The Journal of Arthroplasty 29 (2014) 1457–1462

Contents lists available at ScienceDirect

The Journal of Arthroplasty

journal homepage: www.arthroplastyjournal.org

A Nomogram to Predict Major Complications After Hip and Knee Arthroplasty

Thomas H. Wuerz, MD, MSc [a,b,c,d], David M. Kent, MD, MS [a], Henrik Malchau, MD, PhD [c], Harry E. Rubash, MD [c]

[a] Center for Predictive Medicine Research, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, Massachusetts
[b] Clinical Research Program, Sackler School of Graduate Biomedical Sciences, Tufts University, Boston, Massachusetts
[c] Department of Orthopaedic Surgery, Massachusetts General Hospital, Boston, Massachusetts
[d] Department of Orthopaedic Surgery, Cleveland Clinic Foundation, Cleveland, Ohio

ARTICLE INFO

Article history:
Received 23 April 2013
Accepted 9 September 2013

Keywords:
postoperative complications
primary and revision hip and knee arthroplasty
outcomes
predictive model

ABSTRACT

We aimed to develop a nomogram for risk stratification of major postoperative complications in hip and knee arthroplasty based on preoperative and intraoperative variables, and assessed whether this tool would have better predictive performance compared to the Surgical Apgar Score (SAS). Logistic regression analysis was performed to develop a nomogram. Discrimination and calibration were assessed. Net reclassification improvement (NRI) was used to compare to the SAS. All variables were found to be statistically significant predictors of post-operative complications except race and lowest heart rate. The concordance index was 0.76 with good calibration. Compared to the SAS, the NRI was 71.5% overall. We developed a clinical prediction tool, the Morbidity and Mortality Acute Predictor for arthroplasty (arthro-MAP) that might be useful for postoperative risk stratification.

Quality and safety improvement has become a major focus in healthcare overall and orthopaedic surgery specifically. Reliable, objective and routine assessment of patient condition after surgery may provide important information for optimizing postoperative management and could support efforts to improve quality of care and patient safety by minimizing postoperative complications.

It is believed that intraoperative management contributes importantly to overall outcomes. So far, there are no quantitative metrics of operative care [37]. Intraoperative factors, such as hypotension [12,18,23,32], hypertension [4,5,24,29,30], hypothermia [7,10,28], bradycardia [27,30], tachycardia [9,15,27,29,30], and blood loss [20,39] have been independently associated with adverse outcomes. Although some risk prediction models have integrated intraoperative variables [1,34], there is no agreement on direct evaluation of performance and safety in the operating room [2,13,14].

Efforts in quality improvement and patient safety can be supported by predictive assessment of patient status. Based on data routinely collected as part of the National Surgical Quality Improvement Program (NSQIP), a 10-point Surgical Apgar Score (SAS) was recently developed and validated in general and vascular surgery [11,19,25,26]. This score provides an instant prediction of major complications and death based on intraoperative parameters, including estimated blood loss (EBL), lowest heart rate (HR), and lowest mean arterial pressure (MAP).

In a recent study [40], we tested the SAS in a large database of hip and knee arthroplasty patients. We found that intraoperative performance as measured by the three parameters of the SAS contributes substantially to postoperative risk for major complications independent of preoperative risk. However, the score's predictive performance, as measured by discrimination, did not support its use as a reliable tool to predict major postoperative complications during hospitalization after hip and knee arthroplasty. Discrimination is a function of sensitivity and specificity. It measures how well the score can differentiate patients with and without complications.

For patients undergoing hip and knee arthroplasties, (1) whether we could identify risk factors for major postoperative complications, (2) whether we could develop a nomogram for accurate risk stratification of major postoperative complications based on preoperative and intraoperative variables, and (3) whether this tool would have better predictive performance in comparison to the SAS.

Materials and Methods

Data were collected from all patients undergoing primary and revision hip and knee arthroplasty from March 1, 2003 to August 31, 2006 at a tertiary care center. This represents a unique database developed by manually abstracting outcomes data from electronic medical records. The original SAS study used the NSQIP database from a nationally run and sponsored research consortium, to our

The Conflict of Interest statement associated with this article can be found at http://dx.doi.org/10.1016/j.arth.2013.09.007.

Reprint requests: Thomas H. Wuerz MD, MSc, Massachusetts General Hospital, Orthopaedic Surgery, Suite 3700, 55 Fruit Street, Boston, MA 02114.

http://dx.doi.org/10.1016/j.arth.2013.09.007
0883-5403/

knowledge there is no other database or resources available covering the variables we assessed in orthopaedic surgery. Data on primary diagnosis, procedure, comorbidities, intraoperative variables, and immediate outcomes during hospitalization were collected from electronic clinical data and electronic administrative records. Data on each patient were manually abstracted from discharge summaries, operative notes, and ICD-9 codes by a single investigator (THW). To assess interrater reliability, a clinical fellow independently abstracted data on the same variables from a random subsample of 50 patients.

Primary diagnosis and procedures were abstracted from the operative reports and discharge summaries. Where these sources conflicted, the operative report took precedence.

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); 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), as described in the original SAS [11]. Blood loss estimation can be imprecise, but the broad categories used to calculate the score (< 100 mL, 100 to 600 mL, 601 to 1000 mL, >1000 mL) are well within observers' range of precision in careful volumetric studies [8]. Independent estimation by anesthesiologists improves the reliability and insulates against surgeon bias. However, some imprecision in estimating blood loss can remain. Additional variables, for example volume of IV fluids administered or surgical length were assessed in our initial study [40], but were not found to be significant or not consistently recorded. Intraoperative records on each patient were available since they were routinely stored in an electronic Anesthesia Information Management System (Saturn, Dräger Medical, Telford, PA). This database is accessible via Structured Query Language (SQL). An SQL query was developed to examine the intraoperative physiologic data during the surgery. Electronic anesthesia data differ from handwritten records in multiple aspects [7,28]. 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 [25,28]. The data extraction algorithm excluded extraphysiologic values for HR (data points <20/min or >200/min) and MAP (data points <25 mm Hg or >180 mm Hg) and then selected the median of remaining values in every 5-min period for analysis. The lowest of these medians for each variable, along with the recorded EBL, was used to calculate the score. The type of anesthesia was not recorded in the database.

The primary endpoint 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 included: acute renal failure, bleeding requiring ≥4 U of red cell transfusion within 72 h after surgery, cardiac arrest requiring CPR, coma for ≥24 h, deep venous thrombosis, myocardial infarction, unplanned intubation, ventilator use for ≥48 h, pneumonia, pulmonary embolism, stroke, wound disruption, deep or organ-space surgical site infection, sepsis, septic shock, systemic inflammatory response syndrome, according to definitions from the NSQIP [17].

Basic demographics and summary statistics were calculated overall, and for those with and without major complications. For all variables, including the three intra-operative variables of interest, differences between patients with and without complications were compared using two-sided t-test or Wilcoxon rank-sum test for continuous variables and chi-squared test for categorical variables.

A logistic multivariable regression model was created including all clinically relevant predictor variables in order to build a prognostic model for risk stratification. Confounding is a major concern for studies whose intent is to support causal inference. Risk models can still perform well and reliably even when the variables are non-causally related (e.g. via confounders) to the outcome. Additionally, unlike linear regression, logistic regression is less restrictive than ordinary sum of squares regression. It does not require normally distributed dependent data or homogeneity of variance.

Selection of variables in the final model was driven by purely clinical considerations, not by statistical selection (i.e. all clinically relevant variables were forced into the model). Including clinically relevant, but not necessarily statistically significant variables has been established as a feasible methodology [33]. Only cases with complete data were included. Due to some missing variables, the number of cases included in this model was reduced to N = 3241. Ordinal and continuous variables were fit using restricted cubic splines to relax the linearity assumption between numeric predictors and the outcome. The resulting multivariable regression model served as basis for the nomogram.

We evaluated the nomogram by assessing discrimination and calibration. Discrimination, as a measure of how well the score can differentiate patients with and without complications was quantified with the concordance index. Closely related to sensitivity and specificity, the concordance index 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.

Calibration was visually assessed by plotting the nomogram predicted probability against the observed proportion of complications. Poor calibration is defined as dissimilarity between observed and expected event rates across quantiles of predicted probabilities. Bootstrapping was used to correct for overfitting bias for both discrimination and calibration. This is a standard practice of randomly resampling the observed dataset with replacement thereby reaching better estimates of the examined variables. Predictions derived from our nomogram were compared to those obtained using the SAS in the same patient population.

The net reclassification improvement (NRI) measures improvement in model performance based on reclassification tables constructed separately for participants with and without events, quantifying the correct movement across discrete risk strata [6,22]. NRI represents the difference in proportions moving up and down risk strata among patients with major complications versus patients without major complications. It distinguishes movements in the correct direction (up for patients with major complications and down for patients without major complications). Ideally, the predicted probabilities would move higher (up a category) for patients with major complications and lower (down a category) for patients without major complications.

All statistical analyses were performed using the SAS 9.1 statistical software package (Cary, NC, 2006) and S-plus Professional software (Statistical Sciences, Seattle, WA, 2000) with Design package added. The study protocol, including a waiver of informed consent for individual patients, was approved by the MGH Human Research Committee.

Results

We identified several risk factors for major postoperative complications. The demographics and descriptive statistics for our patient cohort are given in Table 1. All predictor variables in the multivariable logistic analysis except lowest heart rate and race were statistically significant (Table 2). Higher blood loss or blood urea nitrogen levels were associated with worse outcome. Among the different surgical procedures, revision surgery has the worst outcome. As part of the development an initial subgroup analysis was performed finding no significant difference between TKA and THA for the variables under investigation. Patients with an American