


Natural Language Processing Performance for the Identification of Venous Thromboembolism in an Integrated Healthcare System

Clinical and Applied
Thrombosis/Hemostasis
Volume 27: 1-6
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10760296211013108
journals.sagepub.com/home/cat


Bela Woller¹, Austin Daw, BS², Valerie Aston, MBA³, Jim Lloyd, BS⁴, Greg Snow, PhD⁵, Scott M. Stevens, MD⁶, Scott C. Woller, MD⁶ , Peter Jones, MSLS⁷, and Joseph Bledsoe, MD^{8,9} 

Abstract

Real-time identification of venous thromboembolism (VTE), defined as deep vein thrombosis (DVT) and pulmonary embolism (PE), can inform a healthcare organization's understanding of these events and be used to improve care. In a former publication, we reported the performance of an electronic medical record (EMR) interrogation tool that employs natural language processing (NLP) of imaging studies for the diagnosis of venous thromboembolism. Because we transitioned from the legacy electronic medical record to the Cerner product, iCentra, we now report the operating characteristics of the NLP EMR interrogation tool in the new EMR environment. Two hundred randomly selected patient encounters for which the imaging report assessed by NLP that revealed VTE was present were reviewed. These included one hundred imaging studies for which PE was identified. These included computed tomography pulmonary angiography—CTPA, ventilation perfusion—V/Q scan, and CT angiography of the chest/ abdomen/pelvis. One hundred randomly selected comprehensive ultrasound (CUS) that identified DVT were also obtained. For comparison, one hundred patient encounters in which PE was suspected and imaging was negative for PE (CTPA or V/Q) and 100 cases of suspected DVT with negative CUS as reported by NLP were also selected. Manual chart review of the 400 charts was performed and we report the sensitivity, specificity, positive and negative predictive values of NLP compared with manual chart review. NLP and manual review agreed on the presence of PE in 99 of 100 cases, the presence of DVT in 96 of 100 cases, the absence of PE in 99 of 100 cases and the absence of DVT in all 100 cases. When compared with manual chart review, NLP interrogation of CUS, CTPA, CT angiography of the chest, and V/Q scan yielded a sensitivity = 93.3%, specificity = 99.6%, positive predictive value = 97.1%, and negative predictive value = 99%.

Keywords

pulmonary embolism, PE, computerized tomography pulmonary angiography, natural language processing, NLP, venous thromboembolic disease, VTE

Date received: 28 December 2020; revised: 25 February 2021; accepted: 30 March 2021.

¹ Loyola University Chicago, Undergraduate Education, Chicago, IL, USA

² University of Colorado Health Sciences Center, Office of Human Research, Aurora, CO, USA

³ Intermountain Healthcare, Office of Research, Acute Care Research, Salt Lake City, UT, USA

⁴ Intermountain Healthcare, Informatics and Analytics, Salt Lake City, UT, USA

⁵ Intermountain Healthcare, Office of Research, Statistical Data Center, Salt Lake City, UT, USA

⁶ Department of Medicine, Intermountain Medical Center and University of Utah, Salt Lake City, UT, USA

⁷ Intermountain Healthcare, Enterprise Analytics, Salt Lake City, UT, USA

⁸ Department of Emergency Medicine, Intermountain Healthcare, Salt Lake City, UT, USA

⁹ Department of Emergency Medicine, Stanford Medicine, Palo Alto, CA, USA

Corresponding Author:

Joseph Bledsoe, Department of Emergency Medicine, Intermountain Medical Center, 5121 S. Cottonwood Drive, Murray, UT 84157, USA; Stanford Medicine, Palo Alto, CA, USA.

Email: joseph.bledsoe@imail.org



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons

Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use,

reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Introduction

Venous thromboembolism (VTE), including deep vein thrombosis (DVT) and pulmonary embolism (PE) complicates surgical procedures, prolongs hospital stays, and when undiagnosed, increases mortality.^{1,2} While the lifetime risk of VTE approximates 1:1000, VTE disproportionately affects the elderly, hospitalized medically ill and those with cancer.³⁻⁵ The risk of VTE increases as much as 20-fold following surgery.⁶ The detection of VTE among hospitalized patients informs decision-making surrounding VTE risk and thrombosis risk mitigation. Automated methods for the identification of VTE outcomes among hospitalized patients may further enhance improvements in care. Methods of electronic medical record (EMR) interrogation using embedded computer algorithms for the identification of outcome events have been described.^{1,2,7-9}

In a former study we reported the operating characteristics of an EMR-embedded algorithm that employed natural language processing (NLP) in our legacy electronic health record.² Natural language processing, also referred to as "text mining,"¹⁰ is programmed to interrogate free-text reports from different sources such as radiology reports, progress notes, and chart documentation to identify structured language that documents the diagnosis or findings of interest.⁹ Studies reporting NLP used to identify VTE, but also to identify VTE risk factors have been published.^{1,2,7,11-13} Furthermore, NLP has been adopted more broadly to identify other outcomes, such as patient safety indicators¹⁴ and operative complications.⁷ Healthcare systems use these data to enhance patient safety, demonstrate quality in health systems ratings and rankings,^{15,16} and promote cost-conscious care.¹⁷

We formerly reported that in our legacy EMR the NLP interrogation tool that we derived had a 92% sensitivity and a 99% specificity when detecting DVT, and a 100% sensitivity and a 98% specificity when detecting PE.² However, in 2017 Intermountain Healthcare discontinued the use of the legacy EMR. In a joint initiative with Cerner LLC a new EMR, iCentra, that would fully meet requirements of the Affordable Care Act¹⁸ while retaining much of the functionality^{19,20} that was historically integrated in the legacy EMR, was developed. The purpose of this study is to report the operating characteristics of the NLP iCentra-embedded technology to assure that this method continues to reliably identify patients with DVT and PE.

Methods

An electronic random number generator was used to select encounters that occurred between 1 January 2018 and 31 December 2018. One hundred patient encounters for which NLP identified PE by any imaging modality (CT pulmonary arteriography (CTPA) $n = 70$, CT of the chest, abdomen, and pelvis $n = 8$, CT of the thorax with contrast $n = 18$, and ventilation/perfusion (V/Q) scans $n = 4$) were acquired. A random sample of 50 V/Q scans and 50 CTPA for the assessment of suspected PE in which no thrombosis was identified

upon NLP interrogation, was generated for comparison. Similarly, 100 random patient encounters for which NLP identified DVT on comprehensive ultrasound ((CUS) $n = 100$) including compression ultrasound of a unilateral, bilateral, upper, or lower extremity, were selected. A random sample of 100 CUS performed for the assessment of suspected DVT, with negative result per NLP, were identified.

Manual chart review was conducted by 2 authors (AD, IAW) to ascertain if venous thrombosis was present or absent. If uncertainty existed, then 2 authors (JRB, SCW) independently reviewed the charts to ascertain consensus. To calculate the sensitivity and specificity of NLP we ascertained the prevalence of PE and DVT from our healthcare system among CTPA and V/Q scans ordered for suspected PE (8.7%) and CUS ordered for suspected DVT (14.2%).

Results

Natural language processing when compared with the gold-standard of manual chart review for the aggregate outcome of venous thromboembolism (DVT + PE) yielded a sensitivity of 93.3% (95% CI 82.9-99.1), a specificity of 99.6% (95% CI 99.2-99.9), and a positive and negative predictive value of 97.1% (95% CI 94.3-99.8) and 99.0% (95% CI 97.3-99.9); respectively. This is represented in Figure 1, patient demographics in Table 1, and a tabular representation of NLP for all imaging, CT, CUS, and V/Q scan is presented in Table 2.

For the outcome of PE the sensitivity of NLP was 95.9% (95% CI 89.4-99.5) with a specificity of 99.1% (95% CI 97.6-99.9), and a positive and negative predictive value of 98.1% (95% CI 94.7-99.8) and 98.0% (95% CI 94.6-99.8); respectively. CTPA alone for the outcome of PE yielded a sensitivity of 96.1% (86.7-99.9), specificity of 99.1% (97.3-99.9) and a positive and negative predictive value of 98% (94.2-99.8) and 98.1% (93.1-100); respectively. Ventilation/perfusion scans alone demonstrated a sensitivity of 91.3% (77.9-98.8%), a specificity of 94.3% (85.1-99.8) and a positive and negative predictive value of 87% (63.7-99.5) and 96.1% (89.1-99.5); respectively. See Figure 2A-C.

When the performance of NLP was assessed for the ascertainment of DVT among patients that received comprehensive ultrasound, NLP yielded a sensitivity of 96.1% (95% CI 87-99.9), a specificity of 98.8% (95% CI 97.7-99.6), and a positive and negative predictive value of 95.1% (95% CI 90.2-98.4) and 99% (95% CI 96.5-99.9); respectively (Figure 2D).

The distributions of the proportions surrounding the point estimates are not normative and is graphically represented in Figure 2.

Discussion

We report that our ability to reliably ascertain the outcome of VTE using various modalities including CTPA, CT of the chest/abdomen/pelvis, CT of the thorax, V/Q scan, and CUS of the upper or lower extremities is generally excellent. The detection of VTE in a reliable standardized fashion can

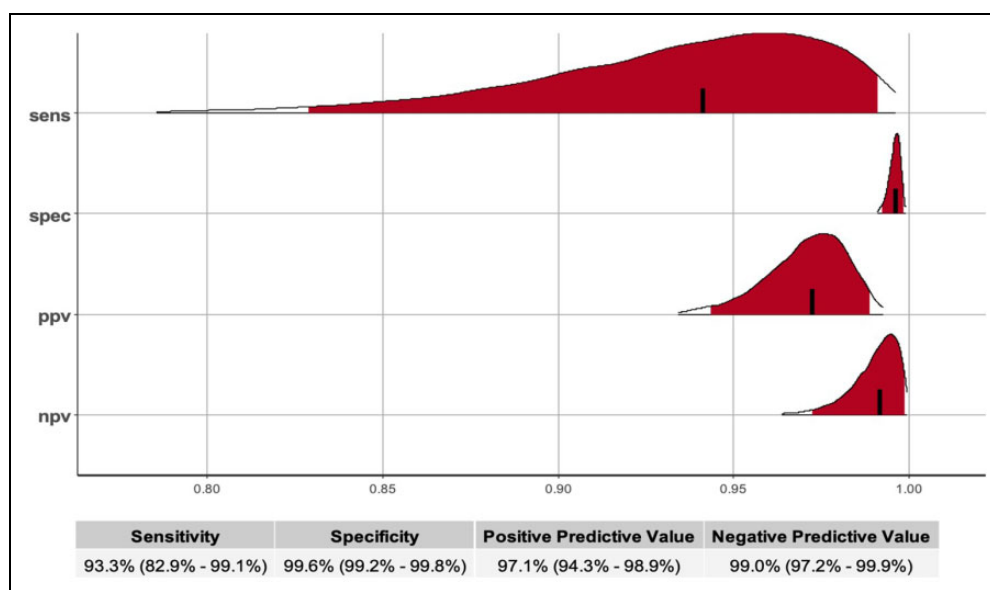


Figure 1. Natural language processing operating characteristics for all imaging in the detection of venous thromboembolism.

Table 1. Demographics of All Patients Studied.

	Percentage (%)
Sex (female)	57.25
Age; mean in years (standard deviation)	60.6 (17.9)
Length of stay in days (standard deviation)	2.5 (4.8)
Cancer	18
Obesity	70
Prior venous thromboembolism	19.8
Hypercoagulability (defined as a laboratory thrombophilia)	8
Hormone replacement therapy	3
Congestive heart failure	15.5
Diabetes	13.75
Current tobacco use	17.25
Surgery in the preceding 30 days	16.75
Infection	9.25
Peripherally inserted central catheter (PICC) line	5.5
Sepsis	8

enhance patient care, facilitate system-wide interventions to improve quality, enhance a system's reputation, and reduce medical cost. Reliable electronic tools that can perform this case identification should enhance patient care improvement efforts. Our observations affirm that upon electing a new EMR we have successfully introduced NLP technology for the outcome of VTE. The operating characteristics that we observed in this study are not dissimilar from the performance that we observed by the NLP tool in the legacy EMR.

While historically the performance of NLP has been compared with ICD codes representative of thrombotic outcomes for accuracy, we elected to report the performance of NLP comparing with manual chart review. To optimally report the performance of our NLP tool and calculate a negative predictive value, we selected random studies that were ordered for the

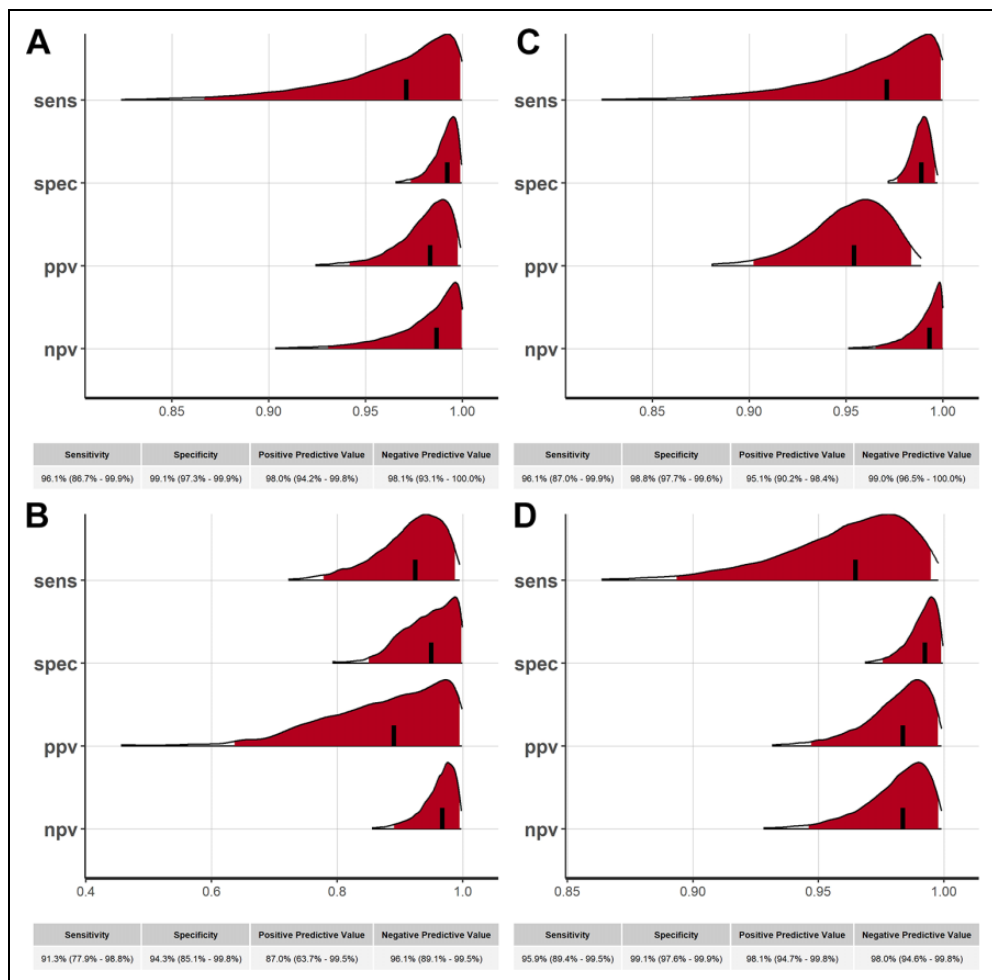
clinical suspicion of VTE in which VTE was refuted. When proportions are close to 1 (or 0) the likely values are not symmetric around the point estimate. For this reason we provide plots of the full posterior distributions (Figures 1 and 2) to show the likelihood of various values in addition to the credible (confidence) interval. We report with a good degree of certainty that we are not missing thrombotic events given our observed NPV = 99%. Additionally, while we report the confidence intervals surrounding the point estimates of the operating characteristics of NLP, inspection of the posterior distribution demonstrates a greater probability that the performance is more favorable. This is evidenced by the asymmetric distributions of the credible intervals skewed to the right in Figures 1 and 2.

The NLP performance that we report is similar to those formerly reported by others. Galvez in a review of 250 charts for the outcome of pediatric DVT reported that their NLP program, Reveal NLP, had a sensitivity of 97.2% and a specificity of 92.5%.¹ A university-affiliated healthcare system, not dissimilar to ours, showed that for DVT a 94% sensitivity and a 96% specificity were achieved and for PE a 94% sensitivity and 96% specificity was reported.²¹ Finally, a study looking at radiology reports to analyze for DVT and PE found similar results, with a 96% sensitivity and a 94% specificity when classifying for both DVT and PE. Subsequent research in this domain may involve enhancement of NLP using machine learning techniques which suggest that even better performance for the detection of thromboembolic disease may be achievable.⁸

One strength of this study is that we now report the false negative rate based upon review of charts for patients that were suspected of having VTE, however that NLP identified as VTE being absent. This observation informs reassurance that our EMR interrogation tool is not missing thrombosis. Likewise,

Table 2. Outcome Results.

	All imaging for VTE, mean % (95% CI)	CT pulmonary arteriography, mean % (95% CI)	Comprehensive ultrasound, mean % (95% CI)	Ventilation/perfusion lung scintigraphy, mean % (95% CI)	PE-specific imaging (CTPA + V/Q scan), mean % (95% CI)
Sensitivity	93.3 (82.9-99.1)	96.1 (86.7-99.9)	96.1 (87.0-99.9)	91.3 (77.9-98.8)	95.9 (89.4-99.5)
Specificity	99.6 (99.2-99.8)	99.1 (97.3-99.9)	98.80 (97.7-99.6)	94.3 (85.1-99.8)	99.1 (97.6-99.9)
Positive predictive value	97.1 (94.3-98.9)	98.0 (94.2-99.8)	95.1 (90.2-98.4)	87.0 (63.7-99.5)	98.1 (94.7-99.8)
Negative predictive value	99.0 (97.2-99.9)	98.1 (93.1-100)	99.0 (96.5-100)	96.1 (89.1-99.5)	98.0 (94.6-99.8)

**Figure 2.** Natural language processing operating characteristics for the detection of pulmonary embolism by computed tomography (A), ventilation/perfusion scan (B), comprehensive ultrasound (C), and computed tomography + ventilation/perfusion scan (D).

our analysis provides insight regarding the performance of our NLP in a “real-world” setting, that includes having identified PE ascertained on studies for which PE may not have been the initial diagnosis sought (e.g. CT of the thorax/chest/abdomen/pelvis). While these studies identify PE rarely, we deliberately created our query to be able to report the comparative frequency with which PE was observed in 100 random sample encounters of PE being found. This observation is informative to our learning healthcare system providing additional insight

regarding PE events that occur in routine care. Yet, because we wished to deliberately report the operating characteristics of those studies for which VTE is most frequently assessed, CUS, CTPA, and V/Q scan, we generated 100 CUS, 50 CTPA and 50 V/Q scans determined to be negative for thrombosis by NLP. It is for this reason that the distribution of studies on which PE is reported is not normal. Yet, to calculate the positive and negative predictive value, we use CTPA and V/Q scan results for which PE was absent on NLP.

Table 3. Discreet Events of Natural Language Processing Found to be in Error.

Imaging modality	Phrase assessed	Error type
CTPA	Resolution of previously noted pulmonary embolic disease.	False positive
CUS bilateral lower extremities	Specifically, and the nonocclusive thrombus in the left common femoral vein detected on prior examination is no longer visualized.	False positive
CUS left lower extremity	The gastrocnemius veins which previously had clot on the prior examination are now patent.	False positive
CUS left lower extremity	The deep vein thrombosis seen previously in the posterior tibial and peroneal veins is not definitively identified this time and is presumed to have resolved.	False positive
CUS left lower extremity	Diffuse Edema without acute or chronic deep vein thrombosis. FINDINGS: The common femoral, femoral, popliteal, posterior tibial and peroneal veins are patent, compressible and free of thrombus.	False positive

Abbreviations: CUS, comprehensive ultrasound; CTPA, computed tomography pulmonary arteriogram.

Iterative assessment and refinement of NLP is a central tenant to using this technology and it is for this reason that we publish the instances when we observed NLP failure (Table 3). In fact, as a result of this study, adjustment of our NLP code was made to improve performance.

Limitations of our study include that we manually reviewed 400 charts. It was our intent to understand the overall performance of NLP for the outcome of VTE. However, to report the performance of individual imaging modalities with a similar level of confidence additional manual chart review would be required. But we are encouraged by the results of subgroup analyses for the individual imaging modalities most used clinically (i.e. CTPA and CUS). It is possible that our assumptions of the prevalence of PE and DVT are imprecise which would affect the operating characteristics of the NLP reported. However these estimates are similar to our formerly reported rates,²² and also the prevalence of VTE that we use was assessed during the same time frame that the study was conducted. Even with these limitations, our study confirmed the effectiveness of NLP in detecting PE and DVT in the iCentra health record. The implications of this work are 2-fold. First, we will use this technique to report real-time rates of thrombosis in our system for clinical care and outcome surveillance. Second, we will with confidence be able to report rates of VTE identified by NLP in future studies that report this outcome in research.

Conclusion

NLP has been a reliable method to detect VTE rates and risk factors in the legacy Intermountain Healthcare EMR as we formerly reported.^{12,13,22,23} Recent adoption of the iCentra EMR prompted our revalidation of the NLP performance. We report favorable performance of NLP for the identification of VTE in iCentra. Validation of this technique permits us to inform others' efforts to adopt an analogous approach, and confidently use the rates of thrombosis that we identify to enhance patient care and conduct future research.


Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Scott C. Woller  <https://orcid.org/0000-0002-2522-2705>

Joseph Bledsoe  <https://orcid.org/0000-0002-3005-2878>

References

- Galvez JA, Pappas JM, Ahumada L, et al. The use of natural language processing on pediatric diagnostic radiology reports in the electronic health record to identify deep venous thrombosis in children. *J Thromb Thrombolysis*. 2017;44(3):281-290.
- Evans RS, Lloyd JF, Aston VT, et al. Computer surveillance of patients at high risk for and with venous thromboembolism. *AMIA Annu Symp Proc*. 2010;2010:217-221.
- Turetz M, Sideris AT, Friedman OA, Tripathi N, Horowitz JM. Epidemiology, pathophysiology, and natural history of pulmonary embolism. *Semin Intervent Radiol*. 2018;35(2):92-98.
- Goldhaber SZ. Venous thromboembolism: epidemiology and magnitude of the problem. *Best Pract Res Clin Haematol*. 2012; 25(3):235-242.
- Heit JA, Spencer FA, White RH. The epidemiology of venous thromboembolism. *J Thromb Thrombolysis*. 2016;41(1):3-14.
- White RH, Henderson MC. Risk factors for venous thromboembolism after total hip and knee replacement surgery. *Curr Opin Pulm Med*. 2002;8(5):365-371.
- Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA*. 2011;306(8): 848-855.
- Pham AD, Neveol A, Lavergne T, et al. Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings. *BMC Bioinformatics*. 2014;15(1):266.
- Pons E, Braun LM, Hunink MG, Kors JA. Natural language processing in radiology: a systematic review. *Radiology*. 2016; 279(2):329-343.
- Hearst M. Untangling text data mining. In: *Proceedings of the 37th Annual Meeting of the Association for Computational*

- Linguistics on Computational Linguistics*. Association for Computational Linguistics. 1999.
11. Spencer FA, Emery C, Joffe SW, et al. Incidence rates, clinical profile, and outcomes of patients with venous thromboembolism. The Worcester VTE study. *J Thromb Thrombolysis*. 2009;28(4):401-409.
 12. Woller SC, Stevens SM, Evans RS, et al. Electronic alerts, comparative practitioner metrics, and education improves thromboprophylaxis and reduces thrombosis. *Am J Med*. 2016;129(10):1124 e1117-1126.
 13. Woller SC, Stevens SM, Evans RS, et al. Electronic alerts, comparative practitioner metrics, and education improve thromboprophylaxis and reduce venous thrombosis in community hospitals. *Res Pract Thromb Haemost*. 2018;2(3):481-489.
 14. Quality. AfHRA. AHRQ Guide to Patient Safety Indicators Version 3.0a. Version 3.1. 2003.
 15. HealthGrades. HealthGrades annual patient safety in American hospitals study. 2020.
 16. Services CfMaM. Centers for Medicare and Medicaid Services, Medicare Quality Monitoring System. 2020.
 17. Goudie A, Dynan L, Brady PW, Fieldston E, Brilll RJ, Walsh KE. Costs of venous thromboembolism, catheter-associated urinary tract infection, and pressure ulcer. *Pediatrics*. 2015;136(3):432-439.
 18. Office of the National Coordinator for Health Information Technology DoH, Human S. Health information technology: initial set of standards, implementation specifications, and certification criteria for electronic health record technology. Final rule. *Fed Regist*. 2010;75(144):44589-44654.
 19. Pryor TA, Gardner RM, Clayton PD, Warner HR. The HELP system. *J Med Syst*. 1983;7(2):87-102.
 20. Evans RS, Pestotnik SL, Classen DC, et al. A computer-assisted management program for antibiotics and other antiinfective agents. *N Engl J Med*. 1998;338(4):232-238.
 21. Tian Z, Sun S, Eguale T, Rochefort CM. Automated extraction of VTE events from narrative radiology reports in Electronic Health Records: a validation study. *Med Care*. 2017;55(10):e73-e80.
 22. Woller SC, Stevens SM, Adams DM, et al. Assessment of the safety and efficiency of using an age-adjusted d-dimer threshold to exclude suspected pulmonary embolism. *Chest*. 2014;146(6):1444-1451.
 23. Woller SC, Stevens SM, Jones JP, et al. Derivation and validation of a simple model to identify venous thromboembolism risk in medical patients. *Am J Med*. 2011;124(10):947-954 e942.