



Published in final edited form as:

Am J Surg. 2020 July ; 220(1): 114–119. doi:10.1016/j.amjsurg.2019.10.009.

Identification of postoperative complications using electronic health record data and machine learning

Michael Bronsert, PhD^{a,b}, Abhinav B. Singh, MD^b, William G. Henderson, PhD^{a,b,c}, Karl Hammermeister, MD^{a,b,d}, Robert A. Meguid, MD, MPH^{a,b}, Kathryn L. Colborn, PhD, MSPH^{c,*}

^aUniversity of Colorado Anschutz Medical Campus, Adult and Child Consortium for Health Outcomes Research and Delivery Science, Aurora, CO, USA

^bSurgical Outcomes and Applied Research program, Department of Surgery, University of Colorado School of Medicine, Aurora CO, USA

^cUniversity of Colorado Anschutz Medical Campus, Colorado School of Public Health, Department of Biostatistics and Informatics, Aurora, CO, USA

^dUniversity of Colorado Anschutz Medical Campus, School of Medicine, Department of Cardiology, Aurora, CO, USA.

Abstract

Background: Using the American College of Surgeons National Surgical Quality Improvement Program (NSQIP) complication status of patients who underwent an operation at the University of Colorado Hospital, we developed a machine learning algorithm for identifying patients with one or more complications using data from the electronic health record (EHR).

Methods: We used an elastic-net model to estimate regression coefficients and carry out variable selection. International classification of disease codes (ICD-9), common procedural terminology (CPT) codes, medications, and CPT-specific complication event rate were included as predictors.

Results: Of 6,840 patients, 922 (13.5%) had at least one of the 18 complications tracked by NSQIP. The model achieved 88% specificity, 83% sensitivity, 97% negative predictive value, 52% positive predictive value, and an area under the curve of 0.93.

Conclusions: Using machine learning on EHR postoperative data linked to NSQIP outcomes data, a model with 163 predictors from the EHR identified complications well at our institution.

Abstract Summary

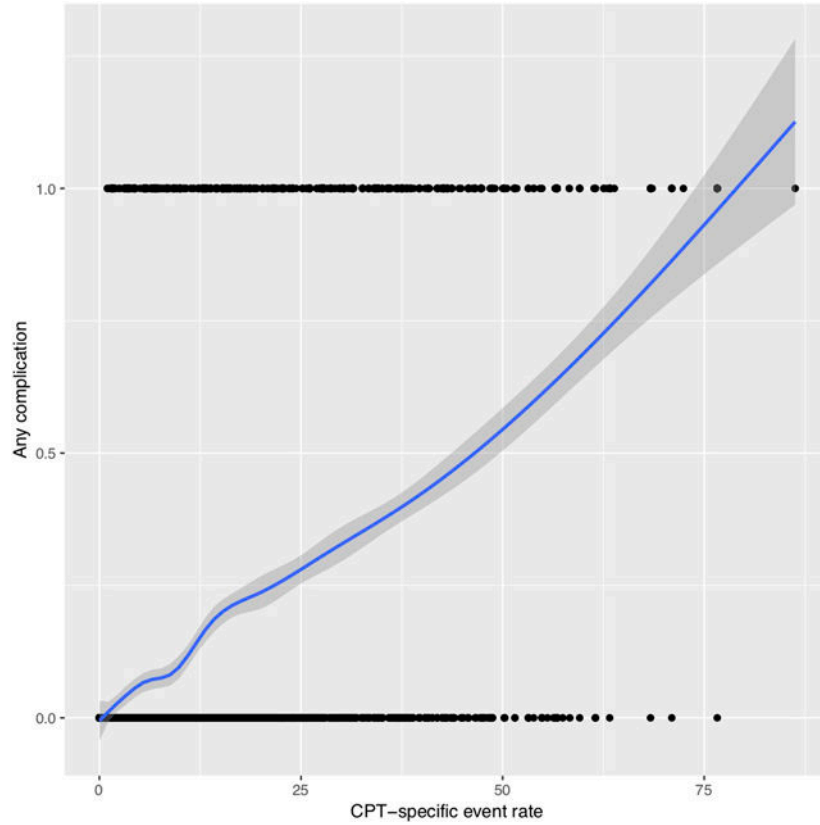
Using the American College of Surgeons National Surgical Quality Improvement Program (NSQIP) complication status of patients who underwent an operation at the University of Colorado Hospital, we developed a machine learning algorithm for identifying patients with one or more complications using data from the electronic health record (EHR). The model achieved 88% specificity, 83% sensitivity, 97% negative predictive value, 52% positive predictive value, and an

*Corresponding author: Kathryn Colborn, PhD, University of Colorado Anschutz Medical Campus, 13001 E. 17th Place, Room W4126, Aurora, Colorado 80045, Phone: 303-724-0648, kathryn.colborn@CUAnschutz.edu.

Conflicts of interest: The authors report no proprietary or commercial interest in any product mentioned or concept discussed in this article.

area under the curve of 0.93. The model developed could be used for electronic postoperative complication surveillance to supplement manual chart review.

Graphical Abstract



Keywords

NSQIP; postoperative complications; elastic-net; machine learning

Introduction

Assessment of quality of surgical care and monitoring of patient postoperative complications is an important concept in current health care delivery. Surveillance of postoperative complications has been traditionally conducted through clinical registries such as the American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP), which began in 2005 with the goal of identifying and preventing surgical complications. The currently-available NSQIP data provide high quality outcomes data on 18 different complications for more than 6.6 million patients undergoing surgery in over 720 hospitals in the United States and internationally. At participating centers, trained surgical clinical reviewers collect preoperative and operative characteristics and 30-day postoperative complications on a representative sample of patients undergoing major surgeries. Thirty-day postoperative outcomes are determined through chart reviews and by patient and family

contact after the index operation. Although the NSQIP data are considered to be of high quality, the data collection methods greatly limit the number of patients who can be assessed (~15% of those patients undergoing surgery at most large hospitals) because the process is time-consuming and costly, and participating hospitals must pay to participate.

There is a large literature regarding the use of statistical models applied to the electronic health record (EHR) to identify surgical complications: surgical site infections,¹⁻⁴ urinary tract infections,⁵⁻¹⁵ sepsis,¹⁶ bleeding,^{17, 18} and any type of complication.¹⁹⁻²¹ Most work on the identification of postoperative complications using EHR data has used structured data for the identification of specific types of complications, but because of the chosen statistical models for these analyses, the authors only explored a small number of explanatory variables.^{1, 2, 22, 23} Other work has included the addition of natural language processing (NLP) of text records in the EHR.^{5, 7, 10, 19, 21} Overall morbidity surveillance has been a major goal of the NSQIP, and it is a good measure of overall quality of care across hospitals.^{24, 25} Furthermore, because each of the 18 complications tracked by the NSQIP occur infrequently, it is difficult to build models for each complication separately, and therefore, a single overall model has potential to achieve higher positive predictive value.

In this study, we used structured data from the EHR and machine learning to identify surgical patients who experienced one or more of the 18 ACS NSQIP postoperative complications: bleeding, superficial surgical site infection (SSI), deep incisional SSI, organ space SSI, wound disruption, sepsis, septic shock, pneumonia, unplanned intubation, ventilator dependence greater than 48 hours after surgery, urinary tract infection (UTI), deep vein thrombosis (DVT)/thrombophlebitis requiring treatment, pulmonary embolism, cardiac arrest requiring cardiopulmonary resuscitation, myocardial infarction, acute renal failure, progressive renal insufficiency, and stroke. This is a novel application of high-dimensional machine learning to identify postoperative complications using EHR data. The model developed could be used for electronic postoperative complication surveillance to supplement manual chart review.

Material and Methods

Data:

In the present study, we used the 6,840 patients who underwent surgery at the University of Colorado Hospital (UCH) between July 1, 2013 and November 1, 2016, whose records were abstracted for inclusion in the ACS NSQIP. These patients' EHR data were obtained and linked by our institution's data repository team, Health Data Compass. The EHR data included demographic characteristics, International Classification of Disease versions nine and ten (ICD-9/10) codes, Current Procedural Terminology (CPT) codes, and medication codes and names. ICD-10 codes reported for patients who underwent treatment after October 1, 2016, were back-coded to ICD-9 codes. We coded ICD-9 and CPT independent variables as positive only if the codes were observed between 0 and 30 days after initial operation. We coded medications as positive only if they were observed between 3 and 30 days after the initial operation to avoid coding prophylactically-delivered medications as indicators of postoperative complications. Surgical complication status, i.e. presence or

absence of the 18 ACS NSQIP complications, came from the UCH ACS NSQIP database. Informed consent was not required for this study. This study was approved by our institutional review board.

Statistical methods:

We designed our analysis to follow the recommendations of the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement.²⁶ The dataset was temporally split into training and test datasets. Using a non-random technique (i.e., using a temporal split rather than resampling techniques, such as bootstrap or cross-validation) to divide the dataset is recommended by TRIPOD. The prevalence of complications did not change considerably over time (Table 1), therefore the temporal split allowed us to test whether the model fit to the training set was valid on a future held-out sample.

The training data consisted of operations performed from 2013 to 2015 (N=5,194; patients with no complication=4,484 [86.3%], patients with 1 complication=710 [13.7%]) and the test set consisted of operations performed in 2016 (N=1,646; patients with no complication=1,434 [87.1%], patients with 1 complication=212 [12.9%]). We formulated a comprehensive model consisting of all ICD-9 codes, CPT codes, and medications that were observed in at least five patients and had a bivariable association with one or more complications of $p < 0.1$. In addition, we included the CPT-specific overall complication rate, defined as the rate of one or more complications experienced by ACS NSQIP patients for each particular surgical procedure calculated from the national ACS NSQIP dataset of more than 6.6 million patients. A binomial generalized linear model with an elastic-net penalty was used to conduct supervised learning on the ACS NSQIP outcomes data, to estimate coefficients and to carry out variable selection. The elastic-net penalty uses a combination of the least absolute shrinkage and selection operator (lasso) and ridge penalties, therefore some coefficients were permitted to be equal to zero (as in the lasso) and correlated covariates were reweighted appropriately (as in the ridge).²⁷ Penalized regression was used primarily because the number of covariates outnumbered the number of individuals who had a complication in this dataset. Furthermore, elastic-net regularization reduces the chance of overfitting to the training data and performs variable selection in a continuous manner by allowing some coefficients to be equal to zero (as opposed to discarding variables based on p-value, like stepwise selection). Ten-fold cross-validation was performed to determine the optimal value for the lasso penalty (i.e., lambda) using the glmnet package²⁸ in R (R Foundation for Statistical Computing, Vienna, Austria). In other words, the training data were divided into ten equally sized subsets, models were fit to 9/10 of the data and tested on the 1/10 held out, and this was repeated ten times for each of the ten subsets. The lambda value that minimized the test set misclassification error from this procedure was chosen. The elastic-net tuning parameter (i.e., alpha) was set at 0.5. We estimated the predicted probabilities of one or more complications in the test set using the fitted model and performed classification using the Youden's *J* threshold²⁹ estimated in the training dataset. Models were compared with respect to sensitivity, specificity, area under the curve (AUC), accuracy, negative predictive value (NPV), positive predictive value (PPV), false negatives and false positives when classifying one or more complications in the test data. We also

evaluated the calibration and discrimination graphically for the final model. Youden's J and all other performance statistics were estimated using the pROC package³⁰ in R.

Results

Of the 6,840 patients who underwent operations at UCH between 2013 and 2016 and who were entered into the UCH ACS NSQIP database, the majority were women (56.2%), white (72.8%), and a little more than half underwent either an orthopedic surgery (26.2%) or a general surgery (24.4%) procedure (Table 1). Patients who had any complications tended to be older (58.6 years vs. 52.7 years, $p<.0001$). Patients undergoing vascular surgery (33.2%, $p<0.0001$) or neurosurgical (17.2%, $p=0.005$) operations had higher rates of complications while patients undergoing otolaryngologic (4.9%, $p<0.0001$) or orthopedic surgery (10.6, $p<0.0001$) operations had lower rates than patients undergoing other surgical specialty procedures. The overall complication rate for 2014 was slightly higher than all other years (14.7% in 2014, vs. 14.3% in 2013, 12.2% in 2015, and 12.9% in 2016, $p=0.04$). There were no significant differences in overall complication rate by race/ethnicity or gender (Table 1).

The most common complications in this patient population were bleeding (6.8% of 6,840 patients), UTI (2%), superficial SSI (1.6%), pneumonia (1.4%), sepsis (1.3%), DVT/thrombophlebitis (1.1%), and organ space SSI (1.0%; Table 2). All other complications occurred in less than 1% of patients (Table 2).

Of the 838 explanatory variables included in the comprehensive model, 163 had non-zero coefficients: 60 ICD-9 codes, 63 CPT codes, 39 medications, and the CPT-specific complication event rate. All variables were dichotomous, except for CPT-specific complication event rate, which was continuous. The relationship between CPT-specific complication event rate and the probability of any complication was approximately linear when visually inspected using a cubic smoothing spline (eFigure 1). The selected variables and their coefficient values are provided in Supplementary Table 1.

Table 3 summarizes any complication status for the binary indicators of any of the 60 ICD-9 codes, any of the 63 CPT codes, and any of the 39 medications. The presence of any of the selected ICD-9 codes (19.7% vs. 7.9%, $p<0.0001$), CPT codes (19.4% vs. 1.8%, $p<0.0001$), or medications (33.1% vs. 3.7%, $p<0.0001$) was highly associated with higher rates of complications. In addition, the median CPT-specific complication rate was much larger (22.7 vs. 5.9, $p<0.0001$) for patients having a complication compared to patients who did not have a complication.

Table 4 summarizes the classification performance of the model fit to the training data and applied to the test data set. Confidence intervals (CI) were estimated using 1,000 bootstrap samples. The AUC was 0.93, sensitivity 0.83 (95% bootstrap CI [0.79, 0.88]), specificity 0.88 (0.87, 0.90), and accuracy 0.88 (0.86, 0.89). PPV was 0.52 (0.48, 0.56) and NPV was 0.97 (0.97, 0.98). In addition, we examined models to test for temporal changes by using the first two years as a training data set and the last two years separately as test sets. The results were very similar (data not shown).

The discrimination plot of the observed and fitted values from the test set displays the predicted probabilities from the model fit to the test set by the observed values for any complication or no complication (Figure 1a). There was good discrimination, as the predicted probabilities were much higher for those with a complication than those without. The discrimination slope was 0.44 (i.e., the difference in mean predicted probabilities between the two classes). The Hosmer-Lemeshow calibration plot of the observed and fitted values from the test set displays the observed and expected rates by deciles of predicted risk of any complications (Figure 1b). The model had good calibration since the expected rates were almost identical to the observed rates, and a Pearson chi-square statistic with eight degrees of freedom was not statistically significant ($p=0.7$).

Discussion

We developed a model for the surveillance of surgical patients with one or more of the 18 ACS NSQIP postoperative complications using patients' EHR data and machine learning that correctly classified 83% of patients with a postoperative complication, 88% of those who did not have a complication, 88% of the overall outcomes, and achieved an area under the ROC curve of 0.93. This model could be used to scale-up surveillance of postoperative complications for all patients undergoing surgery at a medical center beyond the small sample of patients assessed by the ACS NSQIP protocol, without the need to hire additional dedicated staff to do time consuming chart reviews and call backs to patients and their families. Furthermore, implementation of this model at other institutions would be relatively cheap. At least at our institution, EHR data extraction is free, and the data analyst on this project spent about 10 hours cleaning and manipulating the data for analysis. Creation of the variables is also EHR-platform independent; it simply requires creation of dummy indicators for each variable from a vector of codes. For a given patient, these variables can then be multiplied by the coefficients (Supplementary Table 1) and summed, then applying the logit transformation to this sum gives the predicted probability of a complication.

At UCH, we use the ACS NSQIP data of approximately 2,500 cases annually to monitor 30-day postoperative mortality, overall morbidity, readmission, and specific types of complications (cardiac, pneumonia, unplanned intubation, ventilator dependence for greater than 48 hours, VTE, renal failure, UTL, SSI, and sepsis) across all non-cardiac and nontransplantation surgery, for individual surgical specialties and selected specific types of operations. Sample sizes are large enough to obtain reliable results for all of surgery combined and selected large surgical specialties, but not for lower volume specialties, individual providers' outcomes, or for the selected specific types of operations. An automated machine-learning surveillance system using EHR data for all surgical operations would have annual sample sizes of nearly 30,000 operations at UCH and >80,000 operations in the UCHealth system³¹, and could potentially obtain reliable results for the lower volume surgical specialties and for the specific types of operations. In addition, we could also potentially obtain reliable results to support performance measures for individual surgeons. Once machine-learning models are developed for 30-day mortality and all of the individual types of complications (some of which we have already performed^{22, 23}) we plan to use these automated EHR derived data in conjunction with our ACS NSQIP data in our quality improvement efforts.

The strength of the approach we have taken is that it utilizes available structured data from the EHR for all UCH NSQIP patients, unlike previous approaches that applied down-sampling of the majority class (those who did not have a complication).^{1, 19} Furthermore, we used a temporal split validation, which would logically apply to the implementation of this algorithm, requiring identification of future complications in patients. Limitations of our approach include: 1) The lack of claims data and the fact that some patients live out of state and therefore might have missing data; and 2) That we only used data from one hospital. In future studies, we would ideally obtain claims data for each patient, include additional hospitals, and we would attempt to extract additional data from patients' medical records using NLP, as has been done successfully by several other groups.^{5, 7, 10, 21}

This model is likely only generalizable to hospitals with similar outcome definitions (i.e., ACS NSQIP outcomes, not necessarily complications defined by the National Healthcare Safety Network, NHSN or Vizient/University HealthSystem Consortium). However, the statistical methodology and validation approach are sound and generalizable to individuals interested in developing their own models using gold standard data (e.g., ACS NSQIP, VASQIP, NHSN/Vizient) for the purpose of postoperative complication surveillance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support: This project was supported by grant number R03HS026019 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Agency for Healthcare Research and Quality. This study was also supported by a transformational research grant from the University of Colorado School of Medicine's Data-Science-to-Patient-Value initiative and the Surgical Outcomes and Applied Research (SOAR) Program funded by the University of Colorado Department of Surgery. These sources of financial support were allocated to the salaries of authors MB, KLC, KEH, and WGH. The ACS NSQIP and participating hospitals are the source of these data; the ACS has not verified and are not responsible for the statistical validity of the data analysis or the conclusions derived by the authors.

References

1. Branch-Elliman W, Strymish J, Itani KM, et al. Using clinical variables to guide surgical site infection detection: a novel surveillance strategy. *Am J Infect Control* 2014; 42: 1291–1295. 2014/12/04 DOI: 10.1016/j.ajic.2014.08.013. [PubMed: 25465259]
2. Goto M, Ohl ME, Schweizer ML, et al. Accuracy of administrative code data for the surveillance of healthcare-associated infections: a systematic review and meta-analysis. *Clin Infect Dis* 2014; 58: 688–696. 2013/11/13 DOI: 10.1093/cid/cit737. [PubMed: 24218103]
3. Hu Z, Simon GJ, Arsoniadis EG, et al. Automated Detection of Postoperative Surgical Site Infections Using Supervised Methods with Electronic Health Record Data. *Studies in health technology and informatics* 2015; 216: 706–710. 2015/08/12. [PubMed: 26262143]
4. Ju MH, Ko CY, Hall BL, et al. A comparison of 2 surgical site infection monitoring systems. *JAMA surgery* 2015; 150: 51–57. 2014/11/27 DOI: 10.1001/jamasurg.2014.2891. [PubMed: 25426765]
5. Branch-Elliman W, Strymish J, Kudesia V, et al. Natural Language Processing for Real-Time Catheter-Associated Urinary Tract Infection Surveillance: Results of a Pilot Implementation Trial. *Infection control and hospital epidemiology* 2015; 36: 1004–1010. 2015/05/30 DOI: 10.1017/ice.2015.122. [PubMed: 26022228]

6. Choudhuri JA, Pergamit RF, Chan JD, et al. An electronic catheter-associated urinary tract infection surveillance tool. *Infection control and hospital epidemiology* 2011; 32: 757–762. 2011/07/20 DOI: 10.1086/661103. [PubMed: 21768758]
7. Gundlapalli AV, Divita G, Redd A, et al. Detecting the presence of an indwelling urinary catheter and urinary symptoms in hospitalized patients using natural language processing. *Journal of biomedical informatics* 2017; 71s: S39–s45. 2016/07/13 DOI: 10.1016/j.jbi.2016.07.012. [PubMed: 27404849]
8. Hsu HE, Shenoy ES, Kelbaugh D, et al. An electronic surveillance tool for catheter-associated urinary tract infection in intensive care units. *Am J Infect Control* 2015; 43: 592–599. 2015/04/05 DOI: 10.1016/j.ajic.2015.02.019. [PubMed: 25840717]
9. Landers T, Apte M, Hyman S, et al. A comparison of methods to detect urinary tract infections using electronic data. *Joint Commission journal on quality and patient safety* 2010; 36: 411–417. 2010/09/30. [PubMed: 20873674]
10. Sanger PC, Granich M, Olsen-Scribner R, et al. Electronic Surveillance For Catheter-Associated Urinary Tract Infection Using Natural Language Processing. *AMIA Annual Symposium proceedings AMIA Symposium* 2017; 2017: 1507–1516. 2018/06/02.
11. Shepard J, Hadhazy E, Frederick J, et al. Using electronic medical records to increase the efficiency of catheter-associated urinary tract infection surveillance for National Health and Safety Network reporting. *Am J Infect Control* 2014; 42: e33–36. 2014/03/04 DOI: 10.1016/j.ajic.2013.12.005. [PubMed: 24581026]
12. Sopirala MM, Syed A, Jandarov R, et al. Impact of a change in surveillance definition on performance assessment of a catheter-associated urinary tract infection prevention program at a tertiary care medical center. *Am J Infect Control* 2018; 46: 743–746. 2018/03/20 DOI: 10.1016/j.ajic.2018.01.019. [PubMed: 29551201]
13. Tanushi H, Kvist M and Sparrelid E. Detection of healthcare-associated urinary tract infection in Swedish electronic health records. *Studies in health technology and informatics* 2014; 207: 330–339. 2014/12/10. [PubMed: 25488239]
14. Wald HL, Bandle B, Richard A, et al. Accuracy of electronic surveillance of catheter-associated urinary tract infection at an academic medical center. *Infection control and hospital epidemiology* 2014; 35: 685–691. 2014/05/07 DOI: 10.1086/676429. [PubMed: 24799645]
15. Zhan C, Elixhauser A, Richards CL Jr., et al. Identification of hospital-acquired catheter-associated urinary tract infections from Medicare claims: sensitivity and positive predictive value. *Medical care* 2009; 47: 364–369. 2009/02/06 DOI: 10.1097/MLR.0b013e31818af83d. [PubMed: 19194330]
16. Thottakkara P, Ozrazgat-Baslanti T, Hupf BB, et al. Application of Machine Learning Techniques to High-Dimensional Clinical Data to Forecast Postoperative Complications. *PLoS One* 2016; 11: e0155705. 2016/05/28 DOI: 10.1371/journal.pone.0155705.
17. Ngufor C, Murphree D, Upadhyaya S, et al. Effects of Plasma Transfusion on Perioperative Bleeding Complications: A Machine Learning Approach. *Stud Health Technol Inform* 2015; 216: 721–725. 2015/08/12. [PubMed: 26262146]
18. Weller GB, Lovely J, Larson DW, et al. Leveraging electronic health records for predictive modeling of post-surgical complications. *Stat Methods Med Res* 2018; 27: 3271–3285. 2018/01/05 DOI: 10.1177/0962280217696115. [PubMed: 29298612]
19. FitzHenry F, Murff HJ, Matheny ME, et al. Exploring the frontier of electronic health record surveillance: the case of postoperative complications. *Medical care* 2013; 51: 509–516. 2013/05/16 DOI: 10.1097/MLR.0b013e31828d1210. [PubMed: 23673394]
20. Hollis RH, Graham LA, Lazenby JP, et al. A Role for the Early Warning Score in Early Identification of Critical Postoperative Complications. *Annals of surgery* 2016; 263: 918–923. 2015/12/23 DOI: 10.1097/sla.0000000000001514. [PubMed: 26692076]
21. Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *Jama* 2011; 306: 848–855. 2011/08/25 DOI: 10.1001/jama.2011.1204. [PubMed: 21862746]
22. Colborn KL, Bronsert M, Amioka E, et al. Identification of surgical site infections using electronic health record data. *Am J Infect Control* 2018 2018/06/17 DOI: 10.1016/j.ajic.2018.05.011.

23. Colborn KL, Bronsert M, Hammermeister K, et al. Identification of urinary tract infections using electronic health record data. *Am J Infect Control* 2018 2018/12/14 DOI: 10.1016/j.ajic.2018.10.009.
24. Daley J, Khuri SF, Henderson W, et al. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *Journal of the American College of Surgeons* 1997; 185: 328–340. 1997/11/05. [PubMed: 9328381]
25. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. *National VA Surgical Quality Improvement Program. Annals of surgery* 1998; 228: 491–507. 1998/10/28. [PubMed: 9790339]
26. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine* 2015; 162: W1–73. 2015/01/07 DOI: 10.7326/m14-0698. [PubMed: 25560730]
27. Zou H and Hastie T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B (Methodological)* 2005; 67: 301–320.
28. Friedman J, Hastie T and Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 2010; 33: 1–22. [PubMed: 20808728]
29. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; 3: 32–35. [PubMed: 15405679]
30. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12: 77 DOI: 10.1186/1471-2105-12-77. [PubMed: 21414208]

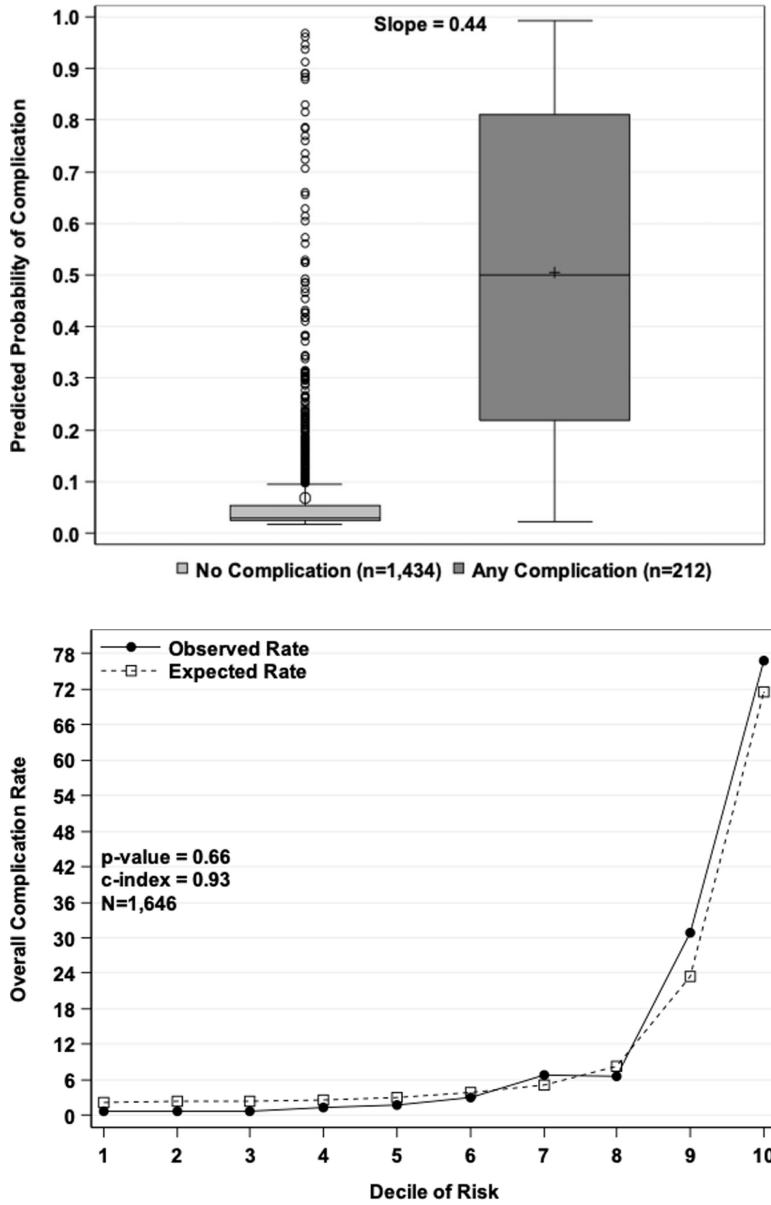


Figure 1.
a) Discrimination plot. Values on the x-axis are postoperative complication status from the NSQIP test set and values on the y-axis are predicted probabilities from the model fit to the test set. **b) Hosmer-Lemeshow Calibration plot.** Values on the x-axis are deciles of predicted risk of any complication and values on the y-axis are rates of complications for each decile. The two different lines are observed and expected rates.

Table 1.

Selected Characteristics of University of Colorado Hospital National Surgical Quality Improvement Program sample by Any Complication Status, July 1, 2013 to November 1, 2016.

| | All Patients N (%) | Patients with Any Complications N (%) | Patients with No Complication N (%) | P value ^a |
|----------------------------------|--------------------|---------------------------------------|-------------------------------------|----------------------|
| Characteristics | 6,840 | 922 (13.5) | 5,918 (86.5) | |
| Age, years, mean (SD) | 53.5 (16.4) | 58.6 (15.0) | 52.7 (16.4) | <0.0001 |
| Gender | | | | |
| Female | 3,843 (56.2) | 527 (13.7) | 3,316 (86.3) | 0.54 |
| Male | 2,997 (43.8) | 395 (13.2) | 2,602 (86.8) | |
| Race/Ethnicity | | | | |
| White, not of Hispanic origin | 4,981 (72.8) | 690 (13.9) | 4,291 (86.2) | 0.14 |
| Hispanic origin | 801 (11.7) | 94 (11.7) | 707 (88.3) | 0.14 |
| Black, not of Hispanic origin | 490 (7.2) | 72 (14.7) | 418 (85.3) | 0.41 |
| Asian or Pacific Islander | 143 (2.1) | 14 (9.8) | 129 (90.2) | 0.22 |
| American Indian or Alaska native | 20 (0.3) | 1 (5.0) | 19 (95.0) | 0.51 |
| Null/unknown | 405 (5.9) | 51 (12.6) | 354 (87.4) | 0.65 |
| Primary surgeon specialty | | | | |
| Orthopedic surgery | 1,795 (26.2) | 191 (10.6) | 1,604 (89.4) | <0.0001 |
| General surgery | 1,666 (24.4) | 239 (14.4) | 1,427 (85.7) | 0.23 |
| Gynecologic surgery | 664 (9.7) | 96 (14.5) | 568 (85.5) | 0.44 |
| Urology | 662 (9.7) | 95 (14.3) | 567 (85.7) | 0.47 |
| Neurosurgery | 640 (9.4) | 110 (17.2) | 560 (82.8) | 0.005 |
| Otolaryngology | 586 (8.6) | 29 (4.9) | 557 (95.1) | <0.0001 |
| Thoracic surgery | 345 (5.0) | 56 (16.2) | 289 (83.8) | 0.12 |
| Vascular surgery | 250 (3.7) | 83 (33.2) | 167 (66.8) | <0.0001 |
| Plastic surgery | 232 (3.4) | 23 (9.9) | 209 (90.1) | 0.12 |
| Year of operation | | | | |
| 2013 | 984 (14.4) | 141 (14.3) | 843 (85.7) | 0.39 |
| 2014 | 2,136 (31.2) | 315 (14.7) | 1,821 (85.3) | 0.04 |
| 2015 | 2,074 (30.3) | 254 (12.2) | 1,820 (87.8) | 0.05 |
| 2016 | 1,646 (24.1) | 212 (12.9) | 1,434 (87.1) | 0.43 |

Values are n (row percent) unless otherwise specified. Abbreviations: SD, standard deviation.

^aFisher's exact or t-test. For multiple categories, p-value indicates comparison to all other categories. P-value in bold if <0.05.

Table 2.

Frequency and Percent of each of the 18 National Surgical Quality Improvement Program (NSQIP) complications in the University of Colorado Hospital NSQIP sample (n=6,840), July 1, 2013 to November 1, 2016.

| Complication | N (%) |
|---------------------------------|--------------|
| Bleeding | 462 (6.8) |
| Urinary tract infection | 134 (2.0) |
| Superficial SSI | 106 (1.6) |
| Pneumonia | 97 (1.4) |
| Sepsis | 86 (1.3) |
| DVT/thrombophlebitis | 77 (1.1) |
| Organ space SSI | 66 (1.0) |
| Unplanned intubation | 55 (0.8) |
| Septic shock | 45 (0.7) |
| Deep incisional SSI | 40 (0.6) |
| Ventilator >48 hours | 38 (0.6) |
| Wound disruption | 37 (0.5) |
| Pulmonary embolism | 35 (0.5) |
| Cardiac arrest requiring CPR | 19 (0.3) |
| Acute renal failure | 18 (0.3) |
| Myocardial infarction | 14 (0.2) |
| Stroke | 14 (0.2) |
| Progressive renal insufficiency | 13 (0.2) |

Abbreviations: CPR, cardiopulmonary resuscitation; DVT, deep vein thrombosis; SSI, surgical site infection.

Table 3.

Electronic Health Record Predictors (Categorized by type) by National Surgical Quality Improvement Program Complication Status

| Variable Type | All Patients N (%) | Patients with Any Complication N (%) | Patients with No Complication N (%) |
|--|--------------------|---|--|
| | (n=6,840) | (n=922) | (n=5,918) |
| At least one of the 60 ICD-9 codes | | | |
| No | 3,596 (52.6) | 283 (7.9) | 3,313 (92.1) |
| Yes | 3,244 (47.4) | 639 (19.7) | 2,605 (80.3) |
| At least one of the 63 CPT codes | | | |
| No | 2,306 (33.7) | 41 (1.8) | 2,265 (98.2) |
| Yes | 4,534 (66.3) | 881 (19.4) | 3,653 (80.6) |
| At least one of the 39 medications | | | |
| No | 4,561 (66.7) | 168 (3.7) | 4,393 (96.3) |
| Yes | 2,279 (33.3) | 754 (33.1) | 1,525 (66.9) |
| CPT-specific Complication event rate, median (IQR) | 7.0 (3.1–15.8) | 22.7 (12.4–38.6) | 5.9 (2.3–13.2) |

Abbreviations: CPT, common procedural terminology; ICD, international classification of disease; IQR, interquartile range.

Fisher's exact p-values for binary variables and Wilcoxon rank sum for CPT-specific complication event rate. All p values were <0.0001.

Table 4.Performance of the Model Fit to the Test Data Set using Youden's *J* Statistic as the Decision Threshold.

| Performance measure | Performance Statistic (95% Bootstrap CI) |
|---------------------------|--|
| Training set years | 2013–2015 |
| Sample size | 5,194 |
| Any complication rate (%) | 710 (13.7) |
| Test set year | 2016 |
| Sample size | 1,646 |
| Any complication rate (%) | 121 (12.9) |
| Threshold | 0.11 |
| Specificity ^a | 88 (87, 90) |
| Sensitivity ^a | 83 (79, 88) |
| Accuracy ^a | 88 (86, 89) |
| NPV ^a | 97 (97, 98) |
| PPV ^a | 52 (48, 56) |
| False negatives | 35 (25, 45) |
| False positives | 165 (142, 190) |
| AUC ^a | 93 |

Abbreviations: AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PV, positive predictive value.

^aValues multiplied by 100.