

Medication Extraction from Electronic Clinical Notes in an Integrated Health System: A Study on Aspirin Use in Patients with Nonvalvular Atrial Fibrillation

Chengyi Zheng, PhD¹; Nazia Rashid, PharmD²; River Koblick, BA²; and Jaejin An, PhD³

¹Department of Research and Evaluation, Kaiser Permanente Southern California, Pasadena, California;

²Drug Information Services, Kaiser Permanente Southern California, Downey, California; and

³Department of Pharmacy Practice and Administration, Western University of Health Sciences, Pomona, California

ABSTRACT

Purpose: The purpose of this study was to investigate whether aspirin use can be captured from the clinical notes in a nonvalvular atrial fibrillation population.

Methods: A total of 29,507 patients with newly diagnosed nonvalvular atrial fibrillation were identified from January 1, 2006, through December 31, 2011, and were followed up through December 31, 2012. More than 3 million clinical notes were retrieved from electronic medical records. A training data set of 2949 notes was created to develop a computer-based method to automatically extract aspirin use status and dosage information using natural language processing (NLP). A gold standard data set of 5339 notes was created using a blinded manual review. NLP results were validated against the gold standard data set. The aspirin data from the structured medication databases were also compared with the results from NLP. Positive and negative predictive values, along with sensitivity and specificity, were calculated.

Findings: NLP achieved 95.5% sensitivity and 98.9% specificity when compared with the gold standard data set. The positive predictive value was 93.0%, and the negative predictive value was 99.3%. NLP identified aspirin use for 83.8% of the study population, and 70% of the low dose aspirin use was identified only by the NLP method.

Implications: We developed and validated an NLP method specifically designed to identify low dose aspirin use status from the clinical notes with high accuracy. This method can be a valuable tool to supplement existing structured medication data. (*Clin Ther.* 2015;■:■■■-■■■) © 2015 Elsevier HS Journals, Inc. All rights reserved.

Key words: aspirin, electronic clinical notes, electronic medical record, integrated health systems, medication status, natural language processing.

INTRODUCTION

Atrial fibrillation (AF) is the most common cardiac arrhythmia, affecting between 2.7 million and 6.1 million people in the United States, and its prevalence is expected to increase in the next few decades.¹ Most AF patients develop nonvalvular atrial fibrillation (NVAf), which is associated with a 5-fold increase in risk of ischemic stroke.² One of the main goals of NVAf treatment is to prevent stroke with the use of antithrombotic therapy. Clinical guidelines recommend either oral anticoagulant or antiplatelet therapy based on a patient's risk of stroke (eg, CHA₂DS₂-VASc score) calculated using clinical factors.¹ However, there may be high-risk patients who take aspirin instead of anticoagulants in a real-world setting because of reasons such as high risk of bleed, falls, or patient preference. Patients may also be taking both an anticoagulant and antiplatelet agent at the same time. Understanding the real-world treatment patterns and outcomes are critical; however, often this is challenging because of data limitations.

Typically, clinicians and researchers access patients' medication history using "structured" medication

Accepted for publication July 6, 2015.

<http://dx.doi.org/10.1016/j.clinthera.2015.07.002>

0149-2918/\$ - see front matter

© 2015 Elsevier HS Journals, Inc. All rights reserved.

databases. Most antithrombotic drugs are dispensed prescriptions and can be captured in structured medication databases. However, not all medications are available in the medication databases, such as over-the-counter (OTC) medications. Aspirin is an OTC drug commonly used as an antiplatelet therapy in patients with NVAF. One study estimated that 50 million people in the United States (36% of the adult population) are taking daily aspirin solely for cardiovascular disease and stroke prevention.³ It has been challenging to capture OTC aspirin use in a systematic way other than survey questionnaires.

One potential method for capturing OTC aspirin use may be collecting information from electronic medical record (EMR) systems. Large amounts of medication data are available to both clinical care and research in EMRs. Aspirin use may be documented by health care professionals in the free text clinical notes, within the EMR system. Manual review of clinical notes or medical records can identify aspirin information; however, this can be expensive and time-consuming.

There has been a growing interest in natural language processing (NLP). NLP is a field of computer science and linguistics that aims to understand human (natural) languages. This technique has been used to identify and extract information from the free-text formatted data. Rule-based and statistical machine learning methods are often used together to deliver a robust system.⁴ Compared with human review of medical records, NLP is more efficient and consistent.⁴ In recent years, NLP has gained wider adoption in the biomedical field.⁵

Understanding the use of both anticoagulants and antiplatelet therapy is important for NVAF patients, but capturing aspirin use has been extremely difficult compared with other prescription anticoagulant or antiplatelet agents. The purpose of this study was to determine aspirin use by applying an NLP algorithm. In this study, we developed and applied an NLP method on a large NVAF population. We identified whether a patient was undergoing aspirin therapy and then determined the status of aspirin use (on/off aspirin therapy) and the aspirin dosage information. We also evaluated the contribution of NLP on reducing missing aspirin data in our current structured medication databases.

MATERIALS AND METHODS

Study Setting

Kaiser Permanente Southern California (KPSC) provides integrated, comprehensive medical services to 3.9 million members through its own facilities. Every member receives a medical record number, which allows the member to be linked to various clinical and administrative databases, such as enrollment, drug benefits, medical services and visits, laboratory results, and pharmacy services. The aspects of care and interactions within this integrated care system were captured in an EMR system (Epic Systems, Epic Systems Corporation, Madison, Wisconsin, US), which is available for research purposes.

Study Population

The study population included all KPSC members aged ≥ 18 years with a new diagnosis of atrial fibrillation (≥ 2 serial *International Classification of Diseases, Ninth Revision Clinical Modification* [ICD-9-CM] codes of 427.31, ≥ 30 days apart) from January 1, 2006, through December 31, 2011. The first diagnosis date is the index date. Patients had to have 12 months of continuous membership eligibility before the index date (Figure 1). Patients were followed up until disenrollment, outcomes of interest, death, or December 31, 2012, whichever occurred first. Patients who had prior anticoagulation prescriptions or an anticoagulation clinic visit were excluded. Additional exclusion criteria were applied to deliver the final incident NVAF cohort ($n = 29,507$). The institutional review board at KPSC approved this study.

Data Collection

After the final cohort was created, all clinical notes from the index date through the end of follow-up date for each patient were extracted. All electronic clinical notes related to the study cohort ($n = 29,507$) were retrieved from our EMR system and available for this study. A total of 3,235,393 notes were retrieved for 29,425 patients (99.7% of the study cohort).

NLP Training and Evaluation Data

Training and gold standard data sets were created from 2 different random samples to evaluate aspirin use status and aspirin dosage. A random sample of 2949 notes was selected for the training data set. These notes were used to refine the NLP method. A second random sample of 5339 notes was selected

Download English Version:

<https://daneshyari.com/en/article/5825168>

Download Persian Version:

<https://daneshyari.com/article/5825168>

[Daneshyari.com](https://daneshyari.com)