American Journal of
Biomedical Science & Research

www.biomedgrid.com
ISSN: 2642-1747

Research Article

Definition of a Novel Imaging Quality Measure for the Evaluation of Emergency Department Patients with Suspected Pulmonary Embolism: Use of AI NLP to Validate and Automate It

Vladimir I Valtchinov[1,2,5]*, Ivan K Ip[1,3], Ramin Khorasani[1,2], Laila Cochon[1,2], Bingxue K Zhai[2], Ronilda Lacson[1], Shawn N Murphy[5,6] and Ali S Raja[1,4]

[1]*Center for Evidence-Based Imaging, USA*

[2]*Department of Radiology, Brigham and Women's Hospital (BWH), Harvard Medical School (HMS), USA*

[3]*Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, USA*

[4]*Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, USA*

[5]*Department of Biomedical Informatics, Harvard Medical School, USA*

[6]*i2b2 tranSMART Foundation, USA*

***Corresponding author:** Vladimir I Valtchinov, Center for Evidence-Based Imaging, Department of Radiology, Department of Biomedical Informatics, Brigham and Women's Hospital (BWH), Harvard Medical School (HMS), USA.

To Cite This Article: *Vladimir I Valtchinov, Definition of a Novel Imaging Quality Measure for the Evaluation of Emergency Department Patients with Suspected Pulmonary Embolism: Use of AI NLP to Validate and Automate It. 2020 - 9(3). AJBSR.MS.ID.001400.* DOI: 10.34297/AJBSR.2020.09.001400.

Received: : May 01, 2020; **Published:** July 06, 2020

Abstract

CT pulmonary angiography (CTPA) utilization rates for patients with suspected pulmonary embolism (PE) in the Emergency Department (ED) have increased steadily with associated radiation exposure, costs and overdiagnosis. A quality measure is needed to precisely assess efficiency of CTPA utilization, normalized to numbers of patients presenting with suspected PE and based on patient signs and symptoms. This study used Artificial Intelligence approaches such as ontology-driven natural language processing (NLP) to develop, automate, and validate SPE ("Suspected Pulmonary Embolism [PE]"), a measure determining CTPA utilization in ED patients with suspected PE. This retrospective study was conducted 4/1/2013-3/31/2014 in a Level-1 ED. A NLP engine processed "Chief Complaint" sections of ED documentation, identifying patients with PE-suggestive symptoms based on four Concept Unique Identifiers (CUIs: shortness of breath, chest pain, pleuritic chest pain, anterior pleuritic chest pain). SPE was defined as proportion of ED visits for patients with potential PE undergoing CTPA. Manual reviews determined specificity, sensitivity and negative predictive value (NPV). Among 5,768 ED visits with 1+SPE CUI, and 795 CTPAs performed, SPE=13.8% (795/5,768). AI and NLP identified patients with relevant CUIs with specificity=0.94 [95%CI (0.89-0.96)]; sensitivity=0.73 [95%CI (0.45-0.92)]; NPV=0.98. Using NLP on ED documentation can identify patients with suspected PE to computate a more clinically relevant CTPA measure. This measure might then be used in an audit-and-feedback process to increase the appropriateness of imaging of patients with suspected PE in the ED.

Keywords: Resource Utilization; Imaging Appropriateness; High-value Care; Pulmonary Embolism; Computed Tomography

Abbreviations: CTPA: Computer Tomography Pulmonary Angiography; PE: Pulmonary Embolism; NLP: Natural Language Processing; AI: Artificial Intelligence; ED: Emergency Department

Background

There is significant interest in the appropriateness of CT pulmonary angiography (CTPA) for patients with suspected acute pulmonary embolism (PE) in the Emergency Department (ED) [1-3]. CTPA utilization rates are variable but, overall, have increased steadily [1,4]. This rise in ED CTPA use results in increased radiation exposure [5], cost, and overdiagnosis – the latter of which can result in more downstream imaging and, potentially, inappropriate treatment (with its associated potential complications) [6].

Evidence-based recommendations exist to guide clinicians in the diagnostic workup of patients with suspected PE; the combination of risk-stratification using a validated tool (e.g., the Wells criteria [4,7]), supplemented by D-dimer measurement [8] has been used for over 15 years [9] and adopted by a number of professional societies [10,11]. However current measures of CTPA utilization or adherence to Wells criteria do not accurately capture providers' adherence to evidence, as patients who are appropriately not imaged are not well represented in existing measures. For example, appropriateness is often determined by using a denominator of patients who underwent CTPA and does not include patients in whom imaging was not ordered (who may have been excluded from imaging using the Pulmonary Embolism Rule-Out Criteria8 or clinical gestalt). Similarly, using overall CTPA use per ED visit does not limit the measure denominator to only patients with suspected PE, so comparisons between EDs with different prevalences of disease are not meaningful.

Thus, there is a need for a quality measure that precisely assesses the efficiency of CTPA utilization normalized to the number of patients with suspected PE who present to the ED, based on patients' signs and symptoms at presentation. The purpose of our study was to develop, automate, and validate a new tool – using unstructured data from clinical notes - to define a cohort of patients with suspected PE, which can then be used to develop a quality measure, Suspected Pulmonary Embolism (SPE).

Methods

Study Setting and Human Subjects Approval

This HIPAA-compliant- retrospective cohort study was conducted between April 1, 2013 and March 31, 2014 in the ED of an urban Level-I adult trauma center with ~60,000 visits annually. It was approved by the Institutional Review Board (Protocol Number: 2013P000267). Subsequent to the study period, the hospital underwent a change in the electronic medical record system, making the data capture necessary for this study more difficult.

Data Sources

Data sources included the ED information system, the radiology information system (RIS), and the computerized physician order entry (CPOE) system. For each ED visit, we obtained the text of the ED attending notes well as the text of the "Chief Complaint" field. From diagnostic imaging exam information, we extracted the imaging accession number, medical record number (MRN), and the final report text. All data fields collected were transferred to several tables in a Microsoft SQL server environment (Microsoft, Redmond, WA).

New SPE Measure Definition

We defined the new SPE ("Suspected Pulmonary Embolism [PE]") measure as:

$$SPE = \frac{number\ of\ ED\ visits\ with\ CTPA\ imaging}{number\ of\ all\ ED\ visits\ with\ suspected\ PE} \quad (1)$$

To construct the denominator for the imaging measure, we sought to quantify the cohort of ED patients with signs and symptoms at presentation suggestive of PE, we have used an Artificial Intelligence approaches such as ontologies-based, natural language processing (NLP) tool [12]. After consulting a multi-disciplinary group of clinical, informatics, and imaging experts, we based our algorithm on four of the most common signs and symptoms of PE as represented by Concept Unique Identifiers (CUIs) extracted from the ED note "Chief Complaint" field: shortness of breath (C0013404), chest pain (C0008031), pleuritic pain (C0008033, C0423632) and anterior pleuritic pain (C3532941).

NLP Engine and Customization

The AI NLP platform cTakes version 3.0.1 [13] [including YTEX [14,15] was customized with RadLex [16] and the latest releases of the SNOMED-CT vocabulary files using the NCI-supported Knowledge Representation languages RDF and process definitions from MetamorhoSys' sub-setting utility [17]. The extraction of the CUIs was done using a SQL query with multiple joins for the unique batch name of the job, resulting in a table, each line of which contained the CUI and the ID of the input "Chief Complaint" snippet of text. We also included polarity in the extraction query; a polarity of -1 corresponded to a negation of the named entity [13].

NLP Validation Process

To assess the accuracy of the AI NLP-based PE cohort discovery process, we conducted a manual validation, in which the results of a human-expert classification were compared to those extracted by the NLP algorithm. A physician research assistant was instructed and trained by an attending emergency physician to perform manual chart review classification while blinded to the results of the AI NLP-based classification. A validation sample size of 245 (5% of) cases was reviewed, and 10% of these were overread by the attending emergency physician.

Outcome Measures

As the primary outcome for this study, we computed the value for the new imaging measure, SPE. Our secondary outcomes were the test characteristics (sensitivity, specificity, positive predictive value [PPV], and negative predictive value [NPV]) of the NLP algorithm.

Statistical Analyses

All analyses and visualizations were carried out in the R statistical programming environment [18], version 3.0.2. We used Pearson r and t-test statistics to quantify correlation and similarity of distributions between monthly series of the specific measures. P-values<0.05 were considered statistically significant. The agreement between PE cohort discovery using the AI NLP algorithm and manual chart review was compared using sensitivity and specific-

ity (and corresponding 95% Confidence Intervals [CIs]), as well as NPV and PPV [19].

Results

Study Cohort

A total of 55,781 ED visits were recorded during the study. The mean patient age was 49±19.9 years, and 60.0% (33,467/55,781) were women. Of all ED patient visits, 38.6% (21,521/55,781) had diagnostic imaging. There were 1,159 CTPA imaging exams performed in ED during the study period.

The 4-CUI based algorithm flagged 5,768 ED visits suspected for PE (i.e., the NLP engine identified at lease one of the 4 CUIs in the Chief Complaint field text). Of these, 4,891 (or 84.8%) were mapped to radiology exams. Of those imaging exams, 795 were CTPAs.

Validation Process

Table 1 contains the results of the validation of the AI NLP algorithm, comparing the results achieved to that of the manual chart review. The specificity was 0.93 [95%CI (0.89-0.96)] and sensitivity was 0.73 [95%CI (0.45-0.92)], while NPV=0.98, and PPV=0.42.

Table 1: Results of cohort discovery validation comparing results achieved from natural language processing (cTakes) vs. the manual classification ("chart review") of the "Chief Complaint" field of 245 randomly selected entries from the Emergency Department visits data set.

		Chart Review		
Algorithm		Suspected Pulmonary Embolism (PE)	PE not suspected	
	Suspected PE	11	15	26
	PE not suspected	4	215	219
		15	230	

sensitivity: 73.33% (CI: 44.90% to 92.21%)
specificity: 93.48 % (CI: 89.47% to 96.30%)
PPV: 42.31%
NPV: 98.17 %

Distribution of the ED "Chief Complaint" Note Text

Figure 1 displays the results of a histogram analysis of the "Chief Complaint" field content across the mapping CTPA <-> ED visit. Notably, the large third bar corresponds to an empty "Chief Complaint" field. In addition, the shape of the distribution has a long tail corresponding to symptoms non-specific for PE, i.e., "fever" or "weakness".

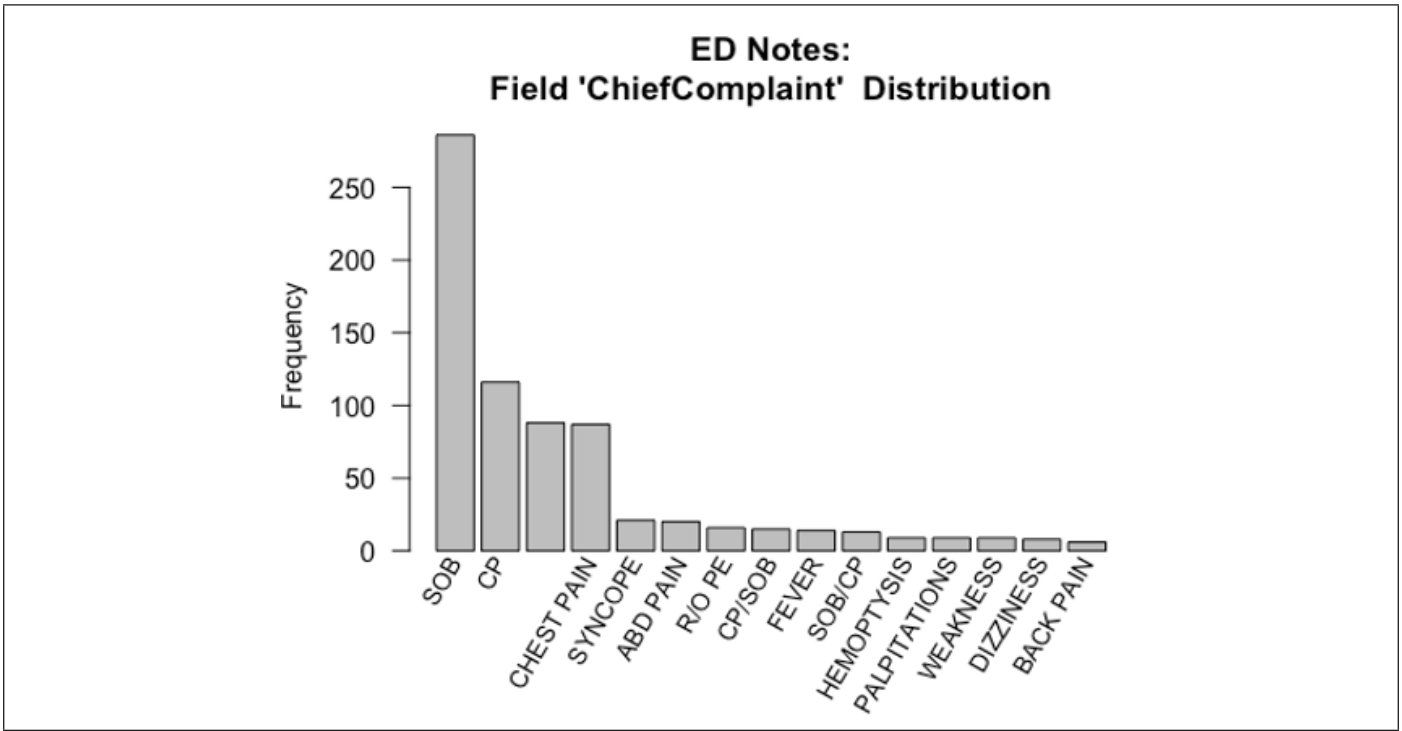

Figure 1: Frequency distribution of the top-15 most common "Chief Complaint" text snippets for all CT pulmonary angiography exams performed in the Emergency Department (ED) during the study period. Of note is the large third bar corresponding to empty "Chief Complaint" field, and the long tail of the distribution corresponding to other Concept Unique Identifiers.

Discussion

We have introduced and computed a new measure of utilization of CTPA imaging in patients with suspected PE in the ED. In contrast with existing imaging utilization metrics, SPE is normalized to the number of patients in whom PE is suspected, a patient cohort whose identification is based on patients' signs and symptoms at presentation. We have automated the calculation of this metric by casting it as an Artificial Intelligence NLP task on unstructured clinical narratives and structured EHR documentation, and then defining the cohort of PE-suspect patients using 4 common CUIs. Calculation of the new measure, SPE, resulted in 13.8% of patients presenting with symptoms of PE who obtained CTPA.

Current imaging quality measures fail to capture the appropriate patient populations. Appropriateness-based measures require resource-intensive calculation of pretest probability and d-dimer measurement, but still exclude patients in whom these data are not available, or who were excluded prior to the determination of these values (e.g., by using PERC.) Conversely, global utilization measures compute the number of CTPAs performed compared to overall ED visit volume, a method that cannot take into account local prevalence of PE.

Our validation of the algorithm for detecting patients with suspected PE had a sensitivity of 73% when compared to manual chart review. This is not surprising, given the other illnesses that can present with a "Chief Complaint" of chest pain or shortness of breath. However, the specificity of 94% and the NPV of 98% are reassuring, in that we likely excluded the vast majority of patients in whom PE would not have been suspected by the treating physician.

In order to determine whether the four CUIs we selected to model SPE patients were an adequate definition for the cohort, we reviewed the most common indications recorded in the "Chief Complaint" field (Figure 1), and found that the terms included in the 4-CUI model are the most common indications. Apart from the case of empty "Chief Complaint" fields (the relatively large 3rd bar in the figure), the rest of the indications are much less common, as well as much less clinically relevant, to PE. Additional studies with longer time spans and at additional institutions might be needed to elucidate this point further.

Limitations

Our study has a number of limitations. First and foremost, our algorithm is dependent on the quality of the data in the electronic health record, notably the presence and completeness of the "Chief Complaint" field in the ED record. For straightforward data gathering, we chose to base the algorithm on a single, pre-parsed text field from the ED notes ("Chief Complaint"), even though additional signs or symptoms might have been present in the free text of the "History of Present Illness" section. In addition, it was conducted in a single academic healthcare center, potentially limiting generalizability. Finally, the data are somewhat dated. However, this was unavoidable due to a change in the electronic medical record system at our hospital.

Implications

Our findings have the potential to improve the quality of care delivery by more accurately measuring the appropriateness of CTPA use for ED patients with suspected PE. Current measures typically only include patients who have undergone CTPA, missing completely those patients who are not imaged by physicians based on clinical criteria. Thus, physicians are unable to accurately determine whether they are appropriately evaluating patients with PE when compared with their peers, limiting the utility of audit-and-feedback reporting meant to improve the appropriateness of imaging.

It would be ideal to verify our findings across different institutions in both community and academic healthcare delivery settings to determine generalizability prior to widespread adoption of this new imaging metric. However, given the potential utility of this model in this imaging modality and indication, performing computation of imaging utilization metrics using appropriate patient cohorts using advanced but existing NLP public tools and ontologies is likely possible for other imaging scenarios as well.

For example, head CT imaging use in ED patients with mild traumatic brain injury (MTBI) has been shown be disproportionally variable [20]. At the same time, mature guidelines for use of imaging in MTBI exist (e.g., the Canadian CT Head Rule) [21]. Determining the rate of head imaging for patients with suspected MTBI, using appropriate CUIs, would be much more appropriate than the broad utilization metrics currently being considered [22]. Similarly, magnetic resonance imaging use in adult primary care patients with low back pain [23] – for which guidelines [24] and point-of-care clinical decision support implementations [23] both exist – might be an appropriate target as well.

Conclusions

Use of AI NLP of physician notes in the ED can help identify patients with suspected PE via flagging specific CUIs in the Chief Complaint field. This should allow for computation of a more clinically-relevant measure of imaging use efficiency of CTPA.

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of the Partners HealthCare and Brigham and Women's Hospital (Protocol Number: 2013P000267). As the study is a retrospective such reading on EMR data, the need for written consent was waived by the IRB as data was extracted from the EMR records of the hospital, specifically the Emergency Department (ED) EMR module.

Consent for Publication

Not Applicable

Availability of Data and material

The datasets generated and/or analyzed during the current study are not publicly available due [as it contains IRB sanctioned patient identified data] but are available from the corresponding author on reasonable request, after a proper de-identification.

Competing Interests

No financial or other, non-financial competing interest declared.

Funding

No funding was obtained for this study.

Authors' Contributions

All authors have read and approved the manuscript (Table)

Acknowledgements

Not Applicable

Reference

1. Lee J, Kirschner J, Pawa S, Wiener DE, Newman DH, et al. (2010) Computed tomography use in the adult emergency department of an academic urban hospital from 2001 to 2007. Ann Emerg Med 56(6): 591-596.
2. Raja AS, Gupta A, Ip IK, Mills AM, Khorasani R (2014) The use of decision support to measure documented adherence to a national imaging quality measure. Acad Radiol 21(3): 378-383.
3. An initiative of the ABIM Foundation. Choosing Wisely.
4. Broder J, Warshauer DM (2006) Increasing utilization of computed tomography in the adult emergency department, 2000-2005. Emerg Radiol 13(1): 25-30.
5. Sodickson A, Baeyens PF, Andriole KP, Luciano M Prevedello, Richard D Nawfel, et al. (2009) Recurrent CT, cumulative radiation exposure, and associated radiation-induced cancer risks from CT of adults. Radiology 251(1): 175-184.
6. Wiener RS, Schwartz LM, Woloshin S (2011) Time trends in pulmonary embolism in the United States: evidence of overdiagnosis. Arch Intern Med 171(9): 831-837.
7. Raja AS, Greenberg JO, Qaseem A, Denberg TD, Fitterman N, et al. (2015) Evaluation of Patients with Suspected Acute Pulmonary Embolism: Best Practice Advice From the Clinical Guidelines Committee of the American College of Physicians. Ann Intern Med 163(9): 701-711.
8. Fesmire FM, Brown MD, Espinosa JA, Richard D Shih, Scott M Silvers, et al. (2011) Critical issues in the evaluation and management of adult patients presenting to the emergency department with suspected pulmonary embolism. Ann Emerg Med 57(6): 628.e75-652.e75.
9. Wells PS, Anderson DR, Rodger M, I Stiell, J F Dreyer et al. (2001) Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and d-dimer. Ann Intern Med 135(2): 98-107.
10. Raja AS, Ip IK, Prevedello LM, Aaron D Sodickson, Cameron Farkas, et al. (2012) Effect of computerized clinical decision support on the use and yield of CT pulmonary angiography in the emergency department. Radiology 262(2): 468-474.
11. Segal JB, Eng J, Tamariz LJ, Bass EB (2007) Review of the evidence on diagnosis of deep venous thrombosis and pulmonary embolism. Ann Fam Med 5(1): 63-73.
12. Valtchinov VI, Lacson R, Wang A, Khorasani R (2020) Comparing Artificial Intelligence Approaches to Retrieve Clinical Reports Documenting Implantable Devices Posing MRI Safety Risks. J Am Coll Radiol 17(2): 272-279.
13. Savova GK, Masanz JJ, Ogren PV, Jiaping Zheng, Sunghwan Sohn et al. (2010) Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 17(5): 507-513.
14. Garla VN, Brandt C (2012) Ontology-guided feature engineering for clinical text classification. J Biomed Inform 45(5): 992-998.
15. Garla V, Re VL, Dorey Stein Z, Farah Kidwai, Matthew Scotch, et al. (2011) The Yale cTAKES extensions for document classification: architecture and application. J Am Med Inform Assoc 18(5): 614-620.
16. (2015) Terminology — CBIIT Welcome to the NCI Center for Biomedical Informatics and Information Technology.
17. (2015) MetamorphoSys Help.
18. R Core Team (2015) R: A Language and Environment for Statistical Computing.
19. Altman DG, Bland JM (1994) Diagnostic tests. 1: Sensitivity and specificity. BMJ 308(6943): 1552.
20. Gupta A, Ip IK, Raja AS, Andruchow JE, Sodickson A, et al. (2014) Effect of clinical decision support on documented guideline adherence for head CT in emergency department patients with mild traumatic brain injury. J Am Med Inform Assoc 21(e2): e347-351.
21. Stiell IG, Wells GA, Vandemheen K, Catherine Clement, Howard Lesiuk et al. (2001) The Canadian CT Head Rule for patients with minor head injury. The Lancet 357(9266): 1391-1396.
22. Schuur JD, Raja AS, Walls RM (2011) CMS selects an invalid imaging measure: déjà vu all over again. Ann Emerg Med 57(6): 704-705.
23. Ip IK, Gershanik EF, Schneider LI, Ali S Raja, Wenhong Mar, et al. (2014) Impact of IT-enabled intervention on MRI use for back pain. Am J Med 127(6): 512-518.e1.
24. Chou R, Qaseem A, Snow V, Donald Casey, J Thomas Cross, et al. (2007) Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med 147(7): 478-491.