American Journal of Infection Control 42 (2014) 333-6

Contents lists available at ScienceDirect

American Journal of Infection Control

journal homepage: www.ajicjournal.org

Brief report

Assessing surgical site infection risk factors using electronic medical records and text mining

James D. Michelson MD [a,*], Jenna S. Pariseau BA [a], William C. Paganelli MD [b]

[a] Departments of Orthopedics and Rehabilitation, Burlington, VT
[b] Departments of Anesthesia, University of Vermont College of Medicine, Burlington, VT

Key Words:
Surgery
Surveillance

Text mining techniques to detect surgical site infections (SSI) in unstructured clinical notes were used to improve SSI detection. In conjuction with data from an integrated electronic medical record, all of the 22 SSIs detected by traditional hospital-based surveillance were found using text mining, along with an additional 37 SSIs not detected by traditional surveillance.

Given the incomplete capture of surgical site infections (SSIs) by traditional surveillance methods,[1-10] the present study was designed to improve SSI detection by combining discrete data from an integrated electronic medical record (EMR) with text mining of clinical notes to screen for evidence of SSIs.[8,10-12]

METHODS

The study cohort comprised patients undergoing elective orthopedic surgery during 2010 in the following divisions: joint replacement, foot and ankle (F&A), sports medicine, and upper extremity. Pediatric patients and patients undergoing spine or trauma surgery were excluded owing to the unique nature of those specialties and our desire to limit the evaluation to elective surgeries. Inpatient and outpatient notes (in plain text), laboratory test results, medications, and problem lists were captured in the EMR (Epic Systems, Verona, WI). The anesthesia record (PICIS, Wakefield, MA) provided information on antibiotic administration, duration of surgery, and the American Society of Anesthesia (ASA) score for perioperative anesthesia risk. Patients with preexisting infection were excluded from the study.

Selection of data elements

A total of 26,947 clinical notes were analyzed. Perioperative data are presented in Table 1. The Charlson Index was calculated using the modification of Deyo et al.[13]

Processing of notes

Notes were screened using a text mining Boolean search algorithm (dtSearch, Bethesda, MD).[10] Forty-two search items were divided into 4 categories suggesting the presence of an SSI in the clinical documentation: superficial infection, deep infection, wound problems, infection treatment. Notes containing 2 or more category terms were manually reviewed for SSIs. The identification of SSIs was based on the clinical judgment of the treating physician (not the Centers for Disease Control and Prevention definition), which was previously shown to be as accurate as the Centers for Disease Control and Prevention criteria for SSI detection.[10] Any wound treated with antibiotics was considered an SSI, whereas wound dehiscence without evidence of infection was not. An SSI was defined as infection occurring within 30 days of the index surgery.

Patients with SSIs were identified by the hospital-based infection control process based on a positive culture on readmission, on reoperation, or during an emergency department visit, or when identified by the surgeon through the department-based quality assurance process. Prospective surveillance data for all joint reconstructions were included.

* Address correspondence to James D. Michelson, MD, Department of Orthopedics and Rehabilitation, University of Vermont College of Medicine, Stafford 418A, 95 Carrigan Dr, Burlington, VT 05401.

E-mail address: james.michelson@uvm.edu (J.D. Michelson).

J.D.M. was responsible for the experimental design, data collection, and analysis and contributed to manuscript preparation. J.S.P. was responsible for the text mining analysis and contributed to manuscript preparation. W.C.P. was responsible for data collection and contributed to manuscript preparation.

This study was partially supported by a research grant from the University of Vermont School of Medicine (to J.S.P.).

Conflicts of interest: None to report.

0196-6553/$36.00 -
http://dx.doi.org/10.1016/j.ajic.2013.09.007

Table 1
Demographic data

Variable	No diabetes*	Diabetes	Total	*P* value†
Sex, n (%)				.283
Female	949 (48)	109 (51.9)	1,058 (48.4)	
Male	1,028 (52)	101 (48.1)	1,129 (51.6)	
Tobacco use (n = 2,140), n (%)				.001
No	1,754 (88.7)	203 (96.7)	1,957 (89.5)	
Yes	176 (8.9)	7 (3.3)	183 (8.4)	
Alcohol use (n = 2,142), n (%)				<.001
No	1,564 (79.1)	195 (92.9)	1,759 (80.4)	
Yes	368 (18.6)	15 (7.1)	383 (17.5)	
Charlson Index ≥3 (n = 2,095), n (%)‡				<.001
No	1,828 (96.7)	173 (84.4)	2,001 (95.5)	
Yes	62 (3.3)	32 (15.6)	94 (4.5)	
Perioperative steroid use, n (%)				>.50
No	1,686 (85.3)	176 (83.8)	1,862 (85.1)	
Yes	291 (14.7)	34 (16.2)	325 (14.9)	
ASA score (n = 2,156), n (%)				<.001
1	457 (23.4)	1 (0.5)	458 (21.2)	
2	1,238 (63.4)	86 (42.2)	1,324 (61.4)	
3	252 (12.9)	114 (55.9)	366 (17)	
4	5 (0.3)	3 (1.5)	8 (0.4)	
Division, n (%)				<.001
F&A	295 (14.9)	39 (18.6)	334 (15.3)	
Hand	740 (37.4)	64 (30.5)	804 (36.8)	
Reconstruction	367 (18.6)	86 (41)	453 (20.7)	
Sports	575 (29.1)	21 (10)	596 (27.3)	
Age, y, mean (95% CI)	52 (51.2-52.8); n = 1,977	67.8 (66-69.5); n = 210	53.5 (52.8-54.3); n = 2,187	<.001
Glucose, mg/dL, mean (95% CI)	123.3 (120.1-126.6); n = 398	156.2 (147.7-164.7); n = 193	134.1 (130.3-137.8); n = 591	<.001
BMI, kg/m^2, mean (95% CI)	28.5 (28.2-28.8); n = 1,697	34.4 (33.3-35.5); n = 184	29.1 (28.8-29.4); n = 1,881	<.001
Hemoglobin, g/dL, mean (95% CI)	11.4 (11.3-11.6); n = 846	10.3 (10-10.7); n = 173	11.3 (11.1-11.4); n = 1,019	<.001
Transfused units, mean (95% CI)	3.4 (3-3.7); n = 88	4.2 (3.3-5); n = 24	3.5 (3.2-3.9); n = 112	.026
Creatinine, mg/dL, mean (95% CI)	0.9 (0.9-0.9); n = 885	1.2 (1-1.3); n = 187	1 (0.9-1); n = 1,072	<.001

*Percentages are with respect to all the patients in either the diabetic cohort or the nondiabetic cohort.
†By the Pearson χ^2 test (for categorical data) or Mann-Whitney *U* test (for continuous data), comparing patients with and without diabetes.
‡The Charlson Index as modified by Deyo et al.[13]

Statistics

Data were analyzed with the χ^2 or Fisher exact test (for categorical variables), or the Mann-Whitney *U* test (for continuous variables) using SPSS version 12.0 (IBM, Armonk, NY). Multivariate binary logistic regression (backward-stepwise, exclusion criterion of $P > .01$) was performed using all of the available parameters. A significance level of $P < .05$ using 2-sided testing was applied.

Ethics

This study was carried out with the approval of the University of Vermont's Institutional Review Board.

RESULTS

During 2010, a total of 2407 surgeries were performed in 2187 patients aged >18 years by 10 orthopedic surgeons at our institution. Only first surgeries were included in the analysis, because second surgeries were usually performed to treat infection.

Compared with the patients without diabetes, those with diabetes were older and more likely to have undergone total joint replacement surgery, and had a higher mean BMI, Charlson Comorbidity Index score, perioperative serum glucose level, and serum creatinine level (Table 1).

SSI was detected in 59 patients. Risk factors identified on univariate analysis included older age, higher BMI, diabetes, more comorbidities, elevated serum creatinine level, and F&A surgery. In F&A surgery, operating on bone (vs soft tissue), anatomic location of the surgery, use of retained hardware, or preoperative vascular concerns were not identified as risk factors for SSI.

Table 2
Logistic regression: Risk factors for SSI

Variable	OR (95% CI)	*P* value
All patients		
F&A	4.3 (1.9-9.7)	<.001
Diabetes	2.2 (1.04-4.6)	.04
ASA score >2	2.5 (1.3-5.0)	.007
Charlson Index ≥3	2.3 (0.95-5.5)	.064
Patients with diabetes		
Alcohol use	4.4 (1.01-19.2)	.048
Creatinine	1.6 (1.1-2.4)	.019
Patients without diabetes		
F&A	7.4 (2.3-24.0)	.001
ASA score >2	3.2 (1.51-6.8)	.002
Age	1.02 (0.998-1.04)	.08

NOTE. Logistic regression was performed using the backward-stepwise method, with a selection criterion of $P > .10$ for exclusion.
Regressions for the overall population and those without diabetes excluded serum creatinine level as a factor owing to a low rate of perioperative measurement.
The numbers of patients included in the regressions were as follows: all patients, n = 1,773 (81.1% of total); diabetic patients, n = 196 (93.3%); nondiabetic patients, n = 1,634 (82.7%).

Risk factors for SSI identified on multivariate analysis included diabetes, ASA score >2, and F&A surgery (Table 2). Variables not predictive of SSI included age, sex, BMI, Charlson Index, tobacco or alcohol use, anemia, duration of surgery, surgical attending, transfusion status, antibiotic timing, use of statin medication, perioperative steroid use, and rheumatoid arthritis. Risk factors for SSI in patients with diabetes included elevated serum creatinine level and alcohol, and those in patients without diabetes were F&A surgery and an ASA score >2.

The text mining screening detected 100% of the infections detected by the traditional infection surveillance system, and also