Published in final edited form as:

Surgery. 2018 December ; 164(6): 1300–1305. doi:10.1016/j.surg.2018.05.008.

Autonomous Detection, Grading, and Reporting of Post-Operative Complications using Natural Language Processing

Luke V. Selby, MD, MS^{1,*}, Wazim R. Narain, PhD, MPH^{2,*}, Ashley Russo, MD¹, Vivian E Strong, MD¹, Peter Stetson, MD, MA²

¹:Departments of Surgery, Memorial Sloan Kettering Cancer Center

²:Health Informatics, Memorial Sloan Kettering Cancer Center

Abstract

Introduction: Natural language processing (NLP), a computer science technique that allows interpretation of narrative text, is infrequently used to identify surgical complications. We designed an NLP algorithm to identify and grade the severity of deep venous thrombosis (DVT) and pulmonary embolisms (PE) (together: VTE).

Methods: Patients from our 2011 – 2014 NSQIP cohorts with a duplex ultrasound or a computerized tomography angiography (CTA) of the chest performed within 30 days of surgery were divided into training and validation datasets. A “bag of words” approach classified the reports; other electronic health record data classified the VTE’s severity.

Results: Of the 10,295 NSQIP patients, 251 were used in our DVT validation cohort (273 total ultrasounds) and 506 in our PE cohort (552 total CTAs). For DVTs the sensitivity and specificity were 85.1% and 94.6%, while for PEs they were 90% and 98.7%. Most discordances were due lack of imaging documentation of a DVT (28/41, 68.3%) or PE (6/6, 100%). Most DVTs (28 patients, 54.6%) and PEs (25 patients, 75.8%) required administration of therapeutic intravenous or subcutaneous anticoagulation.

Conclusions: NLP can reliably detect the presence of post-operative VTEs & its use should be expanded for the detection of other conditions from narrative documentation.

Keywords

Natural Language Processing; Venous Thromboembolism; Surgical Secondary Events; Surgical Complications; Surgical Quality Improvement

Corresponding Author: Vivian E. Strong, MD FACS, Department of Surgery, Gastric and Mixed Tumor Service, Memorial Sloan Kettering Cancer Center, New York, NY 10065, strongv@mskcc.org, (212) 639-5056.

*Contributed equally

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The authors report no financial conflicts of interests related to this work.

Portions of this work were presented at the 12th Annual Academic Surgical Congress, February 7th – 9th, 2017, Las Vegas NV

Introduction:

Modern surgical care requires the accurate tracking of post-operative complications. Not only must a surgeon provide informed consent to their patients detailing the risks of a surgical procedure, but they must also continually strive to improve their care, both of which require a detailed knowledge of one's own surgical outcomes.

Since 2001 the Department of Surgery at Memorial Sloan Kettering Cancer Center (MSKCC) has recorded post-operative complications on every surgical patient. Data entry in to our surgical secondary events (SSE) database occurs at the point of care and at service-specific morbidity and mortality conferences. The SSE database is known to accurately identify the presence and severity of most post-operative complications¹. MSKCC also participates in the American College of Surgeons National Surgical Quality Improvement Project (NSQIP), which tracks post-operative complications on a 10% sampling of patients who undergo operations performed at participating hospitals. We have compared events captured in our internal SSE database to those captured by NSQIP and found that, when accounting for definition differences, the two databases have strong concordance². Our 2009 audit of the SSE database¹, as well as other research³, has shown that while the SSE database does capture most complications, it is most likely to miss low-grade complications.

Natural language processing is a computer science technique that uses computers to extract research data from unstructured narrative text. Since initial reports of using NLP to identify clinical complications^{4,5} numerous reports have highlighted NLP's potential to aid in interpretation of narrative medical documentation.^{6–19} With the national switch to electronic clinical documentation physicians and researchers are searching for ways to search narrative medical documentation on a large scale as individual chart review is a laborious process. Additionally, increasing use of cumbersome electronic medical records and the subsequent reliance on physicians and other healthcare providers to input structured data into the EMR is a known contributor to the rising tide of physician burnout and dissatisfaction²⁰. NLP has been held up as a potential solution to these problems.

As part of a larger institutional quality improvement project on deep venous thrombosis (DVT) and pulmonary embolism (PE) (collectively: venous thromboembolism [VTE])²¹, we developed an NLP algorithm to identify the presence of post-operative VTE, and graded their severity using other institutional data. This project served as a test case to investigate whether or not NLP could be used to augment the post-operative complications surveillance efforts at our institution. We hypothesized that NLP would be able to accurately identify the presence of a post-operative VTE and could be combined with a rule-based algorithm for classifying the severity of the VTE and as a result augment clinician reporting of complications in our SSE database without requiring an extra layer of manual clinician chart review.

Methods:

Data capture for both the SSE database^{1,3} and NSQIP^{22–24} have been described elsewhere. Complications in the SSE database are added at the point of care and at service-specific

morbidity and mortality conferences by residents, fellows, attending surgeons, and research staff at MSKCC. Complications are graded on our 1–5 modification of the Clavien-Dindo classification²⁵ (Table 1), with grades 1 and 2 considered minor complications and grades 3 – 5 considered major complications. NSQIP selects a subset of institutional operations on an eight-day rolling cycle that are then reviewed by a trained surgical case reviewer to identify a number of post-operative complications occurring within 30 days of operation using standardized definitions. At our institution, we additionally cross check between NSQIP and the SSE to ensure that, allowing for definitional differences, both databases are capturing the same complications. NSQIP also tracks the pre-operative presence of several comorbid conditions, including DVT and PE. For the purposes of this study, patients with a known pre-operative DVT or PE who were on an active anticoagulation regimen were excluded, as were cephalic and portal vein thrombosis due to differing definitions between the SSE database and NSQIP.

Following waiver of informed consent from the MSKCC Institutional Review Board, we identified all patients in our 2011 – 2014 NSQIP cohorts who underwent either a computerized tomography angiography (CTA) scan of the chest or a lower extremity duplex ultrasound. These studies were chosen because they are the most frequently utilized studies when there is clinical concern for DVT (in the case of ultrasound) or PE (in the case of CTA).

Radiology reports were transferred from the institution's radiology information system (RIS) to a Microsoft Structured Query Language (SQL) database. Since critical results are contained within the "Impression" section of the radiology report, this section was parsed for analyses.

In total, there were 909 ultrasound reports for 755 patients and 1,837 CT angiogram report for 1,451 patients performed within 30 days of surgery. These were divided into training (70%) and validation sets (30%). The training set was used to train and adjust the classifier while the test set evaluated performance of the classifier (Figure 1), and training set reports were reviewed to ensure the accuracy of both the classifier as well as the graded severity of the VTE.

Analyses were performed using the WEKA machine learning toolkit²⁶. WEKA is an open source software consisting of a collection of machine learning algorithms with tools for data preprocessing, classification, regression, clustering, association rules and visualization. To prepare the dataset for input into the classifiers, the strings of text from radiology reports were converted to numeric data, or document vectors. Documents were represented by rows, sentences were split into tokens and each token represented as a column. For each document and word column, there is a number indicating the absence or presence of that word in the document (0,1) or frequency of that word in the document (Figure 2). Using the training data, each of the methods above were altered and performance subsequently assessed on each iteration for each classifier. For example, N grams were set to one, stop words were included and performance assessed. In another iteration n grams were set to two, stop words were included and performance assessed. Once it was determined which NLP settings performed the best when inputting the preprocessed text into each classifier, the same

methods were applied to the test set, effectively using the model based on training data. Discrepancies between the NLP identification and our NSQIP gold standard were then manually reviewed in a blinded fashion by one of two surgeons (LS and AR).

To classify the severity of a DVT or PE we then queried pharmacy records as well as interventional procedures for all patients with an NLP verified VTE. Prior to examining individual patient records we assigned a grade to all possible medications and interventions used in the treatment of VTE based on the pre-existing DVT and PE definitions in our SSE database (Table 1), and within the training set records were reviewed to ensure grading was appropriate. Prophylactic doses of unfractionated or low molecular weight heparin given prior to VTE diagnosis were excluded for the purposes of grade calculation. Patients with multiple grades (placed on IV unfractionated heparin and later oral warfarin) were assigned the highest grade. All patients in both datasets who underwent operative or radiologic intervention following detection of a VTE were hand reviewed to determine whether or not the intervention was related to their VTE.

Results:

NLP identification of DVT / PE

In total, our 2011 – 2014 NSQIP cohort contained 10,295 patients. Of these, 1,895 had an imaging study for inclusion into either the training or validation datasets. Of the 703 patients in the validation dataset, our NLP algorithm identified 51 DVTs and 33 PEs, compared to 47 DVTs and 30 PEs identified by NSQIP (Table 2a). Using NSQIP as our gold standard, the sensitivity and specificity of our NLP algorithm for the identification of DVT diagnosed in radiology imaging at our institution is 85.1% and 94.6%, respectively, with a positive predictive value and negative predictive value of 78.4% and 96.5%, respectively. For identification of PEs, the sensitivity and specificity is 90.0% and 98.7% respectively, with a positive and negative predictive value of 81.8% and 99.3%, respectively.

Discrepancies between NLP algorithm and NSQIP

Discrepancies between our NLP algorithm and our NSQIP records were hand reviewed (Table 2b and 3). Most false negative cases detected by the NLP algorithm were due to either diagnosis of a VTE by a non-queried imaging modality (for example, diagnosing an incidental PE of the lower lung on a CT of the abdomen / pelvis), diagnosis of a VTE at an outside institution, identification of a chronic, and stable, venous thrombosis that was present pre-operatively, or no documentation of a VTE in the patient record. The NLP process identified three cases that were determined not to be DVTs after review (Table 3).

VTE Grading

The majority of VTEs at our institution are grade two according to our modification of the Clavien-Dindo classification (Figure 3), though this project identified a higher than expected rate of grade 3 VTEs. These were most commonly placement of an IVC filter in a patient who could not receive therapeutic anticoagulation post-operatively due to the extent of their operations. There were no deaths (grade 5 complications) attributable to a post-operative VTE.

Discussion:

We designed a natural language processing (NLP) algorithm that can accurately identify post-operative VTEs and was combined with an automated rule-based algorithm for determining the severity of the VTE. When combined with other electronic data at our institution, this algorithm can accurately score the severity of the VTE. Not surprisingly, the majority of VTEs detected by the NLP algorithm are grade 2 complications on our modification of the Clavien-Dindo classification (Figure 2). We began this project to improve the recording of VTE in our SSE database as well as test, more broadly, whether or not NLP could be used to extract information from our clinical documentation, thus freeing clinical and research staff from manual chart review¹⁹. Previous work^{1,3,27} has shown that if a post-operative complication occurs after discharge or is grade 1 or 2 it is less likely to be entered into our database even though they were identified and appropriately treated at the time. The gap we identified and attempted to narrow using our NLP algorithm was not in the clinical identification of post-operative VTE, but rather in adding these VTE to our internal SSE database.

Like many institutions, the transition to the electronic medical record has forced us to find ways to best leverage the available computing power to ease the clinical and research documentation required of health care providers. By using NLP to screen all radiology reports we sought to eliminate the need for structured data entry and manual chart review by clinical and research staff for detection of some easily identifiable post-operative complications, a task for which a computer is well suited. We have shown that NLP can be used to accurately classify our clinical documentation, and believe that its uses far exceed simply searching radiology documentation for the presence of post-operative VTEs.

The negative predictive value of our NLP algorithm was superior to the positive predictive value. However, due to the setup of our SSE database, which tracks the presence of over 200 post-operative complications, not their absence, we have highlighted the NLP algorithm's ability to rule in the presence of a VTE, not rule in the absence of one. In fact, our algorithm is well suited for both.

We did not compare the NSQIP or NLP algorithms VTE detection to the SSE database because the purpose of this project was not to prove superiority of one detection model over another, but rather to develop an NLP tool to augment data entry into the SSE database for all surgical patients at our institution. The hope was that an additional detection tool would allow for maximal utilization of existing resources and allow our staff to focus on data analysis and practice change rather than data entry, thus increasing their productivity and hopefully slightly decreasing burnout²⁰. This report is part of a larger process to leverage NLP to improve our database's capture of multiple post-operative complications and other clinical conditions. Further work is planned to develop algorithms to identify urinary tract infections and wound infections, among others. We are currently finalizing a process to screen radiology reports for all surgical patients in real time, the final step in implementing our algorithm institution wide.

Our report has several limitations. Because of definitional differences between NSQIP and our SSE database we excluded cephalic and portal vein thrombosis. These thromboses are classified as DVTs by NSQIP but our SSE database records them separately and they were excluded in order to simplify the development of the algorithm. Our algorithm was also additionally too narrow. We restricted it to duplex ultrasound of the extremity and CTA of the chest in order to simplify the development work, but as a result “missed” several DVTs identified on various cross-sectional imaging modalities. We have since expanded the algorithm to include all radiology reports, which should, in practice, improve its positive predictive value.

We have demonstrated the successful development of a NLP algorithm that accurately identifies post-operative VTE in radiology reports and is combined with an automated rule-based scoring system to grade the severity of the VTE. Our initial study examined only duplex ultrasound and CTA of the chest, but we have expanded our algorithm to include all imaging studies in an effort to reduce the generation of false-negative events. While labor intensive to generate, this algorithm, which combined NLP based detection of VTE and an automated determination of the VTE’s severity, allows our clinical and research staff to shift their focus away from data entry, and also has the potential to expedite the work of our NSQIP surgical case reviewers. The algorithm was designed for specific language used in radiology reports at our institution, but NLP as a tool for searching unstructured clinical documentation is easily applicable to a different institution’s specific reporting conventions. Natural language processing can be used to accurately identify the presence of post-operative complications in narrative clinical documentation and, when combined with other freely available clinical data, can accurately grade the severity of these complications.

Funding Sources:

Memorial Sloan Kettering Cancer Center receives institutional support from NIH P30 CA008748.

Abbreviations:

NLP	Natural Language Processing
VTE	Venous Thromboembolism
DVT	Deep Venous Thrombosis
PE	Pulmonary Embolism
CT	Computerized Tomography
NSQIP	National Surgical Quality Improvement Project

References:

1. Strong VE, Selby LV, Sovel M, et al.: Development and assessment of Memorial Sloan Kettering Cancer Center’s Surgical Secondary Events grading system. *Ann Surg Oncol* 22:1061–7, 2015 [PubMed: 25319579]

2. Selby LV, Sjoberg DD, Cassella D, et al.: Comparing surgical infections in National Surgical Quality Improvement Project and an Institutional Database. *J Surg Res* 196:416–20, 2015 [PubMed: 25840487]
3. Grobmyer SR, Pieracci FM, Allen PJ, et al.: Defining morbidity after pancreaticoduodenectomy: use of a prospective complication grading system. *J Am Coll Surg* 204:356–64, 2007 [PubMed: 17324768]
4. Melton GB, Hripcsak G: Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc* 12:448–57, 2005 [PubMed: 15802475]
5. Murff HJ, FitzHenry F, Matheny ME, et al.: Automated Identification of Postoperative Complications Within an Electronic Medical Record Using Natural Language Processing. *Jama-Journal of the American Medical Association* 306:848–855, 2011
6. Rochefort CM, Verma AD, Eguale T, et al.: A novel method of adverse event detection can accurately identify venous thromboembolisms (VTEs) from narrative electronic health record data. *J Am Med Inform Assoc* 22:155–65, 2015 [PubMed: 25332356]
7. Masino AJ, Grundmeier RW, Pennington JW, et al.: Temporal bone radiology report classification using open source machine learning and natural language processing libraries. *BMC Med Inform Decis Mak* 16:65, 2016 [PubMed: 27267768]
8. Ho B, Skaro A, Montag S, et al.: Elementary, My Dear Watson-The Era of Natural Language Processing in Transplantation. *Am J Transplant*, 2016
9. Fritz BA, Escallier KE, Ben Abdallah A, et al.: Convergent Validity of Three Methods for Measuring Postoperative Complications. *Anesthesiology* 124:1265–76, 2016 [PubMed: 27028469]
10. Tien M, Kashyap R, Wilson GA, et al.: Retrospective Derivation and Validation of an Automated Electronic Search Algorithm to Identify Post Operative Cardiovascular and Thromboembolic Complications. *Appl Clin Inform* 6:565–76, 2015 [PubMed: 26448798]
11. Sevenster M, Buurman J, Liu P, et al.: Natural Language Processing Techniques for Extracting and Categorizing Finding Measurements in Narrative Radiology Reports. *Appl Clin Inform* 6:600–110, 2015 [PubMed: 26448801]
12. Liao KP, Cai T, Savova GK, et al.: Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *Bmj* 350:h1885, 2015 [PubMed: 25911572]
13. FitzHenry F, Murff HJ, Matheny ME, et al.: Exploring the frontier of electronic health record surveillance: the case of postoperative complications. *Med Care* 51:509–16, 2013 [PubMed: 23673394]
14. Carrell DS, Halgrim S, Tran DT, et al.: Using natural language processing to improve efficiency of manual chart abstraction in research: the case of breast cancer recurrence. *Am J Epidemiol* 179:749–58, 2014 [PubMed: 24488511]
15. Liu L, Shorstein NH, Amsden LB, et al.: Natural language processing to ascertain two key variables from operative reports in ophthalmology. *Pharmacoepidemiol Drug Saf*, 2017
16. Sohn S, Larson DW, Habermann EB, et al.: Detection of clinically important colorectal surgical site infection using Bayesian network. *J Surg Res* 209:168–173, 2016 [PubMed: 28032554]
17. Patterson OV, Forbush TB, Saini SD, et al.: Classifying the Indication for Colonoscopy Procedures: A Comparison of NLP Approaches in a Diverse National Healthcare System. *Stud Health Technol Inform* 216:614–8, 2015 [PubMed: 26262124]
18. Anderson JE, Chang DC: Using electronic health records for surgical quality improvement in the era of big data. *JAMA Surg* 150:24–9, 2015 [PubMed: 25372451]
19. Deliberato RO, Celi LA, Stone DJ: Clinical Note Creation, Binning, and Artificial Intelligence. *JMIR Med Inform* 5:e24, 2017 [PubMed: 28778845]
20. Robertson SL, Robinson MD, Reid A: Electronic Health Record Effects on Work-Life Balance and Burnout Within the I3 Population Collaborative. *Journal of Graduate Medical Education* 9:479–484, 2017 [PubMed: 28824762]
21. Selby LV, Sovel M, Sjoberg DD, et al.: Preoperative Chemoprophylaxis Is Safe in Major Oncology Operations and Effective at Preventing Venous Thromboembolism. *J Am Coll Surg* 222:129–37, 2016 [PubMed: 26711793]

22. 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. *Ann Surg* 228:491–507, 1998 [PubMed: 9790339]
23. Fink AS, Campbell DA Jr., Mentzer RM Jr., et al.: The National Surgical Quality Improvement Program in non-veterans administration hospitals: initial demonstration of feasibility. *Ann Surg* 236:344–53; discussion 353–4, 2002 [PubMed: 12192321]
24. Khuri SF, Henderson WG, Daley J, et al.: Successful implementation of the Department of Veterans Affairs' National Surgical Quality Improvement Program in the private sector: the Patient Safety in Surgery study. *Ann Surg* 248:329–36, 2008 [PubMed: 18650645]
25. Clavien PA, Barkun J, de Oliveira ML, et al.: The Clavien-Dindo classification of surgical complications: five-year experience. *Ann Surg* 250:187–96, 2009 [PubMed: 19638912]
26. Witten IH, Frank E, Hall MA, Books24x7 Inc. Data mining practical machine learning tools and techniques, third edition [Text.] Burlington, Mass: Morgan Kaufmann Publishers; 2011
27. Selby LV, Vertosick EA, Sjoberg DD, et al.: Morbidity after Total Gastrectomy: Analysis of 238 Patients. *J Am Coll Surg* 220:863–871 e2, 2015 [PubMed: 25842172]

Article Summary:

We used natural language processing (NLP) to identify and grade post-operative DVT and PE. This report demonstrates that NLP can be used to allow clinicians involved in quality improvement to focus on tasks beyond manual chart review.

Figure 1:
Experimental design for derivation of training and validation datasets in 2011 – 2014 NSQIP cohort.

Document	Impression Text
1	Deep venous thrombosis in bilateral calf veins, unchanged.
2	Deep vein thrombosis in the bilateral calf veins.
3	Deep venous thrombosis in left subclavian and axillary veins.
4	Deep venous thrombosis in right deep calf vein at the level of the right mid calf. Left mid calf probable hematoma.
5	No evidence of deep venous thrombosis in left upper extremity veins
6	No evidence of deep venous thrombosis in bilateral femoral popliteal system.

No.	@@class@@ Nominal	bilater Numeric	deep Numeric	evid Numeric	extrem Numeric	femor Numeric	in Numeric	left Numeric	no Numeric	of Numeric	poplite Numeric	system Numeric	thrombos Numeric	upper Numeric	vein Numeric
1	POS	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0
2	POS	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	2.0
3	POS	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0
4	POS	0.0	2.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0
5	NEG	0.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0
6	NEG	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0

No.	@@class@@ Nominal	ven Numeric	and Numeric	at Numeric	axil Numeric	calf Numeric	hematom Numeric	level Numeric	mid Numeric	prob Numeric	right Numeric	subclav Numeric	th Numeric	unchang Numeric
1	POS	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
2	POS	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
3	POS	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
4	POS	1.0	0.0	1.0	0.0	3.0	1.0	1.0	2.0	1.0	2.0	0.0	2.0	0.0
5	NEG	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	NEG	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Figure 2:
Preprocessing of documents from string to word vectors.

Figure 3:

Distribution of VTE Cases Based on Clavien-Dindo Classification * N/A: 5 patients had prophylactic IVC filters placed pre-operative due to a history of prior DVT / PE and then developed a post-operative VTE but required no further treatment because of the prophylactic filter.

Table 1:

Complication Grade for DVT or PE

Grade	Definition	Medication / Intervention
1	Bedside care or oral medication	Warfarin or novel anticoagulant
2	Intravenous medication	Unfractionated Heparin drip or Subcutaneous low molecular weight heparin
3	Radiologic, endoscopic, or operative intervention	Thrombectomy or placement of IVC filter
4	Prolonged disability	Tracheostomy for prolonged respiratory failure related to VTE
5	Death	Death related to VTE

Table 2:

2×2 table for NLP validation. Table 2a shows, separately, 2×2 tables for DVT and PE. Table 2b shows a breakdown of the areas of discordance between our NLP algorithm and NSQIP.

Table 2a: DVT and PE specific 2×2 tables									
DVT					PE				
		NSQIP					NSQIP		
		+DVT	−DVT	TOTAL			+PE	−PE	TOTAL
NLP	+ DVT	40	11	51	NLP	+ PE	27	6	33
	−DVT	7	193	200		−PE	3	470	473
	TOTAL	47	204	251		TOTAL	30	476	506

Table 2b: Breakdown of discrepancies between NLP and NSQIP

VTE Identified by NSQIP not NLP			
		DVT	PE
Total		7	3
NLP correct based on chart review		5	0
NLP incorrect based on chart review		2	3
VTE Identified by NLP not NSQIP			
		DVT	PE
Total		11	6
NLP correct based on chart review		9	5
NLP incorrect based on chart review		2	1

Table 3:

Root cause of false negative results for NLP algorithm in the validation dataset.

DVT Identified by NSQIP but not NLP	N=7	PE Identified by NSQIP but not NLP	N=3
Reasons for Discrepancy		Reasons for Discrepancy	
No indication in EMR	5	Algorithm false negative	3
Report indicates the presence and absence of a DVT for different sites	1		
Algorithm false negative	1		