Journal of Clinical Neuroscience 43 (2017) 247–255

Contents lists available at ScienceDirect

Journal of Clinical Neuroscience

journal homepage: www.elsevier.com/locate/jocn

Tools and techniques

The Seattle spine score: Predicting 30-day complication risk in adult spinal deformity surgery

Quinlan D. Buchlak [a,*], Vijay Yanamadala [a], Jean-Christophe Leveque [a], Alicia Edwards [a], Kellen Nold [a], Rajiv Sethi [a,b]

[a] Neuroscience Institute, Virginia Mason Medical Center, Seattle, WA, USA
[b] Department of Health Services, University of Washington, Seattle, WA, USA

ARTICLE INFO

Article history:
Received 5 February 2017
Accepted 8 June 2017

Keywords:
Risk stratification
Surgical complications
Spine surgery
Predictive clinical decision support

ABSTRACT

Background: Complication rates in complex spine surgery range from 25% to 80% in published studies. Numerous studies have shown that surgeons are not able to accurately predict whether patients are likely to face post-operative complications, in part due to biases based on individual experience. The purpose of this study was to develop and evaluate a predictive risk model and decision support system that could accurately predict the likelihood of 30-day postoperative complications in complex spine surgery based on routinely measured preoperative variables.
Methods: Preoperative and postoperative data were collected for 136 patients by reviewing medical records. Logistic regression analysis (LRA) was applied to develop the predictive algorithm based on patient demographic parameters, including age, gender, and co-morbidities, including obesity, diabetes, hypertension and anemia. We additionally compared the performance of the predictive model to a spine surgeon's ability to predict patient complications using signal detection theory statistics representing sensitivity and response bias (A′ and B″ respectively). We developed a decision support system tool, based on the LRA predictive algorithm, that was able to provide a numeric probabilistic likelihood statistic representing an individual patient's risk of developing a complication within the first 30 days after surgery.
Results: The predictive model was significant ($\chi^2 = 16.242$, $p < 0.05$), showed good fit, and was calibrated by using area under the receiver operating characteristics curve analysis (AUROC = 0.712, $p < 0.01$). The model yielded a predictive accuracy of 75.0%. It was validated by splitting the data set, comparing subset models, and testing them with unknown data. Validation also involved comparing the classification of cases by experts with the classification of cases by the model. The model significantly improved the classification accuracy of physicians involved in the delivery of complex spine surgical care.
Conclusions: The application of technology and data-driven tools to advanced surgical practice has the potential to improve decision making quality, service quality and patient safety.

1. Introduction

The population of patients with spinal deformity requiring surgical treatment is growing [1,2]. With the move towards value-based care, surgical care for these patients is being rewarded for higher quality with controlled cost [2–4]. Efforts to improve the accuracy of surgical decision-making and to develop data-driven risk stratification methods are likely to improve patient safety and outcomes, and thereby increase the overall quality and value of spine surgery care.

Complex spine surgery, defined as a procedure involving six or more levels of spinal fusion, is a high-risk undertaking with high complication rates. Complication rates range from 10 percent up to 80 percent [5–12], and are often associated with increased hospital stay, cost and long-term morbidity [9,10,12]. These complications occur as a result of a complex web of social, physiological and environmental factors [13].

Preoperative assessment of complication risk in complex spine surgery is often based on broad prevalence rates and retrospective percentage statistics. The development of debiasing strategies in high-risk medical decision making has the potential to increase service quality and patient safety. Debiasing involves moving away from intuitive processing towards processing that is more analytical, evidence-based and system-supported [14]. Robust predictive

* Corresponding author at: Virginia Mason Medical Center Neuroscience Institute, 1201 Terry Ave, Seattle, WA 98101, USA.
E-mail address: quinlanbuchlak@gmail.com (Q.D. Buchlak).

http://dx.doi.org/10.1016/j.jocn.2017.06.012
0967-5868/

models are one method to improve risk assessment and achieve gains in service quality. While work on predictive modeling in spine surgery is progressing [15–20], the application of data-driven methods for accurately and reliably predicting surgical risk and patient complications in spine surgery is rare.

The purpose of this study was to generate and calibrate a statistical model to predict the risk of 30-day complications associated with complex spine surgery. The utility of the model was maximized by focusing on preoperative variables that were readily available and easily measurable. We hypothesized that a statistical model developed using preoperative patient characteristics would accurately predict the likelihood of 30-day complications. We used the predictive model to develop a decision support system (DSS) with a quantified output representing the risk of complication within 30 days of surgery. We performed an evaluation experiment to assess the utility of this statistical model-driven DSS in helping physicians involved in the delivery of complex spine surgical care to identify patients who were at higher or lower risk of postoperative complications. We hypothesized that the additional information provided by the DSS would increase the capability of physicians to accurately predict whether patients would or would not go on to experience postoperative complications.

2. Method

2.1. Predictive modeling and DSS development

2.1.1. Participants and data collection

This retrospective predictive modeling study included a total of 136 consecutive spine deformity patients. Inclusion criteria were as follows. Patients were (1) at least 18 years of age, (2) diagnosed with adult spinal deformity with a coronal lumbar or thoracic curve greater than 40 degrees and/or significant sagittal plane imbalance with SVA greater than 10 cm and LL-PI mismatch of 20 degrees or greater, and (3) treated with a spinal fusion procedure involving six or more vertebral levels. All patients underwent a complex spine procedure involving a posterior approach. A subset of cases had a secondary minimally invasive lateral approach for anterior fusion. Surgeries were performed at a single high-volume institution in the United States with a large multistate referral pattern for adult spinal deformity cases. Data was collated for all cases based on queries of the institution's data warehouse and abstraction of electronic medical records (EMR). Abstraction was conducted by two trained abstractors. A random sample of 15% of cases was selected to assess the accuracy of chart abstraction. Inter-rater data concordance was 100%.

2.1.2. Predictive model development

The primary outcome of interest was the occurrence of a complication event within 30 days of surgery. A complication event was defined as a patient experiencing one or more of the following: cardiac event including myocardial infarction, pneumothorax, pneumonia, wound infection, wound dehiscence, urinary tract infection, pulmonary embolism, thromboembolism, unplanned return to surgery and death. The presence of any complication was coded "1" and absence was coded "0". Multiple complications were not additive.

Univariate and multivariate logistic regression analyses (LRA) were conducted to predict the probability of a postoperative complication event. The odds ratio of each risk factor was indicated. A set of theoretically and clinically relevant potential predictive variables was devised based upon the expertise and recommendations of five senior surgeons. Potential predictive variables needed to be captured adequately within the EMR for inclusion in this study. Potential predictors included age, gender, BMI, a history of smoking, and a preoperative diagnosis of hypertension, anxiety, depression, diabetes, bipolar disorder, Parkinson's disease, cancer, and anemia.

Summary statistics were calculated, including frequency and percentage statistics for categorical variables and means and standard deviations for continuous variables. In assessing the magnitude of associations, we calculated odds ratios and 95% confidence intervals. For the multivariate LRA, we included variables that (1) were clinically relevant or (2) achieved a univariate significance level of 0.2 or less, in line with the methods of other predictive modeling researchers [21].

To achieve sufficient power in multivariate LRA, the model must be based on a sample size that is at least 10 times the number of predictors [22]. In this case, the sample size was sufficient to substantially exceed this minimum benchmark. The final model contained seven predictor variables (BMI, age, gender, smoking status, and a preoperative diagnosis of diabetes, hypertension, or anemia) and was based on 136 cases.

Multivariate LRA models were considered significant if they achieved a p value less than 0.05. To calibrate the models and establish an indicator of their performance, discrimination between high- and low-risk patients was assessed using the area under the receiver operating curve (AUROC). The model was developed in line with predictive model development guidelines [13,23–25]. Analyses were conducted using SPSS (SPSS, Chicago, IL).

Three multivariate models were developed and their relative quality was assessed. Model calibration measures how closely actual outcomes align with those predicted by the model. Calibration was measured using the Hosmer-Lemeshow Chi-square statistic [21,26]. Model quality was assessed by reviewing the (1) model's chi-square statistic, (2) percentage of correct predictions, and (3) Nagelkerke's pseudo-R^2. The model that demonstrated the best fit and the highest percentage of correct predictions was selected for subsequent validation and experimental evaluation.

To validate the model, we divided our dataset into five distinct sets. In line with the process articulated by Assman, Cullen & Schulte (2002) [21], combinations of four of these five sets were used for generating the model and training the algorithms. The final set was used for testing the performance of the models on unknown data. This validation process was conducted for every possible 4-part, 1-part combination [21]. This internal validation process showed that the performance of the model was robust. Results in each of the subsets did not differ substantially from the model derived from the full data set.

A predictive algorithm was developed using the beta coefficients and the constant of the model based on the full dataset.

2.1.3. Decision support system development

A DSS was developed to enable the application of the predictive algorithm created. This DSS is an interactive system that applies the exponentiated regression equation, weighting each predictive variable independently. The algorithm was mathematically converted to yield a quantified probability score [24]. The DSS allows for calculation of risk in an individual patient by inputting the value for each of the seven predictor variables. The output of the DSS is a single global percentage statistic, which suggests the likelihood of complications occurring within 30 days for each individual case. Fig. 1 shows the design of the DSS dashboard that was developed. This design adheres to DSS-development guidelines [27,28].

3. Experimental evaluation

3.1. Aims and hypotheses

An experiment was conducted to assess whether the output of the DSS improved the predictive accuracy of expert physicians