HHS Public Access

Author manuscript

Am J Infect Control. Author manuscript; available in PMC 2020 April 01.

Published in final edited form as:

Am J Infect Control. 2019 April ; 47(4): 371–375. doi:10.1016/j.ajic.2018.10.009.

Identification of urinary tract infections using electronic health record data

Kathryn L. Colborn, PhD[1,*], **Michael Bronsert, PhD**[2,3], **Karl Hammermeister, MD**[2,3,4], **William G. Henderson, PhD**[1,2,3], **Abhinav B. Singh, MD**[3], **Robert A. Meguid, MD, MPH**[2,3]

[1]University of Colorado Anschutz Medical Campus, Colorado School of Public Health, Department of Biostatistics and Informatics, Aurora, CO, USA,

[2]University of Colorado Anschutz Medical Campus, Adult and Child Consortium for Health Outcomes Research and Delivery Science, Aurora, CO, USA,

[3]Surgical Outcomes and Applied Research program, Department of Surgery, University of Colorado School of Medicine, Aurora CO, USA,

[4]University of Colorado Anschutz Medical Campus, School of Medicine, Department of Cardiology, Aurora, CO, USA.

Abstract

Background: Population ascertainment of postoperative urinary tract infections (UTI) is time-consuming and expensive, as it often requires manual chart review. Using the American College of Surgeons National Surgical Quality Improvement Program (NSQIP) UTI status of patients who underwent an operation at the University of Colorado Hospital, we sought to develop an algorithm for identifying UTIs using data from the electronic health record (EHR).

Methods: Data were split into training (operations occurring between 2013–2015) and test (operations in 2016) sets. A binomial generalized linear model with an elastic-net penalty was used to fit the model and carry out variables selection. International classification of disease codes, common procedural terminology (CPT) codes, antibiotics, catheterization, and CPT-specific UTI event rate were included as predictors. The Youden's *J* statistic was used to determine the optimal classification threshold.

Results: Of 6,840 patients, 134 (2.0%) had a UTI. The model achieved 92% specificity, 80% sensitivity, 100% negative predictive value, 16% positive predictive value, and an area under the curve of 0.94 using a decision threshold of 0.03.

Conclusions: A model with 14 predictors from the EHR identifies UTIs well, and it could be used to scale up UTI surveillance or to estimate the impact of large-scale interventions on UTI rates.

[*]Corresponding author: Kathryn Colborn, PhD, University of Colorado Anschutz Medical Campus, Mail Stop B119, 13001 East 17th Place, Room C3011, Aurora, Colorado 80045, Phone: 303-724-0648, Kathryn.colborn@ucdenver.edu.

Conflicts of Interest: All authors report no conflicts of interest related to this article.

Keywords

NSQIP; urinary tract infection; UTI; elastic-net; supervised learning; postoperative complication

Introduction

Surveillance of postoperative complications has been conducted at participating hospitals through the National Surgical Quality Improvement Program (NSQIP), developed and implemented in the U.S. Department of Veterans Affairs since 1991 and subsequently in the private sector through the American College of Surgeons (ACS) since 2005. At participating hospitals, trained surgical clinical nurse reviewers collect preoperative through 30-day postoperative data on a representative sample of patients who underwent surgery to determine which patients experienced one or more of the 22 complications tracked by NSQIP. The NSQIP data provide high quality outcomes data on these 22 complications for more than 5.4 million patients. Unfortunately, manual chart review is time-consuming and expensive, so at larger institutions only ~15% of operated patients are typically evaluated. Urinary tract infections (UTI) are one type of infectious complication tracked. According to the NSQIP database, 1.6% of patients developed a UTI within 30 days of their initial operation between 2005–2012.[1]

It is desirable to develop electronic methods for identifying UTIs for several reasons. First, it would augment NSQIP surveillance by permitting assessment of a larger number of patients from each institution. Second, it would allow researchers to evaluate the impact of large-scale interventions to prevent postoperative UTIs without the need for manual chart review. Some effort has been made over the past decade to identify UTIs and catheter-associated-UTIs (CAUTIs). Zhan et al. were specifically interested in identifying CAUTIs in Medicare claims data but found that codes for urinary catheter placement and CAUTI are rarely recorded.[2] To improve sensitivity, they created an approximate catheterization variable based on the patients' surgical procedure, and their deterministic method consisting of international classification of disease version 9 (ICD-9) codes achieved 65% sensitivity with CAUTI identification.[2] Landers et al. developed seven sets of decision rules for classifying patients with a UTI that included symptom codes, diagnosis codes, and procedural codes. All seven rules averaged about 56% sensitivity.[3] They also reported issues with low rates of recording of catheter placement and symptoms, which they attributed to the low sensitivities achieved.[3] In response to various articles reporting low rates of reporting of codes for catheters, CAUTIs and CAUTI symptoms,[2–6] Gundlappali et al. and Branch-Elliman et al. implemented natural language processing (NLP) on the notes of patients' medical charts.[7,8] Although both studies reported the ability to detect catheters in the notes using NLP, they both found that standard methods (i.e., the use of diagnosis codes) performed better.[7,8]

Considering the success with identifying surgical site infections (SSIs) using electronic health record (EHR) data and supervised learning with penalized regression reported by Colborn et. al,[9] we sought to develop a model for identifying UTIs using similar methodology.

Methods

Data:

Using the same data set described by Colborn et al.,[9] we identified 6,840 patients who underwent surgery at University of Colorado Hospital (UCH) from July 1, 2013 to November 1, 2016 in the local NSQIP database. After receiving IRB approval for this study, we sent patient health identifiers (PHI) for these patients to the University of Colorado Health Data Compass (HDC), a University-sponsored data warehouse that contains EHR data on all UCH patients. HDC linked relevant available data to our patients from the local NSQIP database for up to 30 days after their date of surgery. These data included demographic characteristics, ICD-9 codes, current procedural terminology codes (CPT), and drug codes and names. Outcome status, i.e., presence or absence of a UTI, came from the UCH ACS NSQIP database.

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.[10] The dataset was split into training and test datasets, with a temporal split, where the models were trained on the older observations and tested on the more recent observations. 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, and the prevalence of UTI 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 between 2013–15 (N=5,194; no UTI=5,090 [98.0%], UTI=104 [2.0%]) and the test set consisted of operations performed in 2016 (N=1,646; no UTI=1,616 [98.2%], UTI=30 [1.8%]). We formulated a comprehensive model consisting of all variables flagged by two separate clinicians (KH, RM) as potentially important for identifying and/or treating a UTI. This list consisted of ICD-9 codes, CPT codes, and antibiotics. We also included the CPT-specific UTI event rate, defined as the rate of UTIs experienced by NSQIP patients for each particular surgical procedure estimated from data across all 800+ NSQIP-participating hospitals, and a variable indicating whether or not the patient had been catheterized. A binomial generalized linear model (GLM) with an elastic-net penalty was used to estimate coefficients and to carry out variables selection. The elastic-net penalty is a combination of the least absolute shrinkage and selection operator (lasso) and ridge penalties, and it tends to handle correlated covariates better than the lasso.[11] Penalized regression was used primarily because the number of covariates greatly outnumbered the number of individuals who had a UTI in this dataset. Furthermore, penalized regression methods, as opposed to variable selection methods, such as stepwise or best subset selection, achieve the goals of 1) reducing the chance of overfitting to the training data, and 2) conducting variable selection in a continuous manner by allowing some coefficients to be zero (as opposed to just throwing variables out based on a p-value). Ten-fold cross-validation was performed to determine the optimal value for the lasso penalty (i.e., lambda) using the glmnet package[12] in R (R Foundation for Statistical Computing, Vienna, Austria). The elastic-net tuning parameter (i.e., alpha) was set at 0.5. We estimated the predicted probabilities of UTI in the test set

using the fitted model and performed classification using the Youden's *J* statistic[13] estimated in the training data set. 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 UTIs in the test data. Youden's *J* and all other performance statistics were estimated using the pROC package in R.[14] All ICD-9, CPT, and antibiotic independent variables were coded as positive only if the codes were observed between 3 and 30 days after the initial operation. This interval was chosen for several reasons. First, a UTI is unlikely to be attributed to the operation if it is found within two days of the operation. Second, it excludes most antibiotics administered for perioperative prophylaxis, and, finally, any UTIs identified beyond 30 days of the operation were unlikely to be caused by the operation. The only exception to this range was for the catheter variable, which was included if a record of placement of an indwelling urinary catheter occurred between 0 and 30 days from the day of operation.

Sensitivity analysis was performed using random forest[15] and the same variables, training and test sets, and thresholds. Analyses were performed using the random Forest package in R.[16]

Results

Table 1 demonstrates the relationship between individual characteristics and UTI status. Patients who experienced UTIs tended to be older (57.1 years vs. 53.5 years, UTI versus no UTI, respectively, $p<0.001$). Patients undergoing gynecologic and urologic procedures had higher rates of UTIs than patients undergoing other types of operations (gynecology 5.3% [$p<0.001$], urology 3.6% [$p=0.003$]). Females had twice the rate of UTIs as males (2.5% females vs. 1.2% males, $p<0.001$). There were no significant differences by race/ethnicity or year of operation.

Table 2 summarizes a subset of the predictors by UTI status. Presence of an indwelling urinary catheter during surgery was associated with a higher rate of UTI (2.9% with catheter vs. 0.6% no catheter, $p<0.001$). ICD-9 code 599, "UTI, site unspecified", was a clear indicator of UTI in these EHR data (33.8% UTI vs. 1.3% no UTI, $p<0.001$), as well as ICD-9 code 595.9, "cystitis" (40% UTI vs. 1.9% no UTI, $p<0.001$). Of patients with a UTI, 8.5% had an antibiotic prescription vs. 0.3% of those who did not have a UTI ($p<0.001$). Urine culture was also commonly coded among patients with a UTI (11.9% UTI vs. 0.4% no UTI, $p<0.001$). CPT-specific UTI event-rate was more than three times higher in those with a UTI compared to those who did not have a UTI (Table 2, $p<0.001$).

Table 3 summarizes the variables with non-zero coefficient values estimated from the model. Recall, the penalty terms applied allow for some coefficient values to be exactly equal to zero, which results in these variables being removed from the model. Of 44 variables, 14 were selected (i.e., not equal to zero); seven were antibiotics, four were CPT codes, and two were ICD-9 codes. CPT-specific UTI-event-rate was also selected. The variables were sorted by descending odds ratios (OR). All variables were dichotomous, except for CPT-specific UTI-event rate which was continuous; that is one reason why the OR is lower for UTI-event rate compared to the ORs of the other variables. The relationship between CPT-specific UTI-

event-rate and UTI was almost perfectly linear when plotted using a cubic smoothing spline with 4 degrees of freedom (plot not shown). The selected predictors all make clinical sense and are not surprising.

Table 4 summarizes the classification performance of the model in the test set. The results in Table 4 suggest that this is a very good model for classification of UTIs in these data. The AUC was 0.94, sensitivity 0.80, specificity 0.92, and accuracy 0.92 when using the Youden's *J* threshold estimated in the training set.

Figure 1 is a discrimination plot, which has been described by Steyerberg et al.[17] It displays the predicted probabilities from the model fit to the test set by the NSQIP-identified UTIs (i.e., observed outcomes). There is fairly good discrimination, as the predicted probabilities for those with a UTI are much higher than those without. The discrimination slope is 0.22 (i.e., the difference in mean predicted probabilities between the two classes), which suggests good discrimination by the model.

Sensitivity analysis:

We also used random forest[15] and the same variables to obtain predicted probabilities (which were just the proportion of the trees that classified patients as having had a UTI; each individual tree's classification being a majority vote in the terminal node). Random forest performed similarly to the elastic-net when fit to the test set (AUC 0.91; threshold 0.04; specificity 0.95; sensitivity 0.80; accuracy 0.95; NPV 1.00; PPV 0.24; false negatives 6; false positives 74).

Discussion

In this study, we developed a model for identifying UTIs in patients who underwent major surgery at the University of Colorado Hospital using EHR data, NSQIP outcomes data, and a GLM with an elastic-net penalty. This model was tested on a temporal hold-out set and achieved an AUC of 0.94. It correctly classified 80% of those with a UTI and 92% of those who did not experience a UTI. This model could be useful for ascertaining UTI status of patients after large-scale interventions for preventing UTIs. It could also be useful for supplementing NSQIP surveillance activities so that a complete patient dataset can be developed instead of a limited sample.

Using the methodology described by Colborn et al. for identifying SSIs, slightly better results were achieved by a model for identifying UTIs with fewer required variables.[9] This is likely due to the greater strength of association between specific variables and UTI. Interestingly, despite the difference in proportions of patients with urinary catheters between those with a UTI and those who did not have a UTI, the elastic-net model did not select this variable. Therefore, it was excluded when fitting the final model, as 63 patients were missing these data (all of these patients did not have a UTI); and we did not want to exclude these patients from the analysis. We also chose not to perform imputation on this variable, as it was missing for less than 1% of patients. Although we did not include catheterization in our final model reported in Table 3, we suggest including it when applying similar methodology to different UTI data sets, especially given the continued high risk of CAUTIs

reported in the literature.[5,18–20] However, as we pointed out in the introduction, catheter identification from EHR data can be difficult,[2,5–8,21–23] and it was not clear whether the EHR data used in this study properly captured the catheter variable. It is also important to point out that we were not specifically looking for CAUTIs in this analysis. Our goal was to develop a model that identified NSQIP-defined UTIs using EHR data. A recent study by Sopirala et. al highlights the difficulty in conducting CAUTI surveillance.[24] In that study, the authors reported very different rates of CAUTIs in the same population when using two different definitions. Furthermore, because we used NSQIP data to develop this model, it might not be generalizable to other definitions of UTIs. For example, Ju et. al. found that the NSQIP definition of SSI estimated 13.5% of patients had SSIs compared to only 5.7% found using the National Healthcare Safety Network (NHSN) definition in an analysis of 16 hospitals.[25] This might not be surprising, considering they use different definitions for the denominators (NSQIP uses CPT codes, whereas NHSN uses ICD-9 codes to determine operation performed) and numerators (NSQIP uses trained nurse reviewers who call the patients to ascertain additional information that is not seen in the chart). Furthermore, the authors found that collection methods for NHSN differed across hospitals and that outpatient cases were often missed under current NHSN practices. Therefore, we recommend fitting separate models using similar methodology to data that are generated from different definitions than NSQIP's.

Another interesting finding in this study was that there were 98 (1.5%) patients with no UTI according to NSQIP who had an ICD-9 code of 599 ("urinary tract infection; site not specified"), which might be an indication that the ACS NSQIP definition for UTI is more rigorous than that used by providers documenting occurrence of a UTI or by coders assigning ICD-9 codes post hoc. Alternatively, it is possible that the surgical clinical reviewers misclassified these patients.

The findings from this study suggest that UTIs can be identified in the EHR using 14 variables and a linear model. Future work should explore the effect of including claims data and perhaps symptoms and/or catheter data derived from NLP in clinical documentation.

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.

References

1. Meguid RA, Bronsert MR, Juarez-Colunga E, Hammermeister KE & Henderson WG Surgical Risk Preoperative Assessment System (SURPAS): I. Parsimonious, Clinically Meaningful Groups of Postoperative Complications by Factor Analysis. Ann Surg 263, 1042–1048, doi:10.1097/sla.0000000000001669 (2016). [PubMed: 26954897]
2. Zhan C et al. Identification of hospital-acquired catheter-associated urinary tract infections from Medicare claims: sensitivity and positive predictive value. Med Care 47, 364–369, doi:10.1097/MLR.0b013e31818af83d (2009). [PubMed: 19194330]

3. Landers T et al. A comparison of methods to detect urinary tract infections using electronic data. Joint Commission journal on quality and patient safety 36, 411–417 (2010). [PubMed: 20873674]

4. Tanushi H, Kvist M & Sparrelid E Detection of healthcare-associated urinary tract infection in Swedish electronic health records. Studies in health technology and informatics 207, 330–339 (2014). [PubMed: 25488239]

5. Trautner BW et al. Quality gaps in documenting urinary catheter use and infectious outcomes. Infection control and hospital epidemiology 34, 793–799, doi:10.1086/671267 (2013). [PubMed: 23838219]

6. Wald HL, Bandle B, Richard A & Min S Accuracy of electronic surveillance of catheter-associated urinary tract infection at an academic medical center. Infection control and hospital epidemiology 35, 685–691, doi:10.1086/676429 (2014). [PubMed: 24799645]

7. Branch-Elliman W, Strymish J, Kudesia V, Rosen AK & Gupta K Natural Language Processing for Real-Time Catheter-Associated Urinary Tract Infection Surveillance: Results of a Pilot Implementation Trial. Infection control and hospital epidemiology 36, 1004–1010, doi:10.1017/ice.2015.122 (2015). [PubMed: 26022228]

8. Gundlapalli AV et al. Detecting the presence of an indwelling urinary catheter and urinary symptoms in hospitalized patients using natural language processing. Journal of biomedical informatics 71s, S39–s45, doi:10.1016/j.jbi.2016.07.012 (2017). [PubMed: 27404849]

9. Colborn KL et al. Identification of surgical site infections using electronic health record data. Am J Infect Control, doi:10.1016/j.ajic.2018.05.011 (2018).

10. Moons KG et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Annals of internal medicine 162, W1–73, doi:10.7326/m14-0698 (2015). [PubMed: 25560730]

11. Zou H & Hastie T Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B (Methodological) 67, 301–320 (2005).

12. Friedman J, Hastie T & Tibshirani R Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software 33, 1–22 (2010). [PubMed: 20808728]

13. Youden WJ Index for rating diagnostic tests. Cancer 3, 32–35 (1950). [PubMed: 15405679]

14. Robin X et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12, 77, doi:10.1186/1471-2105-12-77 (2011). [PubMed: 21414208]

15. Breiman L Random Forests. Machine Learning 45, 5–32 (2001).

16. Liaw AW, Matthew. Classification and Regression by randomForest. R News 2, 18–22 (2002).

17. Steyerberg E Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. (Springer, 2010).

18. Fletcher KE et al. Qualitative validation of the CAUTI Guide to Patient Safety assessment tool. Am J Infect Control 44, 1102–1109, doi:10.1016/j.ajic.2016.03.051 (2016). [PubMed: 27339790]

19. Meddings J et al. Reducing unnecessary urinary catheter use and other strategies to prevent catheter-associated urinary tract infection: an integrative review. BMJ quality & safety 23, 277–289, doi:10.1136/bmjqs-2012-001774 (2014).

20. Parker V et al. Avoiding inappropriate urinary catheter use and catheter-associated urinary tract infection (CAUTI): a pre-post control intervention study. BMC health services research 17, 314, doi:10.1186/s12913-017-2268-2 (2017). [PubMed: 28464815]

21. Hsu HE et al. An electronic surveillance tool for catheter-associated urinary tract infection in intensive care units. Am J Infect Control 43, 592–599, doi:10.1016/j.ajic.2015.02.019 (2015). [PubMed: 25840717]

22. Sanger PC et al. Electronic Surveillance For Catheter-Associated Urinary Tract Infection Using Natural Language Processing. AMIA … Annual Symposium proceedings. AMIA Symposium 2017, 1507–1516 (2017). [PubMed: 29854220]

23. Shepard 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 42, e33–36, doi:10.1016/j.ajic.2013.12.005 (2014). [PubMed: 24581026]

24. Sopirala MM, Syed A, Jandarov R & Lewis M 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 46, 743–746, doi:10.1016/j.ajic.2018.01.019 (2018). [PubMed: 29551201]

25. Ju MH et al. A comparison of 2 surgical site infection monitoring systems. JAMA surgery 150, 51–57, doi:10.1001/jamasurg.2014.2891 (2015). [PubMed: 25426765]

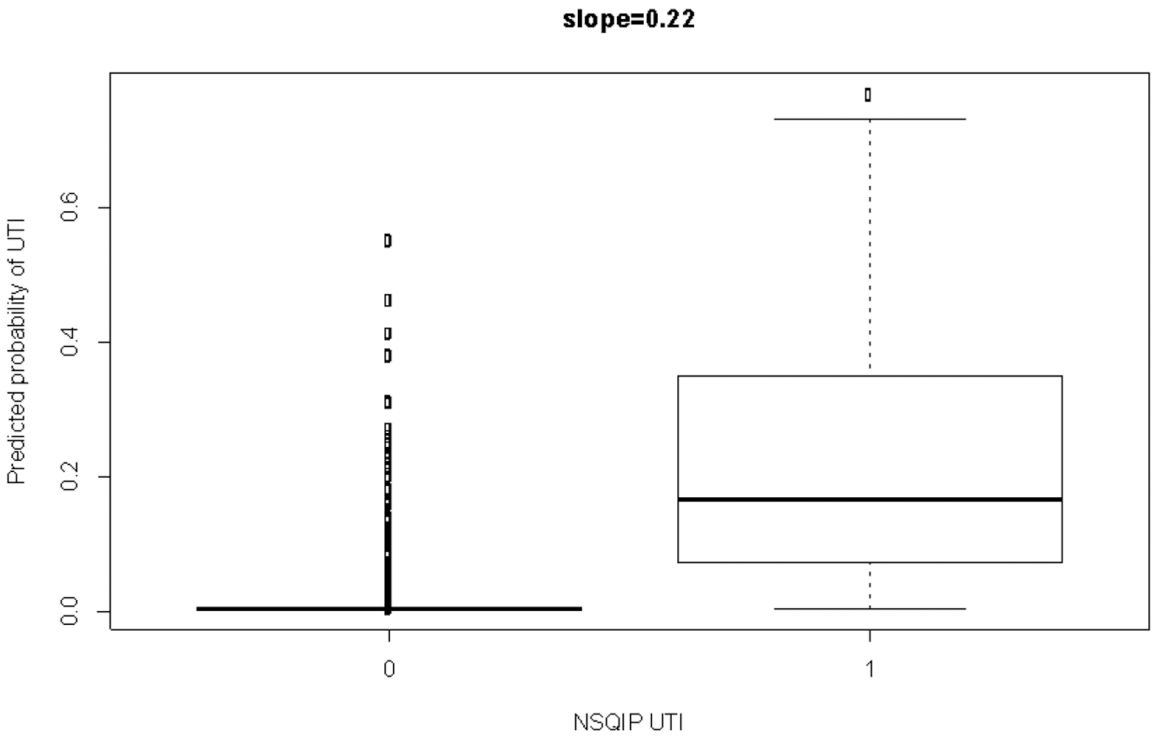

Figure 1. Discrimination plot.

Values on the x-axis are observed UTI status from the NSQIP test set and values on the y-axis are predicted probabilities from the model fit to the test set.

Table 1.

University of Colorado Hospital National Surgical Quality Improvement Program population by UTI status (N=6,840).

	All	UTI	No UTI	
	6,840	**134 (2.0)**	**6,706 (98.0)**	**P value‡**
Age, years, mean (SD)	53.5 (16.4)	57.1 (15.6)	53.5 (16.4)	<0.001
Gender				
Female	3,843 (56.2)	97 (2.5)	3,746 (97.5)	<0.001
Male	2,997 (43.8)	37 (1.2)	2,960 (98.8)	
Race/Ethnicity				
White, not of Hispanic origin	4,981 (72.8)	99 (2.0)	4,882 (98.0)	0.8
Hispanic origin	801 (11.7)	15 (1.9)	786 (98.1)	1.0
Black, not of Hispanic origin	490 (7.2)	11 (2.2)	479 (97.8)	0.6
Asian or Pacific Islander	143 (2.1)	3 (2.1)	140 (97.9)	0.8
American Indian or Alaska native	20 (0.3)	0 (0.0)	20 (100.0)	1.0
Null/unknown	405 (5.9)	6 (1.5)	399 (98.5)	0.6
Primary surgeon specialty				
Orthopedics	1,795 (26.2)	22 (1.2)	1,773 (98.8)	0.01
General surgery	1,666 (24.4)	30 (1.8)	1,636 (98.2)	0.7
Gynecology	664 (9.7)	35 (5.3)	629 (94.7)	<0.001
Urology	662 (9.7)	24 (3.6)	638 (96.4)	0.003
Neurosurgery	640 (9.4)	12 (1.9)	628 (98.1)	1.0
Otolaryngology	586 (8.6)	1 (0.2)	585 (99.8)	<0.001
Thoracic	345 (5.0)	7 (2.0)	338 (98.0)	0.8
Vascular	250 (3.7)	2 (0.8)	248 (99.2)	0.2
Plastics	232 (3.4)	1 (0.4)	231 (99.6)	0.09
Year of operation				
2013	984 (14.4)	25 (2.5)	959 (97.5)	0.2
2014	2,136 (31.2)	48 (2.2)	2,088 (97.8)	0.3
2015	2.074 (30.3)	31 (1.5)	2,043 (98.5)	0.07
2016	1,646 (24.1)	30 (1.8)	1,616 (98.2)	0.7

Values are n (row percent) unless otherwise specified. Abbreviations: SD, standard deviation; UTI, urinary tract infection.

‡Fisher's exact or t-test. For multiple categories, p-value indicates comparison to all other categories.

Table 2.

Electronic Health Record Predictors by National Surgical Quality Improvement Program UTI status.

	All	UTI	No UTI	
	6,840	**134 (2.0)**	**6,706 (98.0)**	**P value**‡
Catheter				
Yes	4,020 (59.3)	117 (2.9)	3,903 (97.1)	<0.001
No	2,757 (40.7)	17 (0.6)	2,740 (99.4)	
ICD-9 599: UTI, site unspecified				
Yes	148 (2.2)	50 (33.8)	98 (66.2)	<0.001
No	6,692 (97.8)	84 (1.3)	6,608 (98.7)	
ICD-9 595.9: Cystitis, unspecified				
Yes	10 (0.1)	4 (40.0)	6 (60.0)	<0.001
No	6,830 (99.1)	130 (1.9)	6,700 (98.1)	
Antibiotic (any)				
Yes	1,365 (20.0)	116 (8.5)	1,249 (91.5)	<0.001
No	5,475 (80.0)	18 (0.3)	5,457 (99.7)	
CPT 87186, 87086, or 87088: Urine culture				
Yes	925 (13.5)	110 (11.9)	815 (88.1)	<0.001
No	5,915 (86.5)	24 (0.4)	5,891 (99.6)	
CPT-specific UTI event rate, median (IQR)	0.91 (0.34, 2.23)	3.14 (1.59, 4.64)	0.89 (0.34, 2.20)	<0.001

Values are n (row percent) unless otherwise specified; Abbreviations: CPT, common procedural terminology; ICD, international classification of disease; IQR, interquartile range; UTI, urinary tract infection.

‡Fisher's exact p-values for binary variables and Wilcoxon rank sum for CPT-specific UTI event rate;

Table 3.

All non-zero coefficient estimates using a binomial generalized linear model with an elastic-net penalty for UTI fit to the training data.

Variable	Coefficient	OR
Cetirizine	1.54	4.66
CPT 87086: Culture, bacterial; quantitative, colony count, urine	1.32	3.74
ICD-9 599: Urinary tract infection, site not specified	1.32	3.76
CPT 87186: Aerobic Susceptibility: Gram Positive Minimum Inhibitory Concentration Panel Urine/Non-Urine	1.30	3.69
Nitrofurantoin	1.10	3.00
Ciprofloxacin	0.79	2.20
CPT 87088: Culture, bacterial; with isolation and presumptive identification of each isolates, urine	0.67	1.95
ICD-9 595.9: Cystitis, unspecified	0.61	1.83
Sulfamethoxazole	0.61	1.85
CPT 81001: Urinalysis, by dip stick or tablet reagent; automated, with microscopy	0.29	1.34
Amoxicillin	0.19	1.22
Ceftriaxone	0.13	1.14
Clindamycin	0.05	1.05
CPT-specific UTI-event-rate	0.01	1.01

Abbreviations: CPT, common procedural terminology; ICD, international classification of disease; OR, odds ratio, UTI, urinary tract infection.

Table 4.

Performance of the model fit to the test data set using Youden's *J* statistic as the decision threshold. (All values multiplied by 100 except the threshold).

Performance measure	Value
Threshold	0.03
Specificity	92
Sensitivity	80
Accuracy	92
NPV	100
PPV	16
False negatives	6
False positives	127
AUC	94

AUC, area under the curve; NPV, negative predictive value;

PPV, positive predictive value.