The American Journal of Surgery xxx (2018) 1–14

Contents lists available at ScienceDirect

The American Journal of Surgery

journal homepage: www.americanjournalofsurgery.com

Enhanced neonatal surgical site infection prediction model utilizing statistically and clinically significant variables in combination with a machine learning algorithm

Marisa A. Bartz-Kurycki [a], Charles Green [a], Kathryn T. Anderson [a], Adam C. Alder [b], Brian T. Bucher [c], Robert A. Cina [d], Ramin Jamshidi [e], Robert T. Russell [f], Regan F. Williams [g], KuoJen Tsao [a, *]

[a] McGovern Medical School at the University of Texas Health Science Center at Houston, 6431 Fannin St, Houston, TX, 77030, USA
[b] Children's Medical Center of Dallas, 1935 Medical District Dr, Dallas, TX, 75235, USA
[c] University of Utah School of Medicine, 30 N 1900 E, Salt Lake City, UT, 84132, USA
[d] Medical University of South Carolina, 180 Calhoun St, Charleston, SC, 29401, USA
[e] Phoenix Children's Hospital, 1919 E Thomas Rd, Phoenix, AZ, 85016, USA
[f] Children's Hospital of Alabama, University of Alabama at Birmingham, 1600 7th Ave. S., Birmingham, AL, 35233, USA
[g] University of Tennessee Health Science Center, 910 Madison Ave, Memphis, TN, 38163, USA

A R T I C L E I N F O

Article history:
Received 17 February 2018
Received in revised form
8 May 2018
Accepted 17 July 2018

Keywords:
Prediction model
Surgical wound infection
Infants
Machine learning algorithm
Risk factors

A B S T R A C T

Background: Machine-learning can elucidate complex relationships/provide insight to important variables for large datasets. This study aimed to develop an accurate model to predict neonatal surgical site infections (SSI) using different statistical methods.
Methods: The 2012–2015 National Surgical Quality Improvement Program-Pediatric for neonates was utilized for development and validations models. The primary outcome was any SSI. Models included different algorithms: full multiple logistic regression (LR), *a priori* clinical LR, random forest classification (RFC), and a hybrid model (combination of clinical knowledge and significant variables from RF) to maximize predictive power.
Results: 16,842 patients (median age 18 days, IQR 3–58) were included. 542 SSIs (4%) were identified. Agreement was observed for multiple covariates among significant variables between models. Area under the curve for each model was similar (full model 0.65, clinical model 0.67, RF 0.68, hybrid LR 0.67); however, the hybrid model utilized the fewest variables (18).
Conclusions: The hybrid model had similar predictability as other models with fewer and more clinically relevant variables. Machine-learning algorithms can identify important novel characteristics, which enhance clinical prediction models.
Summary: This study evaluated risk factors associated with neonatal surgical site infection (SSI) utilizing multiple logistic regression and a random forest classifier. Operative time, open surgical technique, and preoperative supplemental nutrition were associated with SSI. A hybrid multiple logistic regression model was developed based on the random forest and clinical knowledge, and predicted neonatal SSI as well as the other models while being more feasible.

Introduction

Surgical site infections (SSIs) are the most common complication of surgery in children. SSI results in increased morbidity and mortality, additional procedures, longer hospital stays, and increased healthcare system costs.[1] In neonates, the incidence of SSI ranges between 12 and 17% of all surgeries, and may be more frequent in the common non-elective thoracic and abdominal

* Corresponding author. Department of Pediatric Surgery, McGovern Medical School at The University of Texas Health Science Center at Houston, 6431 Fannin St., MSB 5.256, Houston, TX, 77030, USA.
E-mail address: KuoJen.Tsao@uth.tmc.edu (K. Tsao).

https://doi.org/10.1016/j.amjsurg.2018.07.041
0002-9610/

procedures.[1–4] Many factors specific to neonates, including multiple co-morbidities, type of surgery, immunologic immaturity, prematurity, chronic illness and increased length of stay in intensive care units, place this population at increased risk for SSI development.[2,5,6]

Surgical wound classification was the initial system used to stratify patient risks of SSI; however, multiple studies have demonstrated errors in estimating the risk of SSI with this method, especially in children.[7–12] Recently, different models offering improved ability to estimate the risk of post-operative complications have been developed.[2,13–15] However, none specifically evaluate SSI in a large neonatal population. The complexity of this population and presence of multiple risk factors for SSI make it especially difficult to determine which risk factors are the most important.

Statistical methods that utilize machine-learning algorithms may be superior at deciphering complex relationships, providing insight into important variables within large datasets, and improving clinical outcome prediction models compared to methods previously used.[16,17] Random forest classification (RFC) is one example of a machine-learning algorithm that employs multiple classification trees to generate a “forest” of trees with improved outcome prediction.[18] Given the multiple risk factors for SSI in the neonatal population, we hypothesized that combining a machine-learning algorithm with clinical input would result in a more accurate model to predict SSI. This study aimed to determine which patient and clinical characteristics were associated with SSI and develop the most accurate model to predict surgical site infections (SSI) in neonates using RFC and multiple logistic regression.

Material and methods

Data source and patients

The American College of Surgeons National Surgical Quality Improvement Program Pediatric (NSQIP-P) data was queried for a retrospective analysis using Participant Use Files (PUF) from 2012 to 2015. The patient population was narrowed to include all neonates as per NSQIP-P definitions (age <29 days at time of surgery for term infants and <51 weeks post-conceptual age at time of surgery for pre-term infants). All surgical procedures available in the PUF were included.

The NSQIP-P program utilizes trained surgical clinical reviewers (SCR) to abstract patient-level clinical data. Data includes demographics, comorbidities, laboratory values, case type (by surgical specialty and CPT codes) and 30-day outcomes. Stringent variable definitions are adhered to in order to optimize reliability, and random audits are performed by NSQIP-P to check for data validity and definition compliance. Cases are non-consecutive, but systematically sampled across all pediatric surgical specialties at each participating institution per a specified protocol to ensure variability and reduce bias.[2,19]

Predictor variables

The NSQIP-P provides 218 preoperative and intraoperative variables including demographic information, patient co-morbidities, clinical context and perioperative factors. Preoperative variables missing more than 20% of data were excluded from analysis. Binary or categorical covariates deemed clinically significant with less than 20% of data missing were transformed to a categorical variable with an “unknown” category. Fifty-eight variables were available for model development after exclusion based on the above criteria (Appendix, Table 1). Operative technique reports whether the surgery was laparoscopic, laparoscopic converted to open, open, or unknown. This variable was not collected for the 2012 PUF, thus all patients who underwent surgery in 2012 were designated as “unknown” for the operative technique variable. Preoperative sepsis is a categorical variable with 3 levels that is defined by NSQIP-P as: 1) SIRS – fever, leukocytosis, and elevated heart rate or respiratory rate, 2) Sepsis – SIRS with a source of infection including culture, pus or abscess, 3) Septic shock – sepsis criteria plus cardiovascular dysfunction. This variable is determined by the SCR on the data collection worksheet. Only patients with complete datasets after data cleaning as described above were included in the analysis.

Outcomes

The primary outcome was a composite binary variable of any SSI, including superficial, deep, organ space SSI, and wound dehiscence. Each variable included in the composite outcome is based on the American College of Surgeons NSQIP-P User Guide for the 2015 PUF.

Development and validation cohorts

The cohort was divided based on the outcome variable, SSI, using a random number generator to create a data set for model development or training and another for model validation. Two-thirds of the records were randomized to the model training cohort and 33% of records to the validation group. To ensure comparability, variables of interest were compared amongst the cohorts (Table 1; see Appendix Table 2 for comparison of all variables).

Statistical analysis

Model development applied different statistical algorithms to maximize predictive accuracy. One model utilized a RFC, while 3 of the models utilized multiple logistic regression. The first multiple logistic regression model (full logistic regression) evaluated all perioperative variables that were significant ($n = 42$) on univariate analysis utilizing a p-value threshold of <0.25. Next, a second model's (clinical model) variables were selected *a priori* based on previous literature and clinical judgement. The final multiple logistic regression model (hybrid model) combined clinical knowledge and the 20 most important variables from RFC to develop the model. Variables were removed from the model when deemed clinically similar to another variable or felt not to be of clinical importance. Review of the adult and pediatric surgical literature guided this process, which was conducted by one author (MBK). Additional clinical expertise was provided by the senior author (KT).

A RFC was selected for its ability to efficiently analyze large data sets with improved accuracy without limiting or excluding the variables in the data set.[16,18] Additionally, random forest models indicate which predictor variables have the greatest impact on accurate outcome prediction. This machine learning algorithm, developed by *Breiman,* generates a “forest” of many classification trees using random variable selection.[18] Each tree in the forest “votes” for the best classification for a given observation, and the class receiving the majority of votes results in the prediction for that observation. Using resampling, random forest model training selects random subsets of observations in the training set (66.6% of the training sample). At each decision point in a given tree, the RFC algorithm randomly selects a subset of variables among which it identifies the one that provides the greatest classification accuracy. Each tree is grown to its largest potential with a constant number of variables without pruning. Random resampling of observations