

Retrospective Derivation and Validation of an Automated Electronic Search Algorithm to Identify Postoperative Cardiovascular and Thromboembolic Complications

M. Tien¹; R. Kashyap²; G. A. Wilson³; V. Hernandez-Torres²; A. K. Jacob²; D. R. Schroeder⁴; C. B. Mantilla²

¹Mayo Clinic, College of Medicine, Rochester, MN, United States; ²Mayo Clinic, Department of Anesthesiology, Rochester, MN, United States; ³Mayo Clinic, Division of Pulmonary and Critical Care Medicine, Rochester, MN, United States; ⁴Mayo Clinic, Health Sciences Research - Biomedical Statistics and Informatics, Rochester, MN, United States

Keywords

Clinical research informatics, search algorithm, myocardial infarction, venous thromboembolism

Summary

Background: With increasing numbers of hospitals adopting electronic medical records, electronic search algorithms for identifying postoperative complications can be invaluable tools to expedite data abstraction and clinical research to improve patient outcomes.

Objectives: To derive and validate an electronic search algorithm to identify postoperative thromboembolic and cardiovascular complications such as deep venous thrombosis, pulmonary embolism, or myocardial infarction within 30 days of total hip or knee arthroplasty.

Methods: A total of 34 517 patients undergoing total hip or knee arthroplasty between January 1, 1996 and December 31, 2013 were identified. Using a derivation cohort of 418 patients, several iterations of a free-text electronic search were developed and refined for each complication. Subsequently, the automated search algorithm was validated on an independent cohort of 2 857 patients, and the sensitivity and specificities were compared to the results of manual chart review.

Results: In the final derivation subset, the automated search algorithm achieved a sensitivity of 91% and specificity of 85% for deep vein thrombosis, a sensitivity of 96% and specificity of 100% for pulmonary embolism, and a sensitivity of 100% and specificity of 95% for myocardial infarction. When applied to the validation cohort, the search algorithm achieved a sensitivity of 97% and specificity of 99% for deep vein thrombosis, a sensitivity of 97% and specificity of 100% for pulmonary embolism, and a sensitivity of 100% and specificity of 99% for myocardial infarction.

Conclusions: The derivation and validation of an electronic search strategy can accelerate the data abstraction process for research, quality improvement, and enhancement of patient care, while maintaining superb reliability compared to manual review.

Correspondence to:

Carlos B. Mantilla, M.D., Ph.D.
Department of Anesthesiology
Mayo Clinic
200 First Street SW
Rochester, MN 55905
Tel.: (507) 284-7461
Fax: (507) 255-7300
Email: mantilla.carlos@mayo.edu

Appl Clin Inform 2015; 6: 565-576

<http://dx.doi.org/10.4338/ACI-2015-03-RA-0026>

received: March 10, 2015

accepted in revised form: July 28, 2015

published: September 9, 2015

Citation: Tien M, Kashyap R, Wilson GA, Hernandez-Torres V, Jacob AK, Schroeder DR, Mantilla CB. Retrospective Derivation and Validation of an Automated Electronic Search Algorithm to Identify Postoperative Thromboembolic and Cardiovascular Complications. *Appl Clin Inform* 2015; 6: 565-576
<http://dx.doi.org/10.4338/ACI-2015-03-RA-0026>

1. Introduction

Over the course of the last decade, there has been an increase in the use of electronic medical records (EMRs), and the amount of medical information available for epidemiologic and clinical research has rapidly expanded [1]. This poses new obstacles in contemporary research methodology, such as the inability to manually review sufficient amounts of data in a reasonable time period, the use of inadequate search strategies to review the EMR, and the reliance on the accuracy of ICD-9 (International Classification of Diseases, Ninth Revision) billing codes [2, 3].

Recently, there has been an increase in the use of automated electronic search strategies to facilitate data collection. For instance, search algorithms have been successfully developed for extracting information on Charlson comorbidities [2], initiation of emergent intubations in the intensive care unit (ICU) [4], extubation time in the ICU [5], risk factors for acute lung injury [6], and chronic co-morbidity phenotypes from the EMR and Genomics studies [7]. These studies have all demonstrated that electronic searches can achieve sensitivities and specificities greater than 90% when compared to manual search efforts. Additionally, a previous study demonstrated portability of such electronic search tools, potentially allowing for application of search algorithms at external institutions [8]. However, there is limited available literature on the derivation and validation of an electronic search technique for identifying complications in the postoperative setting and the effectiveness of such methodology in this realm, specifically when compared to manual chart review of a prospectively-collected electronic database.

It is well known that critical complications following elective surgeries include thromboembolic and cardiovascular complications such as deep venous thrombosis, pulmonary embolism, and myocardial infarction. Total hip and knee arthroplasties are among the most commonly performed elective surgical procedures [9], and these patients are inherently at a higher risk of these complications because of their demographics, the nature of the operation, and the preexisting comorbidities commonly found in these patients [10–12]. Despite the frequency of deep venous thrombosis, pulmonary embolism, and myocardial infarction following lower extremity arthroplasty, there is no established gold standard available for efficient identification of these important and clinically-relevant events from patient EMRs. This patient population is ideally suited for the derivation and validation of an electronic search algorithm for postoperative thromboembolic and cardiovascular complications.

2. Objectives

The primary objective of this study was to derive and validate an automated electronic search algorithm for identifying the occurrence of postoperative deep venous thrombosis, pulmonary embolism, or myocardial infarction within 30 days of total hip or knee arthroplasty. The secondary objective was to calculate the sensitivity and specificity values of our electronic search algorithm when compared with manual comprehensive review of the patient records (the reference standard) for identifying these thromboembolic and cardiovascular complications within 30 days of surgery. Positive or negative predictive values were not calculated because of their dependence on the incidence rate within the cohort. The automated electronic search algorithm was designed with the goal of expediting clinical research and quality improvement endeavors within an institution, and was not intended for surveillance purposes.

3. Methods

3.1 Identification of Study Population

Mayo Clinic Institutional Review Board approval was obtained for retrospective analysis of patients undergoing total hip or knee arthroplasty surgeries between January 1, 1996 and December 31, 2013. These patient records were available via the Total Joint Registry (TJR) database at Mayo Clinic in Rochester, Minnesota, which is a prospectively collected electronic database containing compre-

hensive information for all joint arthroplasty procedures performed at this institution dating back to 1969 [13]. This database was used to identify a heterogeneous group of patients who underwent total hip or knee arthroplasty during the study period. In patients who may have undergone multiple surgeries during the period of study, review was restricted to the first surgery, and thus only one operative and perioperative course was analyzed for each patient. In accordance with the Minnesota Health Records Act (Minnesota Statute 144.291–144.298), only patients who gave research consent were included in the study. After identification of all patients undergoing their first total hip or knee arthroplasty in the study period and applying exclusion criteria (no research authorization, patients younger than 18 years old, and patients not included in specific Mayo Clinic electronic databases discussed below), two independent subsets of patients were randomly selected to comprise the derivation cohort and the validation cohort (►Figure 1).

3.2 Manual Data Extraction Strategy

Manual review of patient records is the traditional method used for data extraction, as there is no established gold standard available for identification of thromboembolic and cardiovascular complications from the EMR. In this study, two study authors independently reviewed the EMR for patients in the derivation and validation subsets to look for the complications of interest occurring within 30 days of the orthopedic operation. These authors were not involved in the development or utilization of the automated electronic search strategy and were not aware of the results from the automated electronic note search strategy.

Using previously defined criteria for establishing a diagnosis of deep venous thrombosis, pulmonary embolism, and myocardial infarction, the reviewers manually categorized each complication into the highest level of diagnostic certainty (definite, probable, possible, or no event) [10]. Following previously established categorization schemes, patients experiencing complications classified as either definite or probable were combined to represent patients experiencing the complication of interest [10]. For patients experiencing multiple complications of interest or multiple occurrences of complications of interest within the 30-day postoperative period, each manifestation was classified into the highest level of diagnostic confidence. Additionally, it was assumed that patients experiencing pulmonary embolism also had a probable deep venous thrombosis.

3.3 Automated Electronic Data Extraction Strategy

In addition to the TJR database used to identify patients eligible for participation in this study, data was extracted from the Mayo Clinic Intensive Care Unit/Operating Room DataMart, the Mayo Clinic Life Sciences System (MCLSS), and the Anesthesia Quality Database, all of which are comprehensive clinical data warehouses that store information regarding patient demographics and characteristics, operative course, and perioperative management and complication details [14, 15]. A commercially available, web-based query-building tool called Data Discovery and Query Builder (DDQB) was utilized for interrogation of data files contained within the databases listed above (International Business Machines, Corp). The DDQB tool allows clinicians and researchers without extensive programming knowledge to build complex queries based on Boolean logic and free-text searches [4, 6]. These tools were used to search through patient records in the EMR to identify deep venous thrombosis, pulmonary embolism, or myocardial infarction within 30 days of the lower extremity orthopedic procedure.

To develop the electronic search query, all synonyms, abbreviations, medical acronyms, and most common symptoms associated with each complication were entered. Furthermore, a comprehensive list of terms to exclude was developed to make the electronic search algorithm more specific. For instance, words and phrases such as “prior,” “rule out,” or “negative for” were excluded (complete list of inclusion and exclusion terms in ►Appendices 1–3). To establish a more uniform methodology and minimize the number of false positive results, the application of the automated algorithm to note searches was restricted to the Chief Complaint, Diagnosis, and Impression/Report/Plan sections and laboratory results in the patient records. The results from other imaging or diagnostic modalities such as electrocardiography, ventilation-perfusion scans, or computed tomography scans

were only available for abstraction from documentation in note sections in the medical record, as an electronic search was not developed for these modalities.

The automated electronic search algorithm was continuously refined through an iterative process consisting of review of mismatches between the automated electronic search and the manual chart review. Every time any discordant results were identified between the electronic search and the manual search, the search algorithm inclusion and exclusion terms and criteria were updated and re-tested on the derivation cohort. Through various improvements of search and exclusion terms, the sensitivity and specificity for each complication improved to greater than 85%. After several iterations of this process, the last search terms and algorithm used to construct the automated electronic search query were finalized and applied to the validation cohort (►Figure 1). To achieve this validation step, sensitivity and specificity values were calculated through comparison of the applied automated electronic search algorithm to the validation subset versus a comprehensive manual review of the same subset serving as the reference standard. The two study authors who designed and implemented the electronic search algorithm using DDQB did not participate in the extraction of data via manual chart review.

3.4 Statistical Analyses

In this study, point estimates for sensitivity and specificity of the automated electronic search strategy were calculated by comparing the results of the electronic data abstraction method to the results of manual data abstraction method (reference standard) for both the derivation and validation cohorts using JMP version 9.0 (Statistical Analysis System Institute, Inc.). Because the electronic search algorithm was derived over the course of several iterations, the sensitivity and specificity of the search as applied to the derivation cohort was represented by the final iteration of the electronic search algorithm refinement process.

4. Results

After all exclusions were applied, the total final retrospective patient pool included 34517 first lower extremity orthopedic surgeries between January 1, 1996 and December 31, 2013 at Mayo Clinic. Both primary and revision knee and hip arthroplasties were included, as well as both unilateral and bilateral procedures. A subset of 418 random patients was selected as the derivation cohort and another independent subset of 2857 random patients was selected as the validation cohort. Both of these cohorts consisted of patients who experienced deep venous thrombosis, pulmonary embolism, or myocardial infarction, and also patients who did not experience any of these complications within 30 days of total hip or knee arthroplasty.

Initially, the automated electronic search strategy achieved a sensitivity of 70% and specificity of 68% for deep venous thrombosis, a sensitivity of 93% and specificity of 76% for pulmonary embolism, and a sensitivity of 92% and specificity of 68% for myocardial infarction in the derivation cohort when analyzed against a manual review of these patient records. After several revisions were made to the search algorithm, the final iteration of the electronic search algorithm yielded a sensitivity of 91% and specificity of 85% for deep venous thrombosis, sensitivity of 96% and specificity of 100% for pulmonary embolism, and sensitivity of 100% and specificity of 95% for myocardial infarction when applied to the derivation subset. The finalized algorithm and search inclusion and exclusion terms for each complication of interest are provided as Appendices 1–3.

When the finalized automated search algorithm was applied to the validation cohort of 2857 patients, the electronic strategy achieved 97% sensitivity and 99% specificity for deep venous thrombosis, 97% sensitivity and 100% specificity for pulmonary embolism, and 100% sensitivity and 99% specificity for myocardial infarction. These results are summarized in ►Table 1. Additionally, ►Tables 2–4 provide the initial sensitivity and specificity values for the electronic search of each complication of interest, followed by the final sensitivity and specificity values of the derivation cohort after several iterations of updates, and the sensitivity and specificity values when applied to the validation cohort.

5. Discussion

The present study compared two independent methodologies for extraction of data pertaining to the occurrence of thromboembolic and cardiovascular complications within 30 days following lower extremity orthopedic procedures. The results reported in this study support the notion that the development and validation of an automated electronic search algorithm to interrogate the EMR for identifying certain elements pertaining to patient care is a reliable alternative to the reference standard of manual review. Comparison of the two methodologies demonstrated that the sensitivity and specificity of the automated search algorithm could approach 100% in many instances. These findings further corroborate previously published studies demonstrating that the use of automated search strategies produced very accurate results that were agreeable with those produced by manual review and data extraction methods [2, 4, 6, 7]. Additionally, this study demonstrates that this methodology can be a reliable alternative to manual chart review for research pertaining to postoperative complications.

It is important to note that a previous study by Murff et al. [16] compared the use of electronic free-text search algorithms applied to EMRs versus administrative ICD-9 data codes to identify postoperative complications. Searching EMRs to identify postoperative complications performed with higher sensitivity values but lower specificity values when compared with equivalent safety indicators found in administrative discharge codes. These investigators reported sensitivity and specificity values in identifying deep venous thrombosis or pulmonary embolism (59% and 91%, respectively for both thromboembolic complications) and myocardial infarction (91% and 95%, respectively) in the postoperative period using natural language processing [16]. The free-text search algorithm developed in the present study found sensitivity and specificity values between 97–100% for cardiopulmonary and thromboembolic complications within 30 days of total hip or knee arthroplasty surgery. Taken together, these studies solidify the advantage of electronically searching EMRs (compared to other electronic search strategies such as ICD-9 data codes), and demonstrate that this method can approximate the accuracy of manual chart review by experts in prospectively-collected electronic databases across institutions.

There are numerous advantages to utilizing the approach detailed in this current study for application to clinical and translational research in the current era of bioinformatics. For instance, the automated electronic search approach allows for an unprecedented reduction in time spent reviewing patient records compared with manual review for such research studies. The electronic algorithmic search strategy is also able to provide uniform and consistent results that are just as reliable and accurate as those extracted by manual review, and is not prone to mental errors made when researchers get fatigued performing exhaustive manual data extraction. Moreover, because the DDQB is a commercial web-based query-building tool, the current approach is applicable to any EMR database system. The search algorithm was developed using free text and natural language strategies that broadly reflect clinical practice (► Appendices 1–3) and do not rely on the availability of additional codified, electronic information for specialized imaging tests or diagnostic modalities. With the recent trend of transitioning from paper medical records to EMR seen in the United States, the strategy employed in this paper can be widely applicable and used in conjunction with any type of standard or customized software for statistical analysis.

Taking potential differences in data structures, semantics, and documentation language across various institutions and departments into consideration, this search strategy incorporates a broad spectrum of inclusion and exclusion terms (► Appendices 1–3) to minimize the chance of type I errors and maximize the chance of capturing close to 100% of events. At our institution, the search strategy performed with high levels of sensitivity and specificity; however, the point estimates cannot be expected to express the true limits for these values in other settings. In this sense, it is important to note that although the derivation and validation strategy could potentially be enhanced by specifically tailoring the algorithm to better suit other institutions, the portability of free-text search algorithms was previously demonstrated across different institutions and EMRs with minimal need for institution-specific algorithm optimization [8]. The efficiency with which clinical and translational research can now be performed is not only of tremendous value to researchers but can also indirectly enhance the quality of patient care by providing information about risk factors for adverse outcomes of interest. For instance, information regarding perioperative thromboembolic and

cardiovascular complications may permit investigation of the impact of changes in medical practice. In the case of this current derivation and validation of a search algorithm to identify patients experiencing thromboembolic and cardiovascular complications following lower extremity orthopedic surgery, for instance, this search strategy can accelerate any clinical research that may involve analyzing postoperative complications, thromboembolic or cardiovascular disease, or studying new anticoagulants.

There are a few important limitations of this study that should be acknowledged. The quality of the data obtained from automated electronic note search is reliant upon the accuracy of the clinical record and integrity of the database from which it is derived. Any incompleteness or unreliable data points in the database or inconsistencies with the text search phrases can lead to inaccurate results and therefore limit the applicability of this data extraction methodology. Data entry into the database may have been incorrect or the information within the database could potentially become corrupted, but this limitation likely accounts for a very small number of patients in the database [17]. Additionally, it is possible for complications to occur following patient discharge. Thus, the completeness of the EMR will depend on the addition of this information if the patient experiences the complication outside of the institution. While the Mayo Clinic TJR database prospectively ascertains the occurrences of these complications with routine follow-up of patients, it is not known whether or not this conclusion can be generalized to databases used at other institutions [13]. While this study did not determine generalizability to other institutions, appropriate measures were taken to include broad free-text, natural language search criteria that are expected to accommodate for variations in medical terminology across institutions. Furthermore, there is prior evidence to the portability of similar electronic search tools in the literature [8].

The current study only searched through the notes classified under Chief Complaint, Diagnosis, Impression/Report/Plan note sections and laboratory results of the patient records. While this is a thorough and comprehensive search of the most relevant sections, it is by no means all-inclusive, and the occurrence of a complication of interest not documented in these notes would have been missed. However, it is anticipated that this comprises a very small proportion of the study cohort, as the electronic search performed with sensitivities and specificities close to 100% when compared to manual review. Of note, the sensitivity and specificity in the validation cohort were higher than in the derivation cohort, perhaps due to differences in the prevalence of the conditions of interest. In this regard, prevalence of deep venous thrombosis was 8–11%, pulmonary embolism was 5–15%, and myocardial infarction was 20–30% in the two cohorts. Lastly, the acquisition of the data from the database via any electronic search strategy is limited by the timing of when the procedure notes are uploaded or when the databases are updated. Therefore, this methodology is not applicable for real-time use of data collection and analytics.

6. Conclusions

This present study details the derivation and validation of an automated electronic search query algorithm for identifying the occurrence of major thromboembolic and cardiovascular complications in the postoperative period. Complications including deep venous thrombosis, pulmonary embolism, and myocardial infarction occurring within 30 days of lower extremity arthroplasty can be identified with great accuracy using an automated electronic search algorithm. The achieved sensitivity and specificity can approach 100% through continuous improvements in the electronic note search via the iterative process described above. Additionally, the development of this electronic search strategy can be widely applicable to help improve the efficiency and accuracy of clinical research, aid in institutional assessment of surgical and perioperative outcomes, direct quality improvement projects, and ultimately guide physician decision-making to enhance patient recovery.

Clinical Relevance Statement

With the movement towards universal implementation of electronic medical records, automated electronic note search algorithms can serve as effective tools in clinical research for helping investigators accurately and reliably extract data from patient records. This study demonstrates that a free-text search algorithm devised for identifying deep venous thrombosis, pulmonary embolism, or

myocardial infarction within 30 days of total hip or knee arthroplasty was able to successfully perform the search with high degree of sensitivity and specificity. This search strategy can be broadly used with various types of software, may be generalized to other postoperative complications and procedure types, and serves as a highly reliable substitute for manual data extraction. It has the potential to help expedite clinical research and quality improvement projects, thus facilitating the delivery of higher quality patient care.

Conflicts of Interest

The authors have no conflicts of interest.

Financial Support and Disclosure

The Department of Anesthesiology and Division of Pulmonary and Critical Care Medicine at Mayo Clinic, Rochester, Minnesota supported this work with no direct financial support.

Human Subjects Protections

The Institutional Review Board at Mayo Clinic approved the use of existing electronic medical records of patients giving research consent.

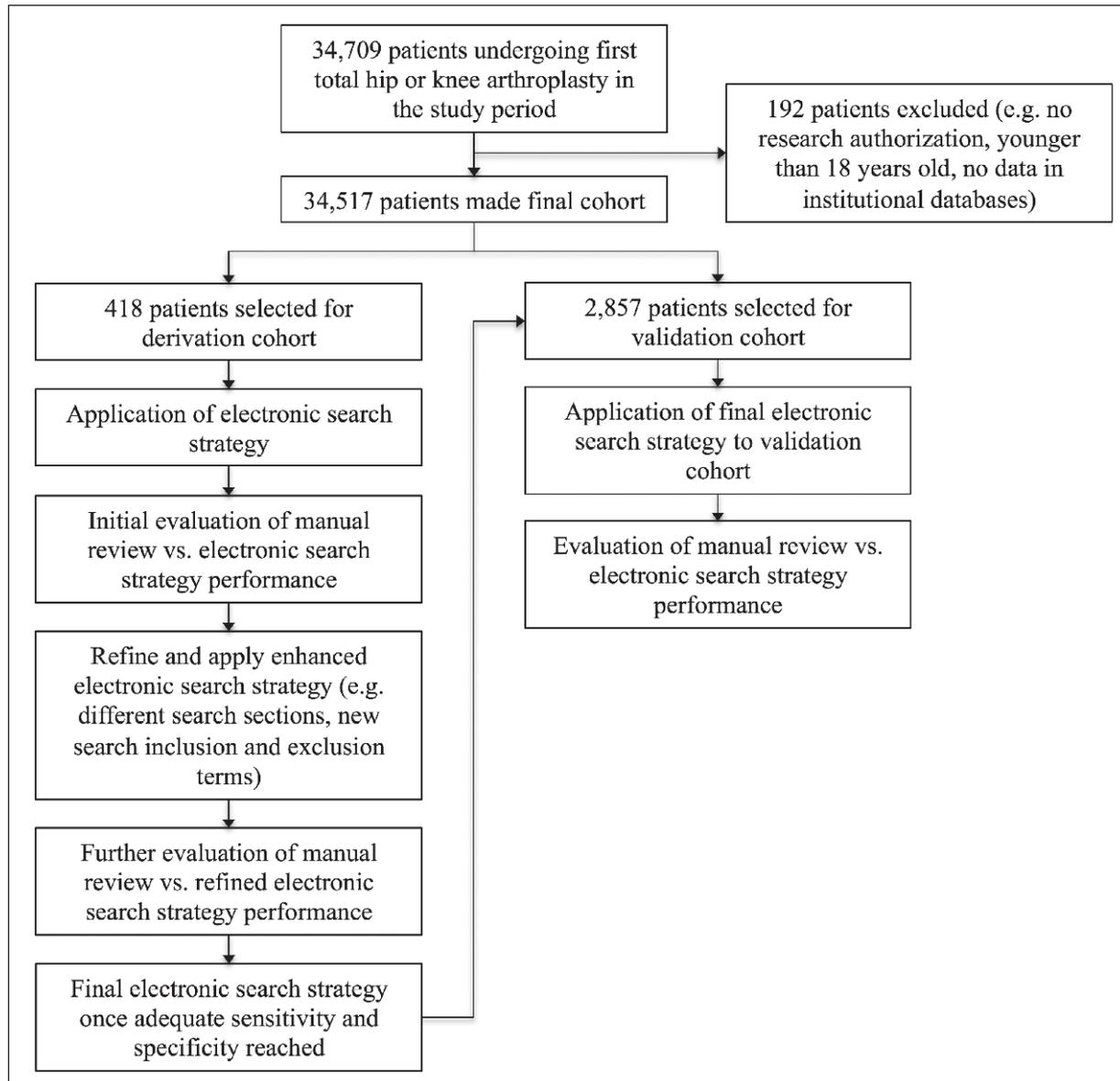


Fig. 1 Selection of patients to the derivation cohort and the validation cohort.

This document was downloaded for personal use only. Unauthorized distribution is strictly prohibited.

Table 1 Derivation cohort vs. validation cohort sensitivity and specificity values

	Derivation Cohort		Validation Cohort	
	Sensitivity	Specificity	Sensitivity	Specificity
Deep venous thrombosis	91%	85%	97%	99%
Pulmonary embolism	96%	100%	97%	100%
Myocardial infarction	100%	95%	100%	99%

Table 2 Derivation and validation of electronic search algorithm for deep venous thrombosis

	Search Algorithm	Sensitivity	Specificity
Initial Derivation	Manual review vs. identification of initial search terms in "diagnosis" section	70%	68%
Final Derivation	Manual review vs. identification of final search terms in "chief complaint" section	91%	85%
Validation	Manual review vs. identification of final search terms in "chief complaint" section	97%	99%

Table 3 Derivation and validation of electronic search algorithm for pulmonary embolism

	Search Algorithm	Sensitivity	Specificity
Initial Derivation	Manual review vs. identification of initial search terms in "diagnosis" section	93%	76%
Final Derivation	Manual review vs. (identification of final search terms in "chief complaint" section OR "diagnosis" section) AND documented imaging changes	96%	100%
Validation	Manual review vs. (identification of final search terms in "chief complaint" section OR "diagnosis" section) AND documented imaging changes	97 %	100%

Table 4 Derivation and validation of electronic search algorithm for myocardial infarction

	Search Algorithm	Sensitivity	Specificity
Initial Derivation	Manual review vs. documented biomarker changes	92%	68%
Final Derivation	Manual review vs. identification of final search terms in any note AND documented biomarker changes AND (documented ECG changes OR documented echocardiography changes)	100%	95%
Validation	Manual review vs. identification of final search terms in any note AND documented biomarker changes AND (documented ECG changes OR documented echocardiography changes)	100 %	99%

References

1. Hsiao CJ, Hing E, Socey TC, Cai B. Electronic health record systems and intent to apply for meaningful use incentives among office-based physician practices: United States, 2001–2011. NCHS data brief 2011: 1–8.
2. Singh B, Singh A, Ahmed A, Wilson GA, Pickering BW, Herasevich V, Gajic O, Li G. Derivation and validation of automated electronic search strategies to extract Charlson comorbidities from electronic medical records. *Mayo Clin Proc* 2012; 87(9): 817–824.
3. Ahmed A, Thongprayoon C, Pickering BW, Akhouni A, Wilson G, Pieczkiewicz D, Herasevich V. Towards prevention of acute syndromes: electronic identification of at-risk patients during hospital admission. *Appl Clin Inform* 2014; 5(1): 58–72.
4. Smischney NJ, Velagapudi VM, Onigkeit JA, Pickering BW, Herasevich V, Kashyap R. Retrospective derivation and validation of a search algorithm to identify emergent endotracheal intubations in the intensive care unit. *Appl Clin Inform* 2013; 4(3): 419–427.
5. Rishi MA, Kashyap R, Wilson G, Hocker S. Retrospective derivation and validation of a search algorithm to identify extubation failure in the intensive care unit. *BMC Anesthesiol* 2014; 14: 41.
6. Alsara A, Warner DO, Li G, Herasevich V, Gajic O, Kor DJ. Derivation and validation of automated electronic search strategies to identify pertinent risk factors for postoperative acute lung injury. *Mayo Clin Proc* 2011; 86(5): 382–388.
7. Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, Basford M, Chute CG, Kullo IJ, Li R, Pacheco JA, Rasmussen LV, Spangler L, Denny JC. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J Am Med Inform Assoc* 2013; 20(e1): e147–e154.
8. Carroll RJ, Thompson WK, Eyler AE, Mandelin AM, Cai T, Zink RM, Pacheco JA, Boomershine CS, Lasko TA, Xu H, Karlson EW, Perez RG, Gainer VS, Murphy SN, Ruderman EM, Pope RM, Plenge RM, Kho AN, Liao KP, Denny JC. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. *J Am Med Inform Assoc* 2012; 19(e1): e162–e169.
9. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am* 2007; 89(4): 780–785.
10. Mantilla CB, Horlocker TT, Schroeder DR, Berry DJ, Brown DL. Frequency of myocardial infarction, pulmonary embolism, deep venous thrombosis, and death following primary hip or knee arthroplasty. *Anesthesiology* 2002; 96(5): 1140–1146.
11. Mantilla CB, Horlocker TT, Schroeder DR, Berry DJ, Brown DL. Risk factors for clinically relevant pulmonary embolism and deep venous thrombosis in patients undergoing primary hip or knee arthroplasty. *Anesthesiology* 2003; 99(3): 552–560.
12. Memtsoudis SG, Della Valle AG, Besculides MC, Esposito M, Koulouvaris P, Salvati EA. Risk factors for perioperative mortality after lower extremity arthroplasty: a population-based study of 6,901,324 patient discharges. *J Arthroplasty* 2010; 25(1): 19–26.
13. Berry DJ, Kessler M, Morrey BF. Maintaining a hip registry for 25 years. Mayo Clinic experience. *Clin Orthop Relat Res* 1997; (344): 61–68.
14. Chute CG, Beck SA, Fisk TB, Mohr DN. The Enterprise Data Trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. *J Am Med Inform Assoc* 2010; 17(2): 131–135.
15. Herasevich V, Kor DJ, Li M, Pickering BW. ICU data mart: a non-iT approach. A team of clinicians, researchers and informatics personnel at the Mayo Clinic have taken a homegrown approach to building an ICU data mart. *Healthc Inform* 2011; 28(11): 42, 44–45.
16. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, Dittus RS, Rosen AK, Elkin PL, Brown SH, Speroff T. Automated identification of postoperative complications within an electronic medical record using natural language processing. *Jama* 2011; 306(8): 848–855.
17. Wisniewski MF, Kieszkowski P, Zagorski BM, Trick WE, Sommers M, Weinstein RA. Development of a clinical data warehouse for hospital infection control. *J Am Med Inform Assoc* 2003; 10(5): 454–462.

Appendix 1

Search terms for deep venous thrombosis

Inclusion terms for deep venous thrombosis:

”DVT“ (Match Case), ”Deep Vein Thrombosis“, ”Deep Vein Thromboses“, ”Venous thromboembolic disease“, ”Vein Thrombosis“, (”lower extremity“ Same sentence as ”thrombus in“), (”lower ex-

tremity“ Same sentence as ”thrombus within“), (”thrombus“ Same sentence as ”lower extremity“), (”positive for thromb“ Same sentence as ”abdomen“), (”pos for thromb“ Same sentence as ”abdomen“), (”positive for DVT“ Same sentence as ”abdomen“), (”pos for DVT“ Same sentence as ”abdomen“), (”positive for deep vein thromb“ Same sentence as ”abdomen“), (”pos for deep vein thromb“ Same sentence as ”abdomen“), (”positive for thromb“ Same sentence as ”pelvis“), (”pos for thromb“ Same sentence as ”pelvis“), (”positive for DVT“ Same sentence as ”pelvis“), (”pos for DVT“ Same sentence as ”pelvis“), (”positive for deep vein thromb“ Same sentence as ”pelvis“), (”pos for deep vein thromb“ Same sentence as ”pelvis“), (”positive for thromb“ Same sentence as ”lower-extremity“), (”pos for thromb“ Same sentence as ”lower-extremity“), (”positive for DVT“ Same sentence as ”lower-extremity“), (”pos for DVT“ Same sentence as ”lower-extremity“), (”positive for deep vein thromb“ Same sentence as ”lower-extremity“), (”pos for deep vein thromb“ Same sentence as ”lower-extremity“), (”positive for deep vein thromb“ Same sentence as ”lower extremity“), (”pos for deep vein thromb“ Same sentence as ”lower extremity“), (”pos for DVT“ Same sentence as ”lower extremity“), (”positive for DVT“ Same sentence as ”lower extremity“), (”pos for thromb“ Same sentence as ”lower extremity“), (”positive for thromb“ Same sentence as ”lower extremity“), ”Deep Venous Thrombosis“, ”vein DVT“, ”lower extremity DVT“

Exclusion terms for deep venous thrombosis:

(”Deep Venous“ Same sentence as ”prior“), (”Deep Venous“ Same sentence as ”negative“), (”Deep Venous“ Same sentence as ”prophylaxis“), (”Deep Venous“ Same sentence as ”risk“), (”Deep Venous“ Same sentence as ”history“), (”Deep Venous“ Same sentence as ”rule out“), (”Deep Venous“ Same sentence as ”rule-out“), (”Deep Vein“ Same sentence as ”prior“), (”Deep Vein“ Same sentence as ”negative“), (”Deep Vein“ Same sentence as ”prophylaxis“), (”Deep Vein“ Same sentence as ”risk“), (”Deep Vein“ Same sentence as ”history“), (”Deep Vein“ Same sentence as ”rule out“), (”Deep Vein“ Same sentence as ”rule-out“), (”DVT“ (Match Case) Same sentence as ”prior“), (”DVT“ (Match Case) Same sentence as ”negative“), (”DVT“ (Match Case) Same sentence as ”prophylaxis“), (”DVT“ (Match Case) Same sentence as ”risk“), (”DVT“ (Match Case) Same sentence as ”history“), (”DVT“ (Match Case) Same sentence as ”rule out“), (”DVT“ (Match Case) Same sentence as ”rule-out“), ”rule-out DVT“, ”rule-out Deep venous“, ”rule-out Deep vein“, ”rule out DVT“, ”rule out Deep venous“, ”rule out Deep vein“, ”history of DVT“, ”history of Deep vein“, ”Increased risk for DVT“, ”Increased risk for Deep venous“, ”Increased risk for Deep vein“, ”Increased risk of DVT“, ”Increased risk of Deep venous“, ”Increased risk of Deep vein“, ”negative for DVT“, ”negative for Deep venous“, ”negative for Deep vein“, ”no DVT“, ”no Deep venous“, ”no Deep vein“

Appendix 2

Search terms for pulmonary embolism

Inclusion terms for pulmonary embolism:

”PE“, ”pulmonary embolism“, ”Pulmonary Embolus“

Exclusion terms for pulmonary embolism:

(”PE“ (Match Case) Same sentence as ”rule-out“), (”PE“ (Match Case) Same sentence as ”rule out“), (”PE“ (Match Case) Same sentence as ”risk“), (”PE“ (Match Case) Same sentence as ”prophylaxis“), (”PE“ (Match Case) Same sentence as ”negative“), (”PE“ (Match Case) Same sentence as ”history“), (”prior“ Same sentence as ”PE“ (Match Case)), (”pulmonary embolism“ Same sentence as ”prior“), (”pulmonary embolism“ Same sentence as ”negative“), (”pulmonary embolism“ Same sentence as ”prophylaxis“), (”pulmonary embolism“ Same sentence as ”risk“), (”pulmonary embolism“ Same sentence as ”history“), (”pulmonary embolism“ Same sentence as ”rule out“), (”pulmonary embolism“ Same sentence as ”rule-out“), ”PE prophylaxis“, ”pulmonary embolism prophylaxis“, ”Increased risk for PE“, ”Increased risk of PE“, ”history of PE“, ”history of Pulmonary embolism“, ”rule out PE“, ”rule out Pulmonary embolism“, ”rule-out PE“, ”rule-out Pulmonary embolism“, ”no PE“, ”no pulmonary embolism“, ”negative for PE“, ”negative for Pulmonary embolism“, ”Increased risk of Pulmonary embolism“, ”Increased risk for Pulmonary embolism“

Appendix 3

Search terms for myocardial infarction

Inclusion terms for myocardial infarction:

"MI" (Match Case), "myocardial infarc", "heart attac", "myocardial ischem", "NSTEMI" (Match Case), "STEMI" (Match Case), "Non-ST elevation infarc", "Non ST elevation infar", "ST elevation infarc", "Non-Q-Wave MI", "Non-Q-Wave Myocardial Infarc", "Non Q Wave Myocardial Infarc"

Exclusion terms for myocardial infarction:

None