October 17–21, 2009, New Orleans, Louisiana.

Address correspondence to Dr. Reynolds: Vanderbilt University Medical Center, Department of Anesthesiology, 1301 Medical Center Drive, 4648 The Vanderbilt Clinic, Nashville, Tennessee 37232-5614. paul.q.reynolds@vanderbilt.edu. This article may be accessed for personal use at no charge through the Journal Web site, www.anesthesiology.org.

Copyright © 2011, the American Society of Anesthesiologists, Inc. Lippincott Williams & Wilkins. Anesthesiology 2011; 114:1305–12

◇ This article is featured in "This Month in Anesthesiology." Please see this issue of ANESTHESIOLOGY, page 9A.

◆ This article is accompanied by an Editorial View. Please see: Gawande AA, Regenbogen SE: Critical need for objective assessment of postsurgical patients. ANESTHESIOLOGY 2011; 114:1269–70.

⊕ Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are available in both the HTML and PDF versions of this article. Links to the digital files are provided in the HTML text of this article on the Journal's Web site (www.anesthesiology.org).

postoperative risk, and it has been criticized for being too simple. It does not account for the age of the patient, nature of the surgical procedure, anesthetic technique, competency or training level of the surgical team, or duration of surgery. Although the ASA classification has proved to be a predictive preoperative risk factor in mortality models, its subjective nature and inconsistent scoring between providers make it less than ideal for performing evidence-based postoperative risk calculation.¹⁻⁴ The POSSUM, APACHE, and SAPS and their later derivations (portsmouth POSSUM, colorectal POSSUM, APACHE II and III, and SAPS II) are more accurate and objective predictive algorithms, but not all of the variables needed are easily and consistently attainable in an operating room setting, making them more practical in their initially intended role as critical care auditing tools rather than predictive tools.^{5,6}

Recently, Gawande *et al.* designed a more ideal postoperative scoring system, the Surgical Apgar Score, patterned after the Apgar obstetric scoring system. Similar to the obstetric scoring system, the Surgical Apgar Score uses a three-item [estimated blood loss (EBL), lowest intraoperative heart rate (HR), and lowest intraoperative mean arterial pressure (MAP)] aggregate to risk-stratify patients undergoing surgery in the postoperative setting.⁷ Gleaned from an initial dataset of 28 variables, these three values were each found to be independent predictors of outcomes. Strengths of the Surgical Apgar Score include the ability to calculate the score quickly and objectively. The provider could then anticipate the need for further or more aggressive interventions or use the data in benchmarking centers by predicted *versus* observed scores. Ultimately, the score may also prove useful in guiding preventive strategies such as optimizing intraoperative heart rate or blood pressure. To date, Gawande's Surgical Apgar Score has been studied in colorectal surgery, vascular surgery, and certain gynecologic and urologic procedures.⁷⁻¹¹

The burgeoning literature on the Surgical Apgar Score also identifies potential weaknesses of the scoring system. For example, calculation of the score relies on EBL, which critics have often tagged as imprecise. However, previous studies have shown that the broad categories used to calculate the Surgical Apgar Score (0–100 ml, 101–600 ml, 60–1,000 ml, >1,000 ml) are easily within observers' range of precision.^{12,13} Another hypothetical weakness lies in the fact that intraoperative hemodynamics may be affected by anesthetic medications and interventions such as induction and intubation, and therefore alter the computation of the Surgical Apgar Score. For example, a transient episode of hypotension associated with anesthetic induction would be treated the same as prolonged hypotension and result in a lower (worse) Surgical Apgar Score. On the other hand, a transient bradycardic episode would contribute to a higher (better) score. Nevertheless, several studies demonstrate that persistent heart rate elevation and hypotension are strongly associated with poorer outcomes, regardless of their cause.¹⁴⁻¹⁶ Finally, other potentially predictive perioperative variables, such as coronary artery disease, intraoperative blood transfusion, ASA

class, sex, volume of intravenous fluids administered, patient age, surgical time, functional status, renal function, and chronic steroid use are excluded from the Surgical Apgar Score. The exclusion of these potentially predictive preoperative risk factors could be construed as a weakness of the score. However, as previously mentioned, an important aspect of the usefulness of the Surgical Apgar Score is its simplicity.

Using data collected through our electronic Perioperative Information Management System from January 1, 2006 to December 31, 2009, we attempt to externally validate and expand the Surgical Apgar Score as a means for predicting postoperative mortality across a wide range of surgical subspecialties and to determine whether this score provides additional information beyond that of the patient's ASA status.

Materials and Methods

Perioperative data for 123,864 procedures including all surgical subspecialties (as defined by the primary service of the attending surgeon) were collected with an electronic Perioperative Information Management System (VPIMS; Acuitec LLC, Birmingham, AL) from January 1, 2006 to December 31, 2009. The data are stored in a perioperative data warehouse, using Microsoft SQL server technology (Microsoft Corporation, Redmond, WA). Preoperative and postoperative data were excluded. All data were collected during normal operations and were retrospectively reviewed. As a matter of routine care, mortality data are obtained from the US Social Security Death Index and linked to patients in the database using both the social security number and the date of birth. Both items have to match to be recognized as a correct linkage. In this study there were eight patients who had data indicating a date of death before the surgery date; when these cases were investigated, it was discovered that they were incorrectly identified in the perioperative electronic record. This error rate is approximately 1 in 10,000 cases. As the queries performed for the study resulted in deidentified information only, the study met criteria for non-human research and was performed with approval by our institutional review board (Vanderbilt University Human Research Protection Program, Nashville, TN).

From the captured intraoperative data, estimated blood loss, blood pressure, and heart rate were analyzed. If EBL was not recorded or was defined as "minimal," it was assumed to be less than 100 ml. MAP was derived from the electronically captured invasive or noninvasive blood pressures, with preference for invasive, unless there was a manually entered provider-override value. HR was derived in order of preference from the provider's manually-entered HR, the electrocardiogram, or the pulse oximeter, depending on availability. As defined by the original article by Gawande *et al.*⁷ on the Surgical Apgar Score, point values associated with the lowest MAP, lowest HR, and EBL were summated to produce the Surgical Apgar Score (table 1).

Table 1. Computing the Surgical Apgar Score

Parameters	Number of Points				
	0	1	2	3	4
Estimated blood loss, ml	>1,000	601–1,000	101–600	1 ≤ 100	—
Lowest mean arterial pressure, mmHg	<40	40–54	55–70	≥70	—
Lowest heart rate/min	>85	76–85	66–75	56–65	≤55

The Surgical Apgar Score is calculated at the end of the 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.

Study Cohort

Cases recorded as ASA 6 were excluded to remove organ donor cases. Patients younger than 18 yr old were also excluded. Death information was verified before September 1, 2009; thus, any patient whose surgery occurred after this date was excluded (9.8%). Fifteen subjects with invalid data such as negative or null days to death were also excluded. All remaining cases recorded in the database were initially included in the data export, but after further review of the specific subspecialties 9 were identified as nonsurgical (e.g., epidurals and bronchoscopies, n = 397), and 7 were associated with low-volume community physician groups that do not regularly operate at our institution but infrequently perform a case for various reasons (n = 1,040). Data from these procedures, along with those that were listed as “unknown” (n = 8,426; unknown cases represent labor and delivery epidurals, offsite procedures in an ambulatory surgery center, bedside cases in the intensive care units, and sedation for gastrointestinal and radiology procedures) were excluded from this analysis (8.8%). See table 2 for a complete breakdown of the omitted categories.

Patient characteristics and surgical summaries (year and operating room duration) were tabulated across the entire dataset. Categorical variables were represented as percentages

Table 2. Excluded Specialties

Specialty	Count
Anesthesiology	54
Bone marrow transplant	69
Cardiology	21
Community dentistry	3
Community general	2
Community gynecology	10
Community ophthalmology	313
Community orthopedic	258
Community plastic	350
Community podiatry	104
Dentistry	39
Dermatology	4
Gastroenterology	125
Obstetrics	39
Pulmonary	29
Radiology	17
Unknown	8,426

Nonsurgical cases (bronchoscopies and epidurals) as well as those associated with low-volume community physician groups or unknown specialties were excluded from the analysis.

and counts whereas continuous variables were summarized by the 5th, 25th, 50th (median), 75th, and 95th quantiles. The patient characteristics were also divided by survey year to determine whether the patient population varied by year (data not presented but available from the authors).

Table 3 summarizes the patient characteristics. Within this adult population, the median (interquartile range) age was 51 yr (38–63 yr). Most of these patients were classified as either ASA 2 or 3 (42.6 and 40.3%, respectively). Mortality after 7, 30, and 90 days was 0.6%, 1.5%, and 2.7%, respectively.

Statistical Analysis

To quantify the association between the Surgical Apgar Score and mortality, a series of regression models was developed. For each surgical subspecialty, three logistic regression models were estimated in which the Surgical Apgar Score was separately regressed on a 7-, 30-, or 90-day mortality. Seven-day mortality was defined as a binary variable that equaled 1 if the subject died

Table 3. Patient and Surgery Characteristics

Characteristics	All Services (No. = 101,907)
Surgery year	
2006	25.9% (26,404)
2007	27.3% (27,828)
2008	27.9% (28,433)
2009	18.9% (19,242)
Age (yr)*	22.0: 38.0: 51.0: 63.0: 78.0
Sex†	—
Male	51.1% (51,728)
Female	48.9% (49,576)
ASA status	—
1	6.3% (6,428)
2	42.6% (43,414)
3	40.3% (41,100)
4	10.4% (10,600)
5	0.4% (365)
ASA emergency status	—
Yes	5.3% (5,406)
OR duration (min)*	44.0: 82.0: 138.0: 222.0: 390.0
Overall mortality	—
Day: 7	0.6% (593)
30	1.5% (1,522)
90	2.7% (2,748)

* Continuous variables: Q₅: Q₂₅: Q₅₀: Q₇₅: Q₉₅. † 603 patients were sex indeterminant (n = 101,304).

ASA Status = American Society of Anesthesiologists Physical Status Classification; OR = operating room.

Table 4. Mortality and Apgar Distributions by Primary Service

Primary Service	N	Deaths (%)			Apgar Score (%)										
		7	30	90	0	1	2	3	4	5	6	7	8	9	10
Burn	1,558	2.7	5.8	10.3	0.0	0.3	1.7	3.8	8.3	16.1	20.6	21.9	17.8	8.3	1.3
Cardiac	3,513	2.2	4.1	6.0	0.0	0.0	0.1	0.9	2.8	6.3	11.3	26.3	35.0	15.8	1.5
Emergency	2,384	2.0	5.2	8.9	0.0	0.1	0.7	1.6	6.1	13.7	18.6	23.3	21.6	12.0	2.2
General oncology	5,513	0.3	1.3	2.8	0.0	0.1	0.2	0.5	2.2	5.6	14.5	27.0	31.7	16.0	2.3
General surgery	8,207	0.3	1.2	2.3	0.0	0.0	0.2	0.5	2.3	6.9	14.2	25.3	29.5	18.1	3.0
Gynecology	5,511	0.1	0.5	0.9	0.0	0.0	0.3	0.6	2.2	6.2	15.6	28.8	31.2	13.7	1.4
Head and neck	2,276	0.1	0.8	2.7	0.0	0.0	0.1	0.7	2.2	6.7	16.5	27.3	29.8	14.9	1.7
Liver transplant	1,217	1.2	3.0	4.6	0.6	1.5	3.9	7.5	10.8	13.6	17.4	18.9	15.4	8.9	1.6
Neurosurgery	7,589	1.0	2.2	4.2	0.0	0.0	0.2	0.8	2.9	7.2	14.3	24.7	30.4	17.0	2.6
Ophthalmology	4,203	0.1	0.3	0.6	0.0	0.0	0.0	0.0	0.3	2.1	8.5	19.8	32.2	28.1	8.9
Oral	1,308	0.4	0.9	1.5	0.0	0.0	0.1	1.1	4.7	10.8	20.4	24.9	23.1	13.0	1.9
Orthopedic	5,988	0.3	1.2	2.5	0.1	0.1	0.3	1.5	4.5	9.7	17.6	25.7	26.1	12.6	1.7
Orthopedic trauma	8,495	0.5	1.0	1.5	0.0	0.2	1.1	3.5	9.5	16.4	21.0	22.3	17.0	7.6	1.2
Ortho sports/hand	11,459	0.1	0.2	0.3	0.0	0.0	0.1	0.2	0.9	4.8	13.0	24.5	31.2	20.8	4.5
Otolaryngology	4,898	0.2	0.6	1.2	0.0	0.0	0.0	0.2	1.6	5.7	14.7	26.1	31.8	17.5	2.4
Plastics	5,862	0.2	0.7	1.4	0.0	0.1	0.2	0.5	2.8	7.1	16.0	28.0	29.9	13.9	1.5
Renal	1,775	0.0	0.4	1.3	0.1	0.0	0.1	0.5	3.0	7.4	17.2	25.0	27.7	15.8	3.3
Thoracic	2,370	0.9	3.3	7.1	0.0	0.1	0.3	1.6	2.9	10.3	17.4	25.4	25.8	13.8	2.5
Trauma	3,491	2.8	5.8	7.5	0.3	1.3	2.3	5.2	12.8	19.0	18.3	18.6	13.9	7.4	1.0
Urology	10,343	0.1	0.5	1.4	0.0	0.0	0.1	0.4	1.4	4.4	11.9	23.9	33.3	21.4	3.2
Vascular	3,947	1.2	3.3	6.0	0.0	0.1	0.5	0.8	3.1	7.9	15.1	23.0	26.4	17.7	5.4
All services	101,907	0.6	1.5	2.7	0.0	0.1	0.4	1.1	3.5	8.0	15.2	24.6	28.2	16.1	2.8

Due to rounding, some of the entries that are represented as percentages are labeled 0.0 and may either be 0 or simply less than 0.5%.

within 7 days after surgery and 0 otherwise. Similar variables were calculated for 30 and 90 days. The Surgical Apgar Score was included in the model as a continuous variable, which assumes that there is a linear relationship between the score and the log odds of death. This assumption was assessed by fitting additional logistic models in which the Surgical Apgar Score was modeled with restricted cubic splines and comparing the models using the likelihood ratio test. The unadjusted odds ratio (OR) and its 95% confidence interval (CI) were then reported for each subspecialty. The OR estimates correspond to the fold-change in the odds of 7-, 30-, or 90-day mortality for each unit increase in the Surgical Apgar Score. As the OR was used to represent the measure of association, values closer to 1 are consistent with weaker associations and values farther away from 1 (either closer to 0 or greater than 1) are consistent with stronger associations. These estimates were then transformed to determine the probability of death at each day of interest and are displayed graphically.

It was also of interest to determine whether the Surgical Apgar Score provided additional information beyond that of a patient's ASA physical status. To address this, a second set of regression models was created in which ASA status was added as a covariate. ASA status was modeled as a linear term, and interactions with the Surgical Apgar Score were not included because of a lack of power among some of the subspecialties. The adjusted estimates associated with the Surgical Apgar Score were transformed and reported in a manner similar to that previously described. All analyses were performed in R version 2.11 (Vienna, Austria).¹⁷

Results

Table 4 summarizes the distributions of mortality and Surgical Apgar Scores by surgical subspecialty. For example, there were 1,558 surgeries involving burn patients. Of these patients, 2.7% died by day 7 and 5.8% by day 30. Approximately 2% had a Surgical Apgar Score ≤ 2 , and nearly 10% had a score of ≥ 9 . The most common subspecialties included orthopedic sports/hand (11%), urology (10%), orthopedic trauma (8%), general surgery (8%), and neurosurgery (7%). Mortality rates ranged from 0% (renal, day 7) to 10.3% (burn, day 90) and were the highest among burn, cardiac, emergency, trauma, and vascular patients. These rates were generally low within the first 7 days after surgery and remained low through day 90 in a few of the subspecialties (ophthalmology, oral, and renal). However, analyses of these subspecialties were still performed despite being underpowered to facilitate interpretation of the association between the Surgical Apgar Score and mortality. The distributions of Surgical Apgar Scores were similar across surgical subspecialties in that most of the scores were between 6 and 9, but the frequencies of low Surgical Apgar Scores (≤ 2) varied slightly. For example, low Surgical Apgar Scores were especially prominent among patients undergoing either liver transplantation or trauma procedures.

Lower Surgical Apgar Scores were associated with an increased risk of death (table 5). For example, the probability of death by day 90 was 32% (fig. 1B, 95% CI: 0.23–0.42) among cardiac patients with a Surgical Apgar Score of 2 and only 4% (0.03–0.05) among those with a score of 8. The

Table 5. Associations between Apgar and Mortality by Primary Service

Primary Service	Day 7		Day 30		Day 90	
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
Burn	0.85 (0.72, 1.01)	1.02 (0.85, 1.22)	0.91 (0.81, 1.03)	1.09 (0.96, 1.24)	0.88 (0.80, 0.96)	1.05 (0.95, 1.17)
Cardiac	0.61 (0.53, 0.70)	0.66 (0.57, 0.76)	0.64 (0.58, 0.72)	0.68 (0.61, 0.75)	0.67 (0.61, 0.73)	0.70 (0.63, 0.76)
Emergency	0.71 (0.61, 0.84)	1.00 (0.83, 1.20)	0.75 (0.67, 0.83)	1.01 (0.89, 1.14)	0.67 (0.62, 0.73)	0.86 (0.78, 0.95)
General oncology	0.45 (0.35, 0.57)*	0.62 (0.48, 0.80)*	0.58 (0.51, 0.67)	0.72 (0.62, 0.83)	0.59 (0.53, 0.65)	0.69 (0.63, 0.77)
General surgery	0.60 (0.47, 0.76)	0.82 (0.65, 1.04)*	0.58 (0.52, 0.66)	0.74 (0.65, 0.84)	0.60 (0.55, 0.66)	0.73 (0.66, 0.80)
Gynecology	0.54 (0.33, 0.89)*	0.72 (0.45, 1.15)*	0.58 (0.46, 0.72)	0.72 (0.57, 0.89)*	0.59 (0.49, 0.69)	0.69 (0.58, 0.82)
Head and neck	0.51 (0.26, 0.98)*	0.75 (0.34, 1.65)*	0.64 (0.48, 0.87)*	0.76 (0.56, 1.03)*	0.66 (0.56, 0.78)	0.73 (0.62, 0.87)
Liver transplant	0.59 (0.46, 0.76)*	0.69 (0.52, 0.92)*	0.76 (0.66, 0.89)	0.86 (0.72, 1.02)	0.81 (0.72, 0.92)	0.88 (0.76, 1.01)
Neurosurgery	0.59 (0.52, 0.67)	0.71 (0.62, 0.81)	0.71 (0.64, 0.78)	0.82 (0.75, 0.90)	0.81 (0.75, 0.87)	0.90 (0.84, 0.97)
Ophthalmology	0.22 (0.10, 0.49)*	0.36 (0.17, 0.73)*	0.49 (0.33, 0.73)*	0.57 (0.39, 0.84)*	0.58 (0.43, 0.78)	0.67 (0.50, 0.90)*
Oral	0.72 (0.41, 1.28)*	0.78 (0.45, 1.37)*	0.76 (0.53, 1.11)*	0.83 (0.57, 1.19)*	0.87 (0.64, 1.17)*	0.93 (0.69, 1.24)*
Orthopedic	0.64 (0.51, 0.82)	0.70 (0.54, 0.91)	0.69 (0.60, 0.79)	0.75 (0.65, 0.87)	0.73 (0.66, 0.80)	0.79 (0.72, 0.88)
Orthopedic trauma	0.76 (0.64, 0.90)	0.93 (0.77, 1.11)	0.88 (0.77, 0.99)	1.05 (0.92, 1.19)	0.89 (0.81, 0.99)	1.06 (0.95, 1.18)
Ortho sports/hand	0.49 (0.34, 0.72)*	0.68 (0.46, 0.99)*	0.51 (0.39, 0.67)	0.71 (0.54, 0.93)*	0.60 (0.48, 0.74)	0.78 (0.63, 0.97)
Otolaryngology	0.36 (0.24, 0.54)*	0.44 (0.29, 0.66)*	0.47 (0.37, 0.61)	0.56 (0.44, 0.73)*	0.55 (0.46, 0.65)	0.63 (0.53, 0.76)
Plastics	0.45 (0.32, 0.64)*	0.59 (0.42, 0.84)*	0.60 (0.50, 0.72)	0.77 (0.64, 0.92)	0.68 (0.60, 0.79)	0.85 (0.74, 0.98)
Renal	—*	—*	0.66 (0.42, 1.03)*	0.75 (0.48, 1.16)*	0.67 (0.52, 0.86)	0.72 (0.55, 0.92)*
Thoracic	0.72 (0.56, 0.93)	0.84 (0.64, 1.11)*	0.73 (0.63, 0.84)	0.85 (0.74, 0.99)	0.80 (0.72, 0.88)	0.90 (0.81, 1.00)
Trauma	0.68 (0.61, 0.75)	0.85 (0.76, 0.95)	0.77 (0.72, 0.83)	0.93 (0.86, 1.01)	0.80 (0.75, 0.85)	0.96 (0.89, 1.03)
Urology	0.86 (0.58, 1.27)*	0.92 (0.63, 1.35)*	0.66 (0.55, 0.78)	0.74 (0.63, 0.88)	0.59 (0.53, 0.65)	0.65 (0.59, 0.72)
Vascular	0.51 (0.44, 0.60)	0.60 (0.51, 0.71)	0.61 (0.55, 0.67)	0.71 (0.64, 0.79)	0.67 (0.62, 0.73)	0.75 (0.69, 0.82)

The odds ratio was used as the measure of association between Surgical Apgar Score and mortality, with values closer to 1 consistent with weaker associations and values away from 1 consistent with stronger associations. Unadjusted and ASA adjusted (simply labeled "Adjusted") estimates associated with the Surgical Apgar Score are presented. An estimate is not provided for the renal subgroup because no deaths were reported by day 7. *P* values were omitted but can be inferred to be less than 0.05 as long as the 95% confidence interval does not contain 1.

* These subspecialties did not have at least 20 (unadjusted) or 30 (adjusted) deaths and hence are potentially underpowered to detect associations between the Apgar score and mortality.

ASA = American Society of Anesthesiologists; CI = confidence interval; OR = odds ratio.

magnitude of this association varied by subspecialty. Only a weak relationship was noted between Surgical Apgar Score and death among burn patients.

For each unit increase in Surgical Apgar Score, the unadjusted risk of death by day 7 decreased by nearly 50% among those undergoing vascular surgery (table 5, 95% CI: 0.44–0.60) but only decreased by 30% among emergency general surgery patients. Similar patterns were observed when modeling death at days 30 and 90. For most of the subspecialties the association between the Surgical Apgar Score and mortality generally decreased as the time since surgery increased (*e.g.*, neurosurgery: 0.59 to 0.71 to 0.81 OR at days 7, 30, and 90), suggesting that at some future point postoperatively the predictive ability ceases to be helpful. However, the relationship between the Surgical Apgar Score and mortality day was also maintained in a few of the subgroups (*e.g.*, cardiac, general surgery, and gynecology) and even strengthened over time in others (*e.g.*, urology). This may simply be an artifact because of sparse data among certain subspecialties, but it could also indicate that the Surgical Apgar Score is a better predictor for early death in some subspecialties and later death in others. After adjusting for ASA status, similar associations, although some were attenuated, were observed between the Surgical Apgar Score and mortality.

Figure 1 summarizes the relationship between the Surgical Apgar Score and the probability of death (at days 7, 30, and 90) for several subspecialties. See figure, Supplemental Digital Con-

tent 1, <http://links.lww.com/ALN/A739>, for additional subspecialties. The *points* correspond to the estimated probability of death, whereas the *vertical bars* correspond to the 95% CI. Points and bars were omitted if there were less than 10 observations for a given Apgar value, whereas only the bars were omitted if there were at least 10 observations but no deaths. Among cardiac patients, there were no Surgical Apgar Scores of 0 (hence no points or bars) and only four patients with a Surgical Apgar Score of 2 (but no deaths at day 7, thus no bar).

Discussion

Our analysis of the risk of postsurgical death at 7, 30, and 90 days expands on the original work published by Gawande *et al.* and others by performing the analysis of the Surgical Apgar Score across all major surgical subspecialties at an academic medical center. As also shown by Regenbogen *et al.*, we have established that the predictive value of the Surgical Apgar Score remains valid at an institution outside of where it was developed and that it can be derived from electronic records.⁸ In addition, we have demonstrated that the magnitude of the relationship of Surgical Apgar Score to the risk of postsurgical death at 7, 30, and 90 days varies by surgical subspecialty and that it contains information above and beyond that of the ASA metric. This difference between subspecialties is important to note and may occur because co-

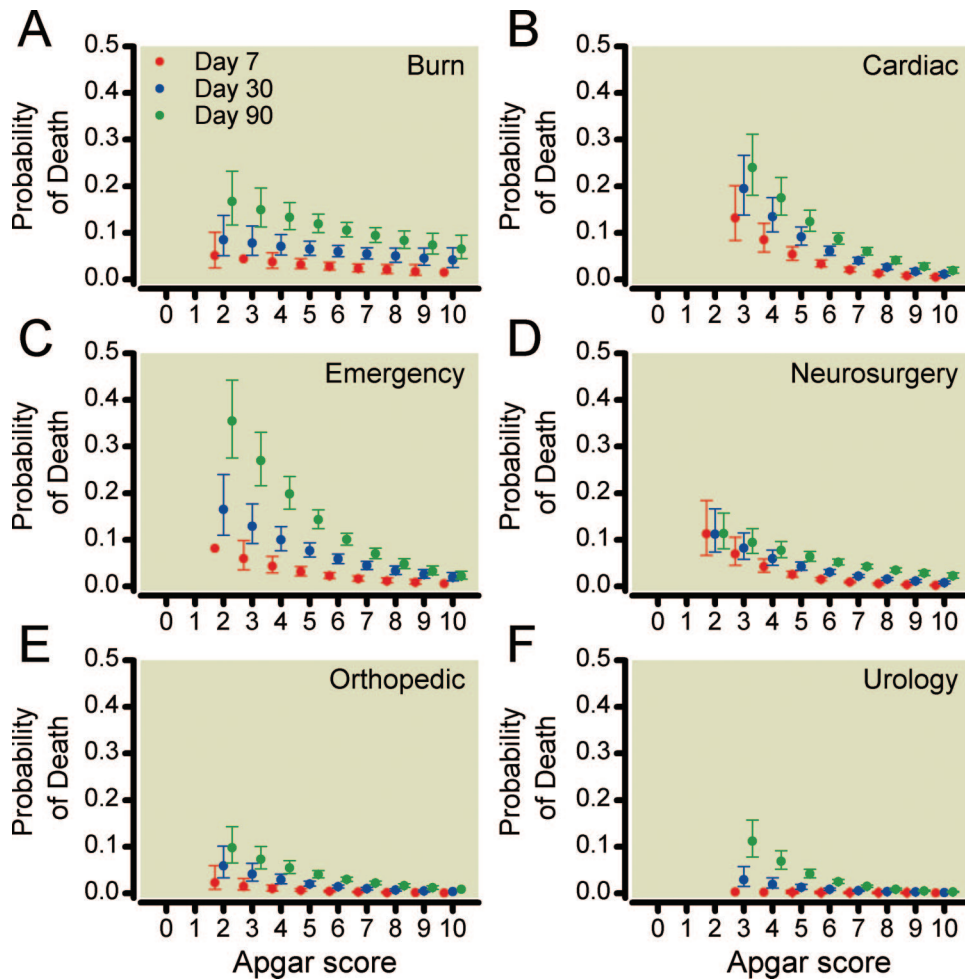


Fig. 1. (A–F) Relationship between Surgical Apgar Score and the probability of death at day 7, 30, and 90 by specialty. Points correspond to the estimated probability of death, whereas the vertical bars correspond to the 95% confidence intervals. Points and bars were omitted if there were less than 10 observations for a given Apgar value, whereas only the bars were omitted if there were at least 10 observations but no deaths.

morbidities and potential causes of death vary among the subspecialty populations. For example, major cardiac events in vascular patients may be well predicted by the Surgical Apgar Score as opposed to sepsis in burn patients.

The Surgical Apgar Score remains a simple, easily calculable score immediately postprocedure for assessing postoperative risk of death. The score may have usefulness in several areas. For example, during the handoff process (the communication between physician services or physician and nursing team members) it can signal the provider taking over care to the overall risk the patient is facing and may indicate the need for additional care measures to minimize risk. The Surgical Apgar Score could be incorporated into electronic documentation packages for real-time calculation either during or at the end of surgery, providing an automated warning to clinicians. This prognostic value may alert the provider that additional diagnostic testing (arterial blood gases, serum lactate, or hematocrit determinations), further resuscitation, one-on-one nursing, or more invasive monitoring is indicated. In fact, several proven risk modification strategies

(such as deep breathing and thoracic epidurals for high risk pulmonary patients, or intraoperative β -blockade for vascular procedures in patients with coronary artery disease) exist in the literature, which suggests that early identification of high-risk patients and implementation of risk modification strategies can decrease hospital stay and mortality.^{18–20} Frequently, decisions to transfer patients to intensive care units are based on clinical impressions rather than hard data. Although improving outcomes through postoperative interventions based on the Surgical Apgar Score is only speculative at this point, the score does provide an objective adjunct to facilitate discussions of the surgeon, anesthesiologist, and intensive care unit physician in determining the need for these further or heightened postoperative care strategies.

Beyond immediate patient care issues, quality improvement initiatives may also be augmented with data from the Surgical Apgar Score. In the event detection process, quality officers may choose to select cases with a low Surgical Apgar Score for additional screening and possible analysis for peer review or other improvement processes. Furthermore, focus-

ing attention on decreasing blood loss, decreasing maximum heart rate, and avoiding hypotension, the three factors that would increase a patient's Surgical Apgar Score, may lead clinicians to create quality and safety systems that are designed to reducing risk through preventing low Surgical Apgar Scores.

Unlike some of the previous studies of the Surgical Apgar Score, our study was limited to all-cause mortality as the primary endpoint. The decision to exclude secondary or surrogate endpoints such as major cardiovascular morbidity was made for several reasons: the difficulty in determining major negative outcomes across a wide variety of surgical services, the subjective and inconsistent reporting of such, and the cost-prohibitive nature of scrutinizing more than 100,000 charts. Furthermore, as shown in the PeriOperative Ischemic Evaluation (POISE) trial, there is literature suggesting that surrogate endpoints may not accurately predict mortality and therefore are not recommended for studies in which the aim is to reduce mortality.^{21,22,18} As such, our findings that the Surgical Apgar Score predicts risk of postoperative death across a wide variety of surgeries may not be applicable to the prediction of other postoperative complications across the same surgical services.

In addition, as described in previous studies, the Surgical Apgar Score is dependent on the preoperative physical condition of the patient. Patients who are hypotensive and tachycardic preoperatively are likely to have poorer Surgical Apgar Scores, potentially based on their preoperative condition rather than the success or difficulty of surgical intervention. This study's population, for example, comes from a large academic referral institution and likely treats patients who have more significant comorbidities than in an average community hospital. As such, the Surgical Apgar Score has not been validated in and of itself to compare physicians or institutions.

The original model of Gawande *et al.* was kept simple so that a human could compute the score. Although the simplicity of the original model is reasonable and in fact a major point of the Surgical Apgar Score, the broad adoption of automatic perioperative information systems could facilitate a potentially more complex but improved model. The additional complexity would be acceptable (if needed) because the score could be computed in real time using the computer. Furthermore, although the original Surgical Apgar Score is now validated across a wide range of surgical subspecialties, it may be possible that the algorithm could be modified for

Saager L, Greenwald S, Kelley S, Schubert A, Sessler D. Hospital Stay and Mortality Are Increased by a "Triple Low" of Blood Pressure, BIS & Anesthetic Level. Presented at the Annual Meeting of the American Society of Anesthesiologists, October 17–21, 2009, New Orleans, Louisiana; abstract number: A880.

** Stapelfeldt W, Greenwald S, Manberg P, Chamoun N, Sessler D. Modeling the Impact of Alert Thresholds of "Triple Low" Conditions on 90-Day Postoperative Mortality. Presented at the Annual Meeting of the American Society of Anesthesiologists, October 16–20, 2010, San Diego, California; abstract number: A922.

better prediction among each subspecialty. Indeed, Gawande *et al.* developed the score for vascular and general cases, and our data suggest that the Surgical Apgar Score is a better predictor of mortality in those subspecialties than it is in, for example, burn excision surgery.

Future work should be directed toward improving the Surgical Apgar Score and examining the usefulness of the score in guiding intraoperative techniques and postoperative interventions, such as intensive care unit admission or other escalation in diagnosis or therapy. As mentioned previously, a potential weakness of the algorithm used to compute the score is that it treats transient and prolonged heart rate and blood pressure fluctuations alike. The score could potentially be improved by excluding for the period surrounding induction, or adding a time factor to the heart rate and blood pressure parts of the algorithm. Furthermore, it would be beneficial to note how the Surgical Apgar Score compares with other current studies relating intraoperative cardiovascular and anesthetic patterns to longer term postoperative outcome. For example, a recent retrospective analysis by researchers at the Cleveland Clinic, Cleveland, Ohio found that patients who simultaneously experience low-normal MAP, bispectral index score, and end-tidal volatile anesthetic concentration have nearly triple the risk for 30-day mortality as those whose numbers are higher. The negative outcome was especially common in patients who spent more than 20 min in a triple-low state, and was ameliorated by vasopressors.^{##} Perhaps a combination of the two algorithms (HR, MAP, EBL, end-tidal volatile anesthetic concentration, and bispectral index score) would result in a more powerful predictive model. Both the Surgical Apgar Score and the triple-low state seem to suggest that patients with autonomic dysfunction, *i.e.*, fragile patients, have poorer outcomes. The two scores seemingly justify and explain the benefit of already-proven therapies such as maintaining a baseline MAP sufficient for end-organ perfusion, avoiding tachycardia with appropriate analgesia and β -blockade, and minimizing the preoperative fasting period with a carbohydrate-rich drink.^{18,23} Future work with these types of predictive models could provide insight into other beneficial intraoperative techniques and postoperative interventions. In so doing, evidence-based protocols may be developed that could potentially decrease morbidity, mortality, and costs.

References

1. Junger A, Engel J, Quinzio L, Banzhaf A, Jost A, Hempelmann G: [Risk predictors, scoring systems and prognostic models in anesthesia and intensive care: Part I: Anesthesia.] *Anesthesiol Intensivmed Notfallmed Schmerzther* 2002; 37:520–7
2. Dripps RD: New classification of physical status (Editorial). *ANESTHESIOLOGY* 1963; 24:111
3. Owens WD: American Society of Anesthesiologists Physical Status Classification System in not a risk classification system (Letter). *ANESTHESIOLOGY* 2001; 94:378
4. Khuri SF, Daley J, Henderson W, Hur K, Gibbs JO, Barbour G, Demakis J, Irvin G 3rd, Stremple JF, Grover F, McDonald G, Passaro E Jr, Fabri PJ, Spencer J, Hammermeister K, Aust JB:

- Risk adjustment of the postoperative mortality rate for the comparative assessment of the quality of surgical care: Results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg* 1997; 185:315-27
5. Jones HJ, de Cossart L: Risk scoring in surgical patients. *Br J Surg* 1999; 86:149-57
 6. Hariharan S, Zbar A: Risk scoring in perioperative and surgical intensive care patients: A review. *Curr Surg* 2006; 63: 226-36
 7. Gawande AA, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ: An Apgar score for surgery. *J Am Coll Surg* 2007; 204: 201-8
 8. Regenbogen SE, Ehrenfeld JM, Lipsitz SR, Greenberg CC, Hutter MM, Gawande AA: Utility of the surgical apgar score: Validation in 4119 patients. *Arch Surg* 2009; 144:30-6; discussion 37
 9. Regenbogen SE, Lancaster RT, Lipsitz SR, Greenberg CC, Hutter MM, Gawande AA: Does the Surgical Apgar Score measure intraoperative performance? *Ann Surg* 2008; 248: 320-8
 10. Zigelboim I, Kizer N, Taylor NP, Case AS, Gao F, Thaker PH, Rader JS, Massad LS, Mutch DG, Powell MA: "Surgical Apgar Score" predicts postoperative complications after cytoreduction for advanced ovarian cancer. *Gynecol Oncol* 2010; 116: 370-3
 11. Prasad SM, Ferreria M, Berry AM, Lipsitz SR, Richie JP, Gawande AA, Hu JC: Surgical apgar outcome score: Perioperative risk assessment for radical cystectomy. *J Urol* 2009; 181:1046-53
 12. Delilkan AE: Comparison of subjective estimates by surgeons and anaesthetists of operative blood loss. *BMJ* 1972; 2:619-21
 13. Gardiner AJ, Dudley HA: The measurement of blood loss at operation. *Br J Anaesth* 1962; 34:653-6
 14. Charlson ME, MacKenzie CR, Gold JP, Ales KL, Topkins M, Shires GT: Intraoperative blood pressure. What patterns identify patients at risk for postoperative complications? *Ann Surg* 1990; 212:567-80
 15. Charlson ME, MacKenzie CR, Gold JP, Ales KL, Topkins M, Shires GT: Preoperative characteristics predicting intraoperative hypotension and hypertension among hypertensives and diabetics undergoing noncardiac surgery. *Ann Surg* 1990; 212:66-81
 16. Monk TG, Saini V, Weldon BC, Sigl JC: Anesthetic management and one-year mortality after noncardiac surgery. *Anesth Analg* 2005; 100:4-10
 17. R Development Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2010.
 18. POISE Study Group, Devereaux PJ, Yang H, Yusuf S, Guyatt G, Leslie K, Villar JC, Xavier D, Chrolavicius S, Greenspan L, Pogue J, Pais P, Liu L, Xu S, Málaga G, Avezum A, Chan M, Montori VM, Jacka M, Choi P: Effects of extended-release metoprolol succinate in patients undergoing non-cardiac surgery (POISE trial): A randomised controlled trial. *Lancet* 2008; 371:1839-47
 19. Celli BR, Rodriguez KS, Snider GL: A controlled trial of intermittent positive pressure breathing, incentive spirometry, and deep breathing exercises in preventing pulmonary complications after abdominal surgery. *Am Rev Respir Dis* 1984; 130:12-5
 20. Liu SS, Wu CL: Effect of postoperative analgesia on major postoperative complications: A systematic update of the evidence. *Anesth Analg* 2007; 104:689-702
 21. Cannon CP: Clinical perspectives on the use of composite endpoints. *Control Clin Trials* 1997; 18:517-29; discussion 546-9
 22. O'Neill RT: Secondary endpoints cannot be validly analyzed if the primary endpoint does not demonstrate clear statistical significance. *Control Clin Trials* 1997; 18:550-6; discussion 561-7
 23. Nygren J, Thorell A, Ljungqvist O: Preoperative oral carbohydrate nutrition: An update. *Curr Opin Clin Nutr Metab Care* 2001; 4:255-9