

# The Surgical Apgar Score in Hip and Knee Arthroplasty

Thomas H. Wuerz MD, MSc, Scott E. Regenbogen MD, MPH,  
Jesse M. Ehrenfeld MD, Henrik Malchau MD, PhD, Harry E. Rubash MD,  
Atul A. Gawande MD, MPH, David M. Kent MD, MS

Received: 22 February 2010 / Accepted: 19 November 2010 / Published online: 4 December 2010  
© The Association of Bone and Joint Surgeons® 2010

## Abstract

**Background** A 10-point Surgical Apgar Score, based on patients' estimated blood loss, lowest heart rate, and lowest mean arterial pressure during surgery, was developed to rate patients' outcomes in general and vascular surgery but has not been tested for patients having orthopaedic surgery. **Questions/purposes** For patients undergoing hip and knee arthroplasties, we asked (1) whether the score provides accurate risk stratification for major postoperative complications, and (2) whether it captures intraoperative

variables contributing to postoperative risk based on the three parameters independent of preoperative risk.

**Patients and Methods** We retrospectively reviewed the electronic records for all 3511 patients who underwent a hip or knee arthroplasty from March 2003 to August 2006 and extracted data to calculate a Surgical Apgar Score. We evaluated the relationship between scores and likelihood of major postoperative in-hospital complications and assessed its discrimination and calibration.

**Results** Complication rates increased monotonically as the score decreased. Even after controlling for preoperative risk, each 1-point decrease in the score was associated with a 34.0% increase (95% confidence interval, 0.66–0.84) in the odds of a complication. The overall discriminatory performance of the score was a c-statistic of 0.61. Seventy-six percent of all major complications occurred in patients classified as low risk with scores of 7 or greater.

**Conclusions** For patients undergoing hip and knee arthroplasties, the score captures important intraoperative

---

One of the authors (THW) has received funding from the National Institutes of Health under Ruth L. Kirschstein National Research Service Award (Agency for Healthcare Research and Quality Grant Number T32 HS00060).

Each author certifies that his or her institution approved the human protocol for this investigation and that all investigations were conducted in conformity with ethical principles of research. This work was performed at Tufts Medical Center and the Massachusetts General Hospital.

---

T. H. Wuerz (✉), D. M. Kent  
Center for Predictive Medicine Research, Institute for Clinical  
Research and Health Policy Studies, Tufts Medical Center,  
35 Kneeland Street, Boston, MA 02111, USA  
e-mail: twuerz@partners.org

T. H. Wuerz  
Clinical Research Program, Sackler School of Graduate  
Biomedical Sciences, Tufts University, Boston, MA, USA

T. H. Wuerz, H. Malchau, H. E. Rubash  
Department of Orthopaedic Surgery, Massachusetts  
General Hospital, Boston, MA, USA

S. E. Regenbogen  
Department of Surgery, Massachusetts General Hospital,  
Boston, MA, USA

S. E. Regenbogen, A. A. Gawande  
Department of Health Policy and Management, Harvard  
School of Public Health, Boston, MA, USA

J. M. Ehrenfeld  
Department of Anesthesia and Critical Care, Massachusetts  
General Hospital, Boston, MA, USA

A. A. Gawande  
Center for Surgery and Public Health, Department of Surgery,  
Brigham and Women's Hospital, Boston, MA, USA

information regarding risk of complications and contributes additional information to preoperative risk, but on its own is insufficient to provide comprehensive postoperative risk stratification for arthroplasties.

*Level of Evidence* Level II, prognostic study. See Guidelines for Authors for a complete description of levels of evidence.

## Introduction

Reliable, objective, and routine assessment of a patient's condition after surgery may provide important information for improving postoperative management and support efforts to improve quality of care and patient safety. Based on data routinely collected as part of the National Surgical Quality Improvement Program (NSQIP), a simple three-item, 10-point Surgical Apgar Score was developed to predict the occurrence of major postoperative complications and 30-day mortality (Table 1) [8]. The Surgical Apgar Score is intended to supply an immediate and easily calculated objective summary evaluation of a patient's condition after surgery to identify patients at high risk for major complications and to provide an objective summary for hand-off communication between different teams. Closer monitoring of postoperative patients with low scores (eg, a score of 4 or less), may allow prevention of complications. The score was intended to be used as a modifiable target for surgical teams and researchers aiming to improve outcomes by serving as a measure for quality improvement programs. Practices were advocated to be developed to reduce the number of patients with low scores. However, it has not been tested in patients undergoing joint arthroplasty, and no similar score has been developed for such patients.

For patients undergoing hip and knee arthroplasties, we asked (1) whether the score provides accurate risk stratification for major postoperative complications, and (2) whether it captures intraoperative variables contributing to postoperative risk independent of preoperative risk.

## Patients and Methods

Data were extracted from the medical records of all 3511 patients undergoing primary revision hip and knee arthroplasties from March 1, 2003, to August 31, 2006, at Massachusetts General Hospital. Data for primary diagnosis, procedure, comorbidities, intraoperative variables, and immediate outcomes during hospitalization were collected from electronic clinical data and electronic administrative records at Massachusetts General Hospital. The study protocol, including a waiver of informed consent for individual patients, was approved by the Massachusetts General Hospital Human Research Committee.

Data for each patient were abstracted manually from discharge summaries, operative notes, and ICD-9 codes by one investigator (THW). Primary diagnoses and procedures were abstracted from the operative reports and discharge summaries. In case of conflicting information, the operative report took precedence. To assess interrater reliability, a clinical fellow (THV) was first instructed on the abstraction techniques and then abstracted data for the same variables regarding comorbidities and major complications from a random sample of 50 patients taken from the database. Abstraction of outcomes by an independent rater yielded 100% agreement for all the variables.

Primary diagnoses and indications were collapsed into the following categories: osteoarthritis (including dysplasia and slipped capital femoral epiphysis); rheumatoid arthritis (including inflammatory etiologies, such as villonodular and psoriatic and heterotopic ossification); infection-related joint arthroplasty; mechanical (including dislocation, aseptic loosening, failed allografts, pseudarthrosis, wear, and osteolysis); avascular necrosis; posttraumatic changes; and benign or malignant tumor-related joint arthroplasty.

Intraoperative parameters included estimated blood loss (EBL), lowest heart rate (HR), and lowest mean arterial pressure (MAP). Intraoperative records were stored in an electronic Anesthesia Information Management System (Saturn; Dräger Medical, Telford, PA, USA). This database is accessible via Structured Query Language (SQL). A SQL query was developed to examine the intraoperative

**Table 1.** The 10-point Surgical Apgar Score

Parameter	0 points	1 point	2 points	3 points	4 points
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 (beats per minute)	> 85*	76–85	66–75	56–65	≤ 55*

\* Occurrence of pathologic bradyarrhythmia, including sinus arrest, atrioventricular block or dissociation, junctional or ventricular escape rhythms, and asystole also receive 0 points for lowest heart rate. (Reprinted with permission of Elsevier from Gawande AA, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ. An Apgar score for surgery. *J Am Coll Surg*. 2007;204:201–208.)

physiologic data during the surgery. Electronic anesthesia data differ from handwritten records in multiple aspects [5, 21]. Specifically, the tendency for inclusion of some artifactual or erroneous values (for example, false pressure readings when an arterial catheter is flushed) is of concern. Therefore, we used a previously validated filtering algorithm to eliminate artifactual readings [19, 21].

The 10-point Surgical Apgar Score was computed for each patient based on three intraoperative parameters (EBL, HR, MAP) (Table 1). Derivation of the score was based on a logistic regression model as previously reported [8].

The primary end point was the incidence of a major postoperative complication or death during hospitalization. Major complications were identified from diagnoses in discharge summaries, operative reports, and ICD-9 codes and are based on definitions from the NSQIP [13] (Table 2).

Basic demographics and summary statistics were calculated overall and for those with and without major complications. There were no missing data for analysis of the intraoperative variables and the outcomes. For development of the preoperative risk model, only complete cases with complete data were used; as some laboratory values were missing for some patients, the overall number of patients for analysis decreased from 3511 to 3236 for this model. For all variables, including the three intraoperative variables of interest, differences between patients with and

without complications were compared using the two-sided t test or Wilcoxon rank-sum test for continuous variables and chi square test for categorical variables.

To assess predictive performance, we used univariable logistic regression, with the 10-point score as a categorical predictor, to evaluate the calibration and discrimination of the score as a comprehensive predictive instrument for major complications or death in the arthroplasty cohort. We evaluated calibration with calibration graphs and the Hosmer-Lemeshow goodness-of-fit test [12]. This test is commonly used to assess for goodness of fit for logistic regression models. It tests whether the observed event rates match expected event rates. If observed and expected event rates are similar, the model is called well calibrated. If the p value of this test is more than 0.05, there is no evidence of lack of fit. Discrimination, as a measure of how well the score can differentiate patients with and without complications, was assessed by the c-statistic. Closely related to sensitivity and specificity, the c-statistic represents the percentage of all possible discordant pairs of cases in which the model correctly assigns a higher probability of having a major complication to the patient with the complication rather than to the patient without the complication.

To investigate the stability of the score in relation to mortality status, sensitivity analysis was performed by subgroup analysis, with patients who had died being excluded.

To analyze whether the score contributes incremental information regarding patient risk independent of baseline clinical variables, we derived a preoperative risk model based on clinical variables, easily obtainable before surgery, using stepwise multivariable logistic regression to predict major complications. Variables tested for inclusion in the model included age, gender, weight, height, American Society of Anesthesiologists (ASA) class, creatinine, estimated glomerular filtration rate, blood urea nitrogen (BUN), and diagnosis of femoral fracture. Preexisting conditions including diabetes mellitus and pulmonary and cardiovascular comorbidities also were evaluated as risk factors (Table 2). 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 (TIA), or stroke, with or without residual neurologic deficit were pooled in one group, called “history of stroke/TIA” and included in cardiovascular comorbidity. Nonphysiologic outliers and influential points were checked individually for the logistic regression models, and variables were truncated accordingly when necessary. Model diagnostics were performed with residual plots. The c-statistic for this

**Table 2.** Individual major complications

Diagnosis	Major complications (number of patients)
Deep venous thrombosis	39 (1.11%)
Bleeding requiring equal or more than 4 units of PRBC	38 (1.08%)
Pneumonia	36 (1.03%)
Acute renal failure	34 (0.97%)
Pulmonary embolism	26 (0.74%)
Myocardial infarction	17 (0.48%)
Unplanned intubation	11 (0.31%)
Cardiac arrest	7 (0.20%)
Systemic inflammatory response syndrome	7 (0.20%)
Surgical site infection	4 (0.11%)
Ventilator dependence	3 (0.09%)
Stroke	3 (0.09%)
Septic shock	3 (0.09%)
Sepsis	3 (0.09%)
Coma	1 (0.03%)
Wound disruption	0

PRBC = packed red blood cells.

model was calculated to summarize discrimination and the Hosmer-Lemeshow goodness-of-fit test was used to assess calibration.

Patients were stratified into equal-sized quintiles based on their preoperative risk as determined by the derived model. A Spearman correlation was used to compare preoperative risk stratification and the score. Outcome rates for increasing levels of the Surgical Apgar Score (categorized as 0–4, 5–6, 7–8, and 9–10) were calculated for each quintile of preoperative risk. A Cochran-Armitage chi square test of trend was performed separately for each preoperative risk quintile to see whether, in each level of preoperative risk, outcome rate was related to the intraoperative risk as measured by the score category. We also used logistic regression to assess whether the score added incremental prognostic information independent of preoperative risk, again using the incidence of major complications as outcome and controlling for preoperative risk predictions. A likelihood ratio test was computed to compare the model with preoperative risk predictions as the only independent variables with a model that included the preoperative risk predictions and the score. The results are presented as the odds ratio, 95% confidence interval, and corresponding *p* value for the score, adjusted for preoperative risk. All analyses were performed using the SAS® 9.1 statistical software package (SAS Institute Inc, Cary, NC, USA).

## Results

Patients who experienced major complications were older and more likely to be male, have cardiovascular disease, pulmonary disease, or diabetes, and an ASA Class III or greater. The likelihood of complications varied: among procedures, the highest rates occurred for hemiarthroplasty and revision THA; fractures and infections were the highest-risk diagnoses (Table 3).

The score did not provide comprehensive risk stratification in hip and knee arthroplasties. The score achieved a *c*-statistic of 0.61. There were no discrepancies between observed and predicted outcome rates (Hosmer-Lemeshow chi square statistic *p* = 0.20); however, calibration curves (Fig. 1) showed the observed outcome rate in this orthopaedic cohort was consistently lower than that predicted by the score for general and vascular surgery. Excluding mortality from the analysis did not appreciably affect model performance (*c*-statistic = 0.60).

The score captured intraoperative contribution to postoperative risk independent of preoperative risk. Among intraoperative measures, the mean lowest MAP was lower and the mean lowest HR and EBL were higher among patients with major complications compared with those

without complications (Table 3). The majority of patients in our cohort had a score ranging from 7 to 10 and few patients had scores between 0 and 4 (Fig. 2). The proportion of major complications or death decreased (*p* < 0.001) with increasing levels of the score (Fig. 2). Stratification by increasing score levels also shows a decreasing trend of likelihood ratios for major complications in arthroplasty. This is consistent with a similar observed trend in general/vascular surgery (Table 4) [18].

We identified four independent preoperative predictors (ASA class, BUN, cardiovascular comorbidity, diagnosis of fracture), which together provided important prognostic information regarding postoperative risk (*c*-statistic = 0.73). The score correlated with preoperative risk predictions (*r* = − 0.074; *p* < 0.001), indicating patients with a higher burden of preoperative risk factors tended to have slightly higher intraoperative risk as well (represented by lower scores), but most of the variation in the scores (almost 99%) was not explained by preoperative risk. In four of the five quintiles of preoperative risk, there was a trend of increasing major complication rates with decreasing score values (Table 5). After controlling for preoperative factors independently associated with increased risk of complications (ASA class, BUN, cardiovascular comorbidity, diagnosis of fracture), the odds ratio for the score was 0.746 (95% confidence interval, 0.66–0.84; *p* < 0.0001). This suggests independent intraoperative contribution to postoperative risk because, for each 1-point decrease in the score, the odds of a major complication increased by 34.0%. Adding the 10-point score as an independent predictor to the preoperative risk model improved model discrimination (*c*-statistic increased from 0.73 to 0.75).

## Discussion

The 10-point Surgical Apgar Score has been developed for providing an estimate of patients' risk for major complications and death immediately after general and vascular surgery. We assessed whether the Surgical Apgar Score provides accurate risk stratification for hip and knee arthroplasties and whether the score captures the intraoperative variables contributing to postoperative risk for major complications independent of preoperative risk.

There are several limitations to our study. First, unlike the NSQIP in general and vascular surgery, no national, comprehensive, validated outcomes data collection and assessment system exists for hip and knee arthroplasties. We could evaluate only the in-hospital complications reliably available in our hospital's electronic medical record. However, in-hospital complications reportedly account for greater than 90% of 30-day complications for

**Table 3.** Preoperative risk factors and in-hospital outcomes.

Patient characteristics	Arthroplasty patients (n = 3511)	No major complications (n = 3313)	Major complications (n = 198)	p Value*
Age (years) <sup>†</sup>	66.2 (12.9)	65.8 (12.8)	71.5 (13.9)	< 0.001
Female gender (number of patients)	1940 (55.3%)	1847 (55.8%)	93 (46.7%)	0.02
Cardiovascular comorbidity (MI, CHF, PVD, stroke, prior revascularization) (number of patients)	434 (12.4%)	363 (11.0%)	71 (35.9%)	< 0.001
Pulmonary comorbidity (pneumonia, COPD, ventilator dependent) (number of patients)	142 (4.0%)	117 (3.5%)	25 (12.6%)	< 0.001
Diabetes mellitus (number of patients)	322 (9.2%)	288 (8.7%)	34 (17.2%)	< 0.001
ASA Class III, IV, or V (number of patients)	826 (23.5%)	719 (21.7%)	107 (54.0%)	< 0.001
Outcomes (number of patients)				
Major complications/death	198 (5.6%)			
Deceased	30 (0.9%)	0 (0%)	30 (15.2%)	
Procedures (number of patients)				< 0.0001
Primary TKA	1318 (37.5%)	1269 (38.3%)	49 (24.8%)	
Revision TKA	113 (3.2%)	106 (3.2%)	7 (3.5%)	
Primary unicompartmental knee arthroplasty	63 (1.8%)	63 (1.9%)	0 (0%)	
Primary THA	1481 (42.2%)	1421 (42.9%)	60 (30.3%)	
Revision THA	328 (9.3%)	296 (8.9%)	32 (16.2%)	
Hemiarthroplasty	207 (5.9%)	157 (4.7%)	50 (25.3%)	
Primary diagnosis (number of patients)				< 0.0001
Osteoarthritis	2580 (73.5%)	2498 (75.4%)	82 (41.4%)	
Rheumatoid arthritis	37 (1.0%)	37 (1.1%)	0 (0%)	
Fractures	213 (6.1%)	154 (4.7%)	59 (29.8%)	
Infection	74 (2.1%)	65 (1.9%)	9 (4.6%)	
Mechanical	385 (11.0%)	352 (10.6%)	33 (16.7%)	
Posttraumatic	40 (1.1%)	38 (1.2%)	2 (1.0%)	
Tumor	72 (2.0%)	68 (2.0%)	4 (2.0%)	
Nonunion	14 (0.4%)	11 (0.3%)	3 (1.5%)	
Avascular necrosis	95 (2.7%)	89 (2.7%)	6 (3.0%)	
Intraoperative characteristics				
Estimated blood loss (mL) <sup>‡</sup>	225 (100–400)	200 (100–400)	300 (100–600)	0.002
Lowest heart rate (beats per minute) <sup>†</sup>	55.7 (9.3)	55.6 (9.2)	58.6 (10.9)	< 0.001
Lowest arterial pressure (mm Hg) <sup>†</sup>	67.6 (10.5)	67.7 (10.5)	65.9 (10.65)	0.02

\* The p values were obtained from chi square tests, t tests, or Wilcoxon rank-sum tests; <sup>†</sup>values are expressed as mean, with SD in parentheses; <sup>‡</sup>values are expressed as median, with interquartile range in parentheses; MI = myocardial infarction; CHF = congestive heart failure; PVD = peripheral vascular disease; COPD = chronic obstructive pulmonary disease; ASA = American Society of Anesthesiologists.

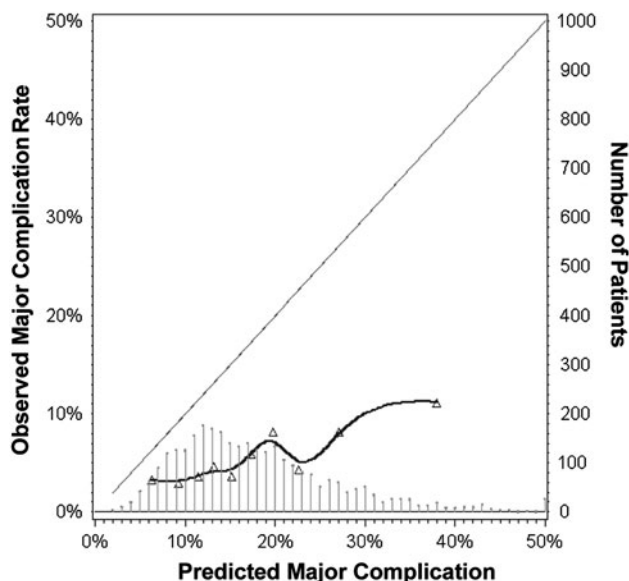
arthroplasties [14], so this difference alone is unlikely to account for the large discrepancy between predicted and observed complication rates when comparing the score's performance in our cohort with the original general and vascular surgery cohort. Second, the lack of full-time research nurses abstracting outcome data from full chart reviews of individual patients (as is performed for the NSQIP) precluded our obtaining the same levels of comprehensive abstraction of outcomes. However, unlike minor complications (eg, urinary tract infections), major complications and death can be expected to be documented

in discharge summaries, operative notes, and ICD coding [3, 21]. Third, this is a retrospective and not a prospective data collection. For assessing the utility of the score, a prospective collection would be required with followup regarding the potential impact of the score on outcomes.

The score on its own does not provide sufficient discrimination between patients who will or will not have complications to be used as a comprehensive risk stratification tool for clinical triage. Only 6.1% of patients with major complications fall into the high-risk category with a score of 4 or less. Most patients with complications

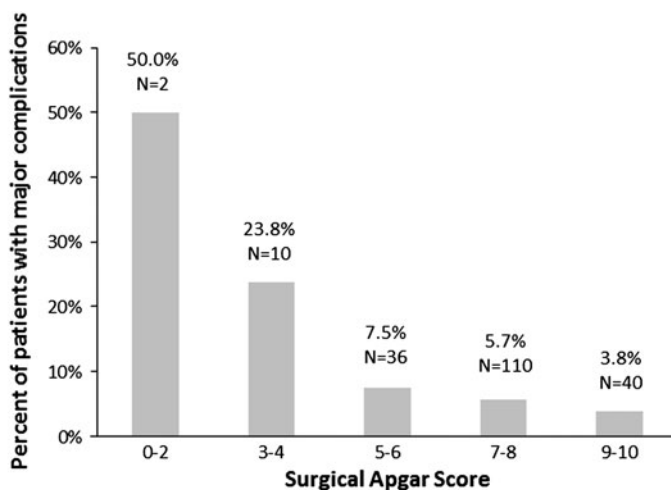


(75.8%) fall into the low-risk group with a score of 7 or more. Thus, for patients undergoing joint arthroplasty, the score cannot supplant physicians' judgment in determining



**Fig. 1** A calibration plot compares the model's predicted probabilities based on the original general and vascular surgery cohort and observed proportions in our arthroplasty cohort. Triangles display deciles of the predicted probabilities in the arthroplasty cohort. The diagonal line reflects the ideal situation (predicted probability = observed proportion). The curve represents the relation nonparametrically. The histogram in the lower part of the figure shows the distribution of predicted probabilities in this sample where the height of each line represents the number of patients shown on the y-axis on the right side of the figure. Probabilities greater than 50% were truncated. This plot shows the score constantly underpredicted major complications and as such is insufficiently accurate to serve as a comprehensive risk stratification tool.

**Fig. 2** Major complication and death rates are shown according to the 10-point Surgical Apgar Score from the operation. The graph shows a monotonic increase in risk across decreasing values of score or categories, indicating the score provides prognostically meaningful information regarding the risk of complications. Below the graph, the absolute number of patients distributed across the score categories is shown. The majority of patients have high score values and only very few have low score values.



Total number of patients (N=3511)	4	42	478	1919	1068
Patients with major complications (N=198)	2	10	36	110	40

whether a given patient merits added monitoring or other measures taken when complications are a concern. It achieved a c-statistic of only 0.61, a modest discrimination. This is substantially less compared with the general/vascular surgery validation cohort (c = 0.73) [17]. It is not surprising discrimination was poorer in this population than in the previous validation cohorts, given the important differences in the patients and operations included in the samples. The performance of prognostic models typically degrades somewhat in new populations, particularly populations that differ considerably from the derivation cohort [11].

However, even though the score on its own is not sufficient to be used as a comprehensive risk stratification tool, we found the score captures meaningful prognostic information based on intraoperative variables in patients undergoing an arthroplasty. In combination with other potential preoperative or postoperative risk factors, some of the score's variables may serve as a foundation for a comprehensive score. Although this arthroplasty cohort was substantially different in many important dimensions from the derivation cohort of general and vascular surgery patients in the NSQIP and had an overall complication rate for arthroplasty less than 1/3 that of the NSQIP cohorts (5.6% versus 22%), similarities in trends of the likelihood ratios support the notion that the relationship between the score and overall relative risk for major complications is similar for arthroplasty compared with general/vascular surgery. Additionally, the score appears to describe risk information of patients based on intraoperative variables that is at least partly independent from preoperative risk factors, indicating the score may capture aspects of the relative performance in the operating room,

**Table 4.** Likelihood ratios in relation to Surgical Apgar Scores

Likelihood ratio	Surgical Apgar Score				
	0–2	3–4	5–6	7–8	9–10
Arthroplasty likelihood ratio (with confidence intervals)	16.86 (2.3–117.6)	5.27 (2.5–10.7)	1.37 (0.9–2.0)	1.02	0.67 (0.4–0.9)
General/vascular surgery likelihood ratio	18.27	7.03	2.36	0.9	0.32

**Table 5.** Postoperative outcomes by preoperative risk stratum and Surgical Apgar Score

Preoperative risk predictors			Surgical Apgar Score								p Value*
			0–4		5–6		7–8		9–10		
Quintile	Mean risk (%)	Range (%)	Number of patients	% major complication	Number of patients	% major complication	Number of patients	% major complication	Number of patients	% major complication	
1	3.2	2.1–3.6	14	7.1	115	11.3	367	3.3	185	1.6	0.0001
2	3.8	3.6–3.9	6	16.7	84	3.6	345	2.3	188	5.3	0.29
3	4.1	3.9–4.3	2	50	79	2.5	353	3.4	176	1.1	0.039
4	4.6	4.4–5.2	4	25	74	5.4	361	3.9	227	1.3	0.005
5	11.0	5.3–59.8	13	38.5	83	15.7	358	15.1	202	8.9	0.002

\* Cochran-Armitage chi square trend test.

where more than  $\frac{2}{3}$  of surgical adverse events occur [9, 10, 20]. Patients with low scores should be recognized to be at an elevated risk for major complications, even patients whose baseline characteristics would suggest they otherwise are at low risk. Characteristics of patients with and without major complications in our cohort are comparable to those reported previously [14–16], confirming our cohort is representative. Although not providing comprehensive risk stratification, there are several potential applications for the score in patients undergoing hip and knee arthroplasties, which should be evaluated in additional studies. The score still may provide a simple summary of how an operation went, thereby supporting postoperative communication. So far, there has been no consistent way to communicate with the team managing the patient postoperatively regarding how a patient did during surgery. The data suggest the score does provide a simple, objective, easily communicated indicator of the magnitude of the procedure. The clear association of higher scores with lower complication rates suggests future studies should evaluate whether the score provides a useful target end point for safety initiatives to improve operative management in arthroplasty. As with the Apgar score in obstetrics [1, 4, 6, 7], another setting with low risk of complications and a score that comprises at best moderate discriminatory ability but nonetheless a strong indication of relative risk of early complications, the Surgical Apgar Score could provide practical information. If a reduction in the percentage of patients with poor

scores could produce a reduction in the number of complications in a given setting, high score levels may be a useful modifiable target for quality improvement programs. The score could be an important internal benchmark for departments to assess relative operative performance with time.

This simple and pragmatic surgical outcome score captured relevant intraoperative prognostic information, integrating components of patient susceptibility and operative performance distinct from standard risk adjustment, but overall is only modestly predictive of patient-specific complication risk in this relatively low-risk cohort. Future studies should focus on modifications to the score that would improve its predictive performance in this low-risk population. Based on our findings, postoperative risk prediction for hip and knee arthroplasties may require different outcome measures compared with general and vascular surgery. There has been an increased effort in building local or national total joint registries for improving quality and safety. One of the many challenges has been to find the most relevant outcome measures to be included. NSQIP has been proposed as a model for orthopaedics [2]. Our results indicate other outcome measures in addition to or in substitution of the ones collected for the NSQIP may be of importance in arthroplasty.

**Acknowledgments** We thank Robin Ruthazer for statistical support and advice. We also thank Thomas H. Vikoren for performing the abstraction to rate interobserver reliability. This study was published previously in similar form [22].

## References

1. Apgar V, Holaday DA, James LS, Weisbrot IM, Berrien C. Evaluation of the newborn infant: second report. *J Am Med Assoc.* 1958;168:1985–1988.
2. Auerbach A. Healthcare quality measurement in orthopaedic surgery: current state of the art. *Clin Orthop Relat Res.* 2009;467:2542–2547.
3. Best WR, Khuri SF, Phelan M, Hur K, Henderson WG, Demakis JG, Daley J. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the Department of Veterans Affairs National Surgical Quality Improvement Program. *J Am Coll Surg.* 2002;194:257–266.
4. 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.
5. Cook RI, McDonald JS, Nunziata E. Differences between handwritten and automatic blood pressure records. *Anesthesiology.* 1989;71:385–390.
6. Feinstein AR. Multi-item “instruments” vs Virginia Apgar’s principles of clinimetrics. *Arch Intern Med.* 1999;159:125–128.
7. Finster M, Wood M. The Apgar score has survived the test of time. *Anesthesiology.* 2005;102:855–857.
8. Gawande AA, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ. An Apgar score for surgery. *J Am Coll Surg.* 2007;204:201–208.
9. Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery.* 1999;126:66–75.
10. Gawande AA, Zinner MJ, Studdert DM, Brennan TA. Analysis of errors reported by surgeons at three teaching hospitals. *Surgery.* 2003;133:614–621.
11. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* 1996;15:361–387.
12. Hosmer DW, Lemeshow S. *Applied Logistic Regression.* 2nd ed. New York, NY: John Wiley and Sons; 2000.
13. Khuri SF, Daley J, Henderson W, Barbour G, Lowry P, Irvin G, Gibbs J, Grover F, Hammermeister K, Stremple JF, 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;180:519–531.
14. Mantilla CB, Horlocker TT, Schroeder DR, Berry DJ, Brown DL. Frequency of myocardial infarction, pulmonary embolism, deep venous thrombosis, and death following primary hip or knee arthroplasty. *Anesthesiology.* 2002;96:1140–1146.
15. Parvizi J, Johnson BG, Rowland C, Ereth MH, Lewallen DG. Thirty-day mortality after elective total hip arthroplasty. *J Bone Joint Surg Am.* 2001;83:1524–1528.
16. Parvizi J, Mui A, Purtill JJ, Sharkey PF, Hozack WJ, Rothman RH. Total joint arthroplasty: when do fatal or near-fatal complications occur? *J Bone Joint Surg Am.* 2007;89:27–32.
17. 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;144:30–36; discussion 37.
18. Regenbogen SE, Lancaster RT, Lipsitz SR, Greenberg CC, Hutter MM, Gawande AA. Does the Surgical Apgar Score measure intraoperative performance? *Ann Surg.* 2008;248:320–328.
19. Reich DL, Wood RK Jr, Mattar R, Krol M, Adams DC, Hossain S, Bodian CA. Arterial blood pressure and heart rate discrepancies between handwritten and computerized anesthesia records. *Anesth Analg.* 2000;91:612–616.
20. Rogers SO Jr, Gawande AA, Kwaan M, Puopolo AL, Yoon C, Brennan TA, Studdert DM. Analysis of surgical errors in closed malpractice claims at 4 liability insurers. *Surgery.* 2006;140:25–33.
21. Romano PS, Chan BK, Schembri ME, Rainwater JA. Can administrative data be used to compare postoperative complication rates across hospitals? *Med Care.* 2002;40:856–867.
22. Wuerz TH, Regenbogen SE, Jesse M, Ehrenfeld JM, Herndon JH, Malchau H, Rubash HE, Ruthazer R, Gawande AA, Kent DM. The Surgical Apgar Score in hip and knee arthroplasty. Available at: <http://proquest.umi.com/pqdlink?did=1692491171&Fmt=14&VType=PQD&VInst=PROD&RQT=309&VName=PQD&TS=1288277781&clientId=79356>. Accessed October 28, 2010.