2022

Addressing Ascertainment Bias in the Study of Cardiovascular Disease Burden in Opioid Use Disorders - Application of Natural Language Processing of Electronic Health Records

Jade Huang Singleton
University of Kentucky, jhsingle88@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0002-7736-8570
Digital Object Identifier: https://doi.org/10.13023/etd.2022.025

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Jade Huang Singleton, Student

Dr. Anna M. Kucharska-Newton, Major Professor

Dr. Heather Bush, Director of Graduate Studies

Addressing ascertainment bias in the study of cardiovascular disease burden in opioid use disorders - application of natural language processing of electronic health records

---

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Public Health
at the University of Kentucky

By

Jade Huang Singleton

Lexington, Kentucky

Co- Directors: Dr. Anna M. Kucharska-Newton, Professor of Epidemiology

and Dr. Erin L. Abner, Professor of Epidemiology

Lexington, Kentucky

2021

https://orcid.org/0000-0002-7736-8570

ABSTRACT OF DISSERTATION

ADDRESSING ASCERTAINMENT BIAS IN THE STUDY OF CARDIOVASCULAR DISEASE BURDEN IN OPIOID USE DISORDERS - APPLICATION OF NATURAL LANGUAGE PROCESSING OF ELECTRONIC HEALTH RECORDS

In the United States, the prevalence of long-term exposure to opioid drugs, for both medically and nonmedically indicated purposes, has increased considerably since the mid-1990's. Concerns have emerged about the potential health effects of opioid use. There is also growing interest in other possible connections with opioid use including cardiovascular disease. Electronic health records (EHR) contain information about patient care in the form of structured codes and unstructured notes. Natural language processing (NLP) provides a tool for processing unstructured textual data in EHR clinical notes and extracts useful information for research with structured formats. The purpose of this dissertation was to 1) to summarize peer-reviewed literature on the association between non-acute opioid and cardiovascular disease (CVD) and identify the gap of this research topic; 2) to apply NLP algorithm to estimate the extent of opioid use disorder (OUD) among hospital inpatients that cannot be identified using ICD-10-CM codes; and 3) to determine the extent to which estimates of the association between OUD and CVD may be biased by misclassification of OUD cases that are not identifiable using ICD-10-CM codes.

First, we conducted a scoping review of the epidemiological literature on nonacute opioid use and CVD. We summarized the current evidence about the association between NOU and CVD, and identified some open questions on this topic. Then, we developed a Natural Language Processing algorithm to identify cases of OUD in electronic healthcare records that were not assigned an ICD-10-CM code for OUD by medical records coders, but for which strong evidence of OUD exists in the unstructured clinical notes. Lastly, we estimated the association between OUD and six types of CVD, arrythmia, myocardial infarction, stroke, heart failure, ischemic heart disease, and infective endocarditis, classifying OUD in two ways: defining OUD cases by ICD-10-CM codes alone, and using a combination of cases identified by ICD-10-CM codes and cases identified using NLP algorithm. We assessed the effect of misclassification of OUD status when using ICD-10-CM codes alone.

KEYWORDS: Opioid use disorder, Cardiovascular disease, Natural language processing, electronic healthcare records

Jade Huang Singleton

11/05/2021

Addressing ascertainment bias in the study of cardiovascular disease burden in opioid use disorders - application of natural language processing of electronic health records

By
Jade Huang Singleton

Anna M. Kucharska-Newton, PhD
Co-Director of Dissertation

Erin L. Abner, PhD
Co-Director of Dissertation

Heather Bush, PhD
Director of Graduate Studies

12/03/2021
Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ADDITIONAL FILES

CHAPTER 1. INTRODUCTION

Exposure to opioid drugs among United States (U.S.) residents has increased exponentially over the past 30 years (Hedegaard et al., 2018). This includes the use of prescription opioids for medical purposes, as directed by a physician – such as for treatment of opioid dependence, cancer-related pain, or noncancer chronic pain – as well as non-medically indicated use of both prescription and illicit opioids. A major driver of increasing opioid use was the perception of under-treatment of pain (Bernard et al., 2018; Meldrum, 2016). Initially, opioid treatment was restricted to cancer-related pain and then later expanded to non-cancer related pain. As a result, opioid prescribing in the U.S. increased nearly seven-fold between 1997 and 2007, from 100 to nearly 700 morphine milligram equivalents per capita (Paulozzi et al., 2011). This increase in opioid availability was accompanied by steep increases in fatal and nonfatal overdoses and an increase in the prevalence of opioid use disorders (OUD) (Haight et al., 2018; Martins et al., 2017). Intravenous use of both prescription and illicit opioids has contributed to outbreaks of infectious diseases including HIV, hepatitis, and infective endocarditis (National Academies of Sciences and Medicine 2020).

Opioid use disorder is a chronic relapsing disorder that increasingly engages anti-reward neurocircuits that drive adverse emotional states and relapse (Strang et al., 2020). Several consequences of OUD cause substantial burden to the individual, their family and the community. For example, OUD itself carries a substantial health burden owing to the disability associated with OUD and the risk of over-dose. People who developed OUD have an increased risk of different health issues, for example, injuries, suicide, homicide,

blood-borne virus infections compared to the general population (Larney, 2020 ; Maloney, 2007; Pierce, 2015; Suryaprasad, 2014; Volkow, 2016; Zibbell, 2015).

In recent years, concerns have emerged about possible cardiovascular effects of opioid use. The risk of infective endocarditis associated with injection drug (including opioid) use has been well-documented (Behzadi et al., 2018; Mihm et al., 2020; Sinner et al., 2021) and they published a review on the relationship between opioid use and cardiac arrhythmia. The authors reported that methadone posed a high risk of QT interval prolongation and arrhythmogenicity, even at low doses. A small number of studies have documented a positive association between opioid exposure and the risk of acute myocardial infarction (MI) (Carman et al., 2011; Jobski et al., 2017; Li et al., 2013; Roberto et al., 2015). Given the high prevalence of opioid use in the U.S. population, even a relatively small effect of opioids on the risk of cardiovascular disease (CVD) would have significant public health implications.

Research on the possible cardiovascular effects of long-term opioid use presents many challenges. Randomized trials would clearly be unethical. CVD has a long latency period. Exposure to prescription opioids for medical purposes would be difficult enough to measure accurately over a such a long period of time, not to mention nonmedical use of prescription and illicit opioids.

One approach is to leverage administrative hospital records to identify the burden of CVD in association with opioid drug use. Both OUD and cardiovascular conditions can be identified in these data sources using International Classification of Diseases (ICD) codes, which are commonly used for this purpose due to their widespread use in medical record coding and their accessibility to researchers (Beam et al., 2021; Mezzich, 2002).

The limitations of ICD codes, including low sensitivity and specificity for many conditions, have been well-documented (Hughes Garza et al., 2021; Kurbasic et al., 2008; O'Malley et al., 2005; Quan et al., 2008), although there has been little investigation of the extent to which this may be true for OUD. If ICD codes have low sensitivity to identify OUD when it is present, then many patients with OUD would be misclassified, in this type of study, as not having OUD. This could result in biased estimates of the association between OUD and CVD.

Electronic health records (EHR) contain – in addition to structured ICD codes – substantial information about patient care in the form of unstructured, narrative text entered by providers, nurse, lab technician or any other member of patient's healthcare team (Spasic and Nenadic, 2020). These clinical notes include information on patient symptoms, conditions, behaviors, and healthcare advice and plans (Wang et al., 2018). Generally, information in unstructured notes includes demographics, medical encounters, developmental history, obstetric history, medications and medical allergies, family history, social history, habits, immunization records (Gliklich et al., 2019).

Natural language processing (NLP) provides a tool for processing unstructured textual data in EHR clinical notes. (Pendergrass and Crawford, 2019). Information extraction is a subtask of NLP that is focused on the extraction of structured data from text (Ford et al., 2016; Meystre et al., 2008). Typically, information extraction involves splitting text into basic units called tokens, which are individual words, punctuation marks, etc. (Ford et al., 2016). Rule-based approaches to information extraction attempt to identify matches of pre-specified sequences of tokens (Nadkarni et al., 2011). Statistical approaches rely on probabilistic models, or on supervised learning methods applied to very

large corpuses of text that have been labeled to indicate which instances do, and which do not, contain the information of interest (Carrell et al., 2015). In supervised learning approaches, systems can be trained to recognize entities within text documents by seeing many correctly-labelled examples and "learning" features of the text that accurately predict the presence of those entities (Spasic and Nenadic, 2020). Although powerful, a limitation of this approach is the enormous effort required to create a sufficiently large, prelabeled corpus of examples (Nadkarni et al., 2011; Velupillai et al., 2018).

Whereas NLP has been successfully applied to EHR records to extract information on certain disease conditions, including cancer and diabetes, there has been little previous work on OUD (Sheikhalishahi et al., 2019; Wang et al., 2018). Carrell et al (2015) investigated the potential to apply NLP to EHR records to increase the identification of problem use of prescription opioids (POU) among patients undergoing chronic opioid therapy. POU was defined as indications of addiction, abuse, misuse or overuse, and is thus more broadly defined and less specific than OUD. Using NLP methods, Carrell et al identified 33% more cases of POU than could be identified by ICD-9-CM codes alone (Carrell et al., 2015).

There is an opportunity to utilize NLP methods to answer important questions about the sensitivity of ICD codes in the identification of OUD from administrative hospital data sources, and the effect on research into the association between OUD and CVD.

The aims of this dissertation were 1) to summarize peer-reviewed literature on the association between non-acute opioid use and CVD; 2) to use NLP methods to estimate the extent of OUD among hospital inpatients that cannot be identified using ICD codes; and 3) to determine the extent to which estimates of the association between OUD and

CVD may be biased by misclassification of OUD cases that are not identifiable using ICD codes.

Chapter Two represents a scoping review of the epidemiological literature on nonacute opioid use (Kivimaki et al.) and CVD, summarizing the current evidence about the association between NOU and CVD, and identifying open questions on this topic. In Chapter Three, we develop a Natural Language Processing (NLP) pipeline for identifying cases of OUD in electronic patient records that were not assigned an ICD-10-CM code for OUD by medical records coders, but for which strong evidence of OUD exists in the unstructured clinical notes. In Chapter Four, we estimate the association between OUD and six types of CVD. We classify OUD in two ways: using ICD codes alone, and using a combination of ICD codes and cases of OUD identified using NLP. We assess the effect of misclassification of OUD status using ICD codes alone. Chapter Five presents an integrative summary of our findings.

CHAPTER 2. ASSOCIATION OF NONACUTE OPIOID USE AND CARDIOVASCULAR DISEASES: A SCOPING REVIEW OF THE LITERATURE

2.1 Abstract

In this scoping review, we identified and reviewed twenty-three original articles from the PubMed database that investigated the relationship between nonacute opioid use (Kivimaki et al.) and cardiovascular outcomes. We defined NOU to include both long-term opioid therapy and opioid use disorder. We summarized the association between NOU and five classes of cardiovascular disease (CVD), including infective endocarditis, coronary heart disease (including myocardial infarction), congestive heart failure, cardiac arrythmia (including cardiac arrest), and stroke. The most commonly studied outcomes were coronary heart disease and infective endocarditis. There was generally consistent evidence of a positive association between community prevalence of injection drug use (with opioids being the most commonly injected type of drug) and community prevalence of IE, and between (primarily medically indicated) NOU and MI. There was less consensus about the relationship between NOU and CHF, cardiac arrhythmia, and stroke. There is a dearth of high-quality evidence on the relationship between NOU and CVD. Innovative approaches to the assessment of opioid exposure over extended periods of time will be required to address this need.

2.2 Background

Exposure to opioid drugs in the United States (U.S.) has increased exponentially over the past 30 years (Haight et al., 2018; Hedegaard et al., 2018; Porter and Jick, 1980; Singh and Cleveland, 2020). This includes use of prescription opioids for medical purposes as directed by a physician – such as, treatment of opioid dependence, cancer-related pain,

or non-cancer chronic pain – as well as non-medically indicated use of prescription opioids and illicit opioid use. Opioid treatment, initially restricted to cancer patients, expanded over time to include the treatment of non-cancer related pain. A growing awareness of the problem of undertreated pain resulted in standards, issued by the Joint Commission on Accreditation of Healthcare Organizations in 2001, requiring greater monitoring and treatment of pain (Max, 1990; Phillips, 2000). Pharmaceutical companies aggressively marketed opioid medications for treatment of chronic pain (Maxwell, 2011), citing flawed research studies as evidence of the safety of these medications (Portenoy and Foley, 1986). As a result of these and other factors, opioid prescribing in the U.S. increased nearly seven-fold between 1997 and 2007. The increase in opioid availability was accompanied by steep increases in fatal and nonfatal overdoses (Paulozzi et al., 2011) and opioid use disorder (OUD) (Haight et al., 2018; Martins et al., 2017).

Long-term opioid therapy – such as for the treatment of chronic pain or opioid addiction – has been defined as use of opioids on most days for more than 3 months (Dowell et al., 2016). Long-term exposure to opioids may also result from the nonmedical use of prescription or illicit opioids due to dependence or addiction, leading to an opioid use disorder (OUD). We use the term "nonacute opioid use" (Kivimaki et al.) to encompass both long-term opioid therapy and OUD.

Researchers have begun to investigate possible effects of long-term opioid use on health outcomes other than addiction and misuse. Of specific interest is cardiovascular disease, which remains a leading cause of death, physician and emergency department visits, and hospitalization in the United States (Geidrimiene and King, 2017; Mensah and Brown, 2007). Trends in hospitalized cases of infective endocarditis (Beck, 2019) have

been shown to mirror trends in opioid overdose and injection drug use (National Academies of Sciences and Medicine, 2020). Observational studies have also reported an association of opioid drug use with increased risk of cardiovascular events, including MI and heart failure.

However, it is unclear what the biological pathways between long-term exposure to opioids and CVD might be. Opioid receptors have been discovered in the heart, and their activation by short-term administration of opioid drugs prior to acute ischemic events has been shown to have a cardioprotective effect (Schultz and Gross, 2001). However, the association between chronic opium use and increased levels of low-density lipoproteins and triglycerides could provide a pathway to coronary artery disease (Aghadavoudi et al., 2015; Zagaria, 2018). High and increasing prevalence of NOU and a sustained high burden of cardiovascular disease have prompted this scoping review of the literature to systematically examine the association of NOU with cardiovascular outcomes.

2.3 Methods

We identified original, peer-reviewed research articles on the relationship between NOU involving any prescription medication containing opioids, or any illicit opioid drug, and cardiovascular disease (CVD). We conducted a keyword search and a MeSH term search of the PubMed database for articles published on or before September 2, 2019. The keyword search included the following strings and logic: (("Heart Failure" OR Endocarditis OR "Myocardial infarction" OR "Atrial fibrillation" OR "cardiac arrhythmia" OR "myocardial ischemia" or "coronary heart disease" or "cardiac arrest" or “stroke” or "coronary artery disease")) AND Opioid AND epidemiology. For the MeSH term search, the strings and logic were: (("Analgesics, Opioid"[Majr] OR "Opioid-Related

Disorders/epidemiology"[Majr]) AND ("Cardiovascular Diseases/epidemiology"[Majr] OR "Stroke/epidemiology" [Majr])). We included the term "epidemiology" in both key word and MeSH term searches to exclude basic science and non-human studies. Additional articles were identified from the reference lists of retrieved articles.

2.3.1 Inclusion and exclusion criteria

We included original articles that investigated the association of NOU with one or more of the following five cardiovascular outcomes: infective endocarditis (IE); coronary heart disease, including myocardial infarction (MI); congestive heart failure (CHF); cardiac arrythmia, including cardiac arrest; and stroke. We excluded studies that lacked an appropriate comparison group. In most cases, this meant individuals who did not experience NOU. In the case of endocarditis, it meant either individuals who did not inject opioids, or a time period during which injection opioid use was expected to be substantially lower due to a policy change. The details of inclusion and exclusion are listed in Table 2.1.

Table 2. 1 Inclusion and exclusion criteria for studies

Inclusion Criteria	Exclusion Criteria
• Investigates the relationship between a history of long-term (medical or nonmedical) exposure to opioids (prescription or illicit) and subsequent cardiovascular disease • Investigates the relationship between history of opioid use and cardiovascular outcomes following a (not necessarily cardiovascular) medical procedure (e.g. kidney transplant, orthopedic surgery, etc.)	• Basic science studies • Non-human studies • Case study or case series • Review articles • Non-original research including editorials, letters and protocols • Short term opioid use • Studies of the effect of brief exposures to opioid medications (e.g. for analgesia or anesthesia related to a surgical procedure or other medical events) • Studies of the relationship between opioid use and any cardiovascular events other than those of interest

One author (JS) reviewed the entire list of identified references, while two authors (AKN and ELA) each reviewed a mutually exclusive half of the references. Disagreement in the classification of records by the two independent reviewers was adjudicated by group consensus. A flow diagram summarizes article selection procedures (Figure 2.1).

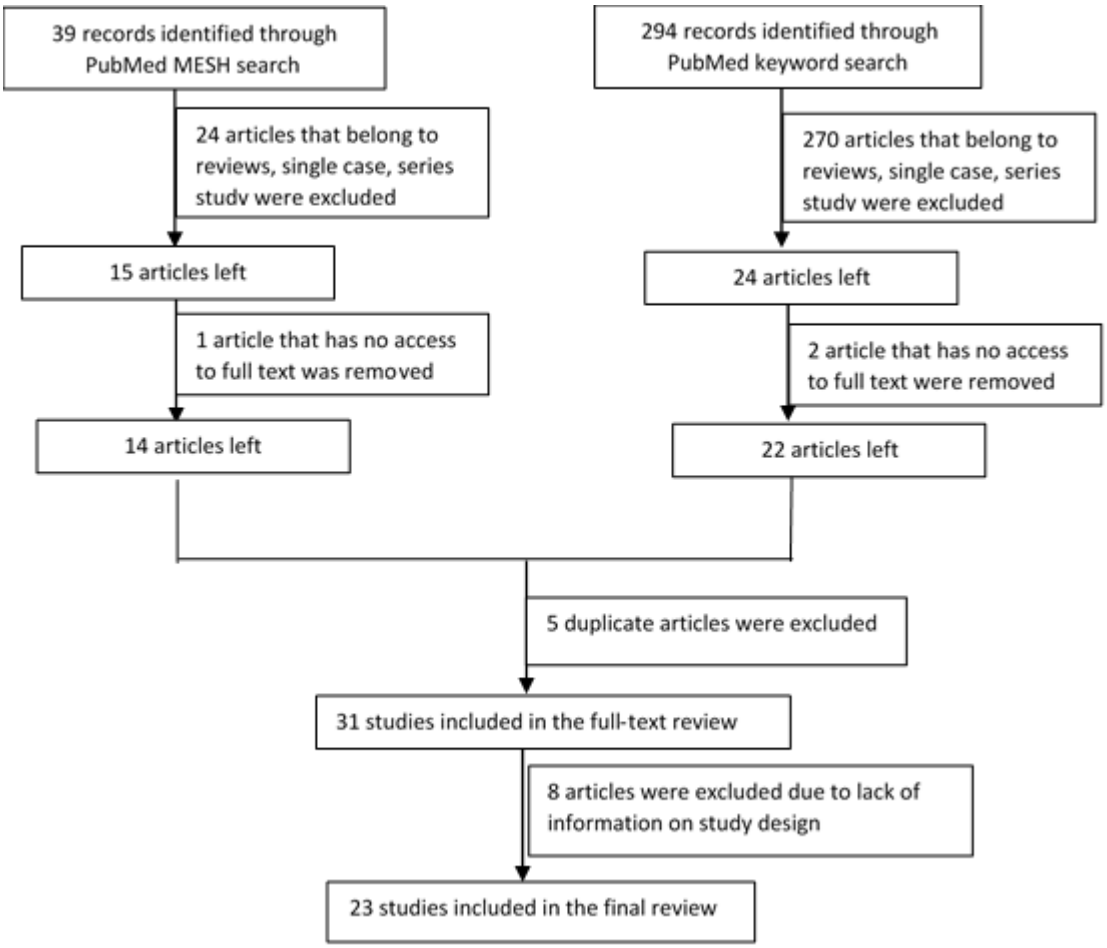


Figure 2. 1 Flow diagram of included studies

2.4 Results

A total of 39 articles were identified from the MeSH term search and 294 articles from the keyword search. After excluding reviews and case series, articles that did not address NOU or any of our outcomes of interest, those that did not include a comparison group, and those for which full text could not be retrieved; and after and resolving duplicates that were retrieved through both search protocols, 23 studies remained for review. Fourteen articles were from the U.S., 2 from Iran, 2 from Canada and 1 each from of the following countries: United Kingdom, Italy, Germany, Spain, Taiwan. There were 10 cohort studies (Carman et al., 2011; Jain et al., 2008; Keeshin and Feinberg, 2016; Lentine et al., 2015; Marmor et al., 2004; Meisner et al., 2019; Menendez et al., 2015; Mirzaiepour, 2012; Omran et al., 2019; Solomon et al., 2010), 5 case-control (Jobski et al., 2017; Lee et al., 2013; Li et al., 2013; Porter and Jick, 1980; Roberto et al., 2015), 3 cross-sectional (Khodneva et al., 2016; Sadeghian et al., 2009; Vozoris et al., 2017), and 5 trend analyses (Bates et al., 2019; Gray et al., 2018; Hartman et al., 2016; Lewer et al., 2017; Weir et al., 2019) (Table 2). All included studies utilized retrospective designs.

Table 2. 2. Selected major studies of the association of nonacute opioid use with Endocarditis, MI, CHF, Arrhythmia, and Stroke

Study	Study design	Year frame	N	Exposure	Outcome	Effect Estimate
Bates, 2019(Bates et al., 2019), United States	RTA	2008-2015	462	Illicit drug use	Endocarditis	Relative Risk increase = 0.06%, P=0.001
Carman, 2011(Carman et al., 2011), United States	RCS	2002-2005	148,657	Overall chronic opioid [I] therapy for non-malignant pain	Myocardial infarction	IRR = 2.66, 95% CI 2.3 - 3.08
				Low dose chronic opioid therapy for non-malignant pain	Myocardial infarction	IRR = 1.21, 95% CI 1.02 - 1.45
				High dose chronic opioid therapy for non-malignant pain	Myocardial infarction	IRR = 1.89, 95% CI 1.54 - 2.33
Gray, 2018(Gray et al., 2018), United States	RTA	2000-2016	510	Injection drug use (no mention of opioid name)	Endocarditis	Prevalence ratio of IDU per year = 1.09, 95% CI: 1.05–1.14
Hartman, 2016(Hartman et al., 2016), United States	RTA	2009-2014	127	Injection drug use [II]	Endocarditis	Percentage of endocarditis increases from 14% in 2009 to 56% in 2014
Jain, 2008(Jain et al., 2008), United States	RCS	1996-2003	238	Injection drug use [III]	Tricuspid valve (TV) IE	OR 4.37, p=0.001
				Injection drug use	mitral valve (MV) IE	OR 4.37, p=0.001
				Heroin	Tricuspid valve (TV) IE	OR 4.37, p=0.001
	CC	2006-2011	309,936			OR = 1.17,

Jobski, 2017 (Jobski et al., 2017), Germany				Current or recent use of ER-HPO[IV] (referent: past use)	Myocardial infarction	95% CI 1.09 - 1.26
				Recent discontinuation of any ER-HPO (referent: past use)	Myocardial infarction	OR = 1.11, 95% CI 0.98 - 1.25
				Recent switch of substance (referent: past use)	Myocardial infarction	OR = 1.38, 95% CI 1.02 - 1.86
				Current or recent use of ER-HPO (referent: past use)	Stroke	OR = 0.95, 95% CI 0.88 - 1.02
				Recent discontinuation of any ER-HPO (referent: past use)	Stroke	OR = 1.14, 95% CI 1.02 - 1.27
				Recent switch of substance (referent: past use)	Stroke	OR = 1.19, 95% CI 0.89 - 1.58
Keeshin, 2016 (Keeshin and Feinberg, 2016), United States	RCS	1999-2009	392	Injection opioid use [V]	Endocarditis	Endocarditis cases increase 2-fold
				Injection opioid use	HCV antibody prevalence	HCV antibody prevalence increase 3-fold
				Injection opioid use	Positive opiate toxicology screens	Positive opiate toxicology screens increase 6-fold
Khodneva, 2016 (Khodneva et al., 2016), United States	CS	2003-2007	29,025	Prescription opioid [VI] use for nonmalignant chronic pain	Stroke	HR = 1.04, 95% CI 0.78 - 1.38
				Prescription opioid use for nonmalignant chronic pain	Coronary artery disease in all	HR = 1.03, 95% CI 0.83 - 1.26
				Prescription opioid use for nonmalignant chronic pain	Coronary artery disease in female	HR = 1.38, 95% CI 1.05 - 1.82
				Prescription opioid use for nonmalignant chronic pain	Coronary artery disease in male	HR = 0.7, 95% CI 0.5 - 0.97
Lee, 2013(Lee et al., 2013),	CC	1998-2010	6,040	Treatment with morphine for all cancer-related pain	Stroke: All stroke	OR = 1.13, 95% CI 0.97 - 1.31

Taiwan				Treatment with morphine for prostate cancer-related pain	Stroke: All stroke	OR = 3.02, 95% CI 1.68 - 5.42
				Treatment with morphine for prostate cancer-related pain	Stroke: Hemorrhagic	OR = 4.24, 95% CI 1.03 - 17.4
				Treatment with morphine for prostate cancer-related pain	Stroke: Ischemic	OR = 2.9, 95% CI 1.58 - 5.35
Lentine, 2015 (Lentine et al., 2015), United States	RCS	2006-2008	16,322	Pre-transplant prescription narcotic use with living donor	Ventricular arrhythmia	HR = 1.38, 95% CI 0.14 - 13.42
				Pre-transplant prescription narcotic use with deceased donor	Ventricular arrhythmia	HR = 5.58, 95% CI 2.19 - 14.21
				Pre-transplant prescription narcotic use with living donor	Cardiac arrest	HR = 1.83, 95% CI 0.94 - 3.54
				Pre-transplant prescription narcotic use with deceases donor	Cardiac arrest	HR = 1.31, 95% CI 0.85 - 2.01
Lewer, 2017 (Lewer et al., 2017), UK	RTA	1997-2016	1,052,444	Injection opioid use	Endocarditis	Hospital admissions for infections related to injection drug use increased annually from 2012 to 2016.
Li, 2013(Li et al., 2013), United States	CC	1990-2008	56,590	Any opioid prescription [VII] at current use (<=30 days)	Myocardial infarction	OR = 1.28, 95% CI 1.19 - 1.37
				Any opioid prescription cumulative use 11 to 50 Rx	Myocardial infarction	OR = 1.38, 95% CI 1.28 - 1.49
				Any opioid prescription cumulative use > 50 Rx	Myocardial infarction	OR = 1.25, 95% CI 1.11 - 1.4
				Buprenorphine prescription	Myocardial infarction	OR = 1.71, 95% CI 1.09 - 2.68

				Morphine prescription	Myocardial infarction	OR = 2.15, 95% CI 1.24 - 3.74
				Meperidine prescription	Myocardial infarction	OR = 1.46, 95% CI 1.22 - 1.76
Marmor, 2004 (Marmor et al., 2004), United States	RCS	1998	98	Serologic evidence of methadone or opiates (as proxy for long-term exposure to opioids)	Coronary artery disease	OR = 0.43, 95% CI 0.2 - 0.94
Meisner, 2019(Meisner et al., 2019), United States	RCS	2013-2017	1,921	Injection drug use	Endocarditis	Endocarditis percentage change = 238%
Menendez, 2015 (Menendez et al., 2015), United States	RCS	2002-2011	9,307,348	Opioid use disorder (opioid abuse or dependence)	Myocardial infarction	OR = 1.9, 95% CI 1.3 - 2.6
Mirzaiepour, 2012 (Mirzaiepour, 2012), Iran	RCS	2010-2011	200	Opium addiction (as defined by DSM-IV criteria for substance dependence)	Arrhythmia and cardiac arrest	OR = 21.9, 95% CI 9.58 - 50.01
Omran, 2019 (Omran et al., 2019), United States	RCS	1993-2015	5,283	Opioid	Stroke	Stroke percentage change from 1993 to 2008 = 1.9%, 95% CI -2.2% - 6.1%
					Stroke	PCT change from 2008 to 2015 = 20.3%, 95% CI 10.5% - 30.9%
Pontes, 2018(Porter and Jick, 1980),	CC	2008-2012	22,652	Opioid analgesic therapy for OA-related pain	Myocardial infarction	OR = 1.13, 95% CI 1.03 - 1.24

Spain						
Roberto, 2015(Roberto et al., 2015), Italy	CC	2002-2012	12,483	Current use of acetaminophen or/and an acetaminophen-codeine combination (0-90 days) (referent = nonuse, defined as more than 365 days since last use)	Myocardial infarction	OR = 1.22, 95% CI 0.92 - 1.63
				Recent use	Myocardial infarction	OR = 1.12, 95% CI 0.8 - 1.55
				Past use	Myocardial infarction	OR = 1.13, 95% CI 0.86 - 1.48
Sadeghian, 2009 (Sadeghian et al., 2009), Iran	CS	2006-2008	4,398	Opioid dependence (according to DSM-IV criteria)	Myocardial infarction	RR = 0.34, 95% CI 0.02 - 3.23
				Opioid dependence (according to DSM-IV criteria)	Arrhythmia and cardiac arrest	RR = 0.65, 95% CI 0.43 - 1.03
Solomon, 2010(Solomon et al., 2010), United States	RCS	1996-2005	31,375	Incident opioid therapy [VIII] for nonmalignant pain	Nonspecific	RR = 1.62, 95% CI 1.27 - 2.06
Vozoris, 2017 (Vozoris et al., 2017), Canada	CS	2008-2013	149,094	Incident opioid use [IX]	Heart failure	HR = 0.84, 95% CI 0.73 - 0.97
				Incident opioid use	Coronary artery disease	HR = 2.15, 95% CI 1.5 - 3.09

Weir, 2019(Weir et al., 2019), Canada	RTA	2006-2015	60,529	Intervention: removal of controlled-release oxycodone from Canadian market (2011 Q4)	Endocarditis	No quantitative estimates provided

Abbreviations in Table 2.2:

CC: Case control;

CS: Cross sectional;

RCS: Retrospective cohort study;

RTA: Retrospective trend analysis;

TV: Tricuspid valve;

MV: Mitral valve;

IRR: Incidence rate ratios;

OR: Odds Ratio;

RR: Relative risk;

HR: Hazard ratio;

Cl: Confidence Interval;

HCV: Hepatitis C virus;

ER-HPO: Extended-release high-potency opioid;

PCT: Percentage;

DSM-IV: Diagnostic and Statistical Manual - 4.

Notes: Opioid names listed in Exposure in Table 2 with roman numerals are listed in Supplemental Table 1.

2.4.1 Myocardial infarction (MI)

Of the 10 studies that reported on the association of opioid use with MI, four used data from retrospective cohorts (Carman et al., 2011; Khodneva et al., 2016; Sadeghian et al., 2009; Vozoris et al., 2017) , 2 were cross-sectional (Marmor et al., 2004; Menendez et al., 2015) , and 4 were nested case-control studies (Jobski et al., 2017; Li et al., 2013; Porter and Jick, 1980; Roberto et al., 2015). Among older adults with chronic obstructive pulmonary disease, when restricting to opioid-only formulation, positive associations were observed for CAD-related mortality: HR=1.83, 95% CI (1.32-2.53), and for CAD-related ER visit or hospital admission: HR=1.38, 95% CI (1.08-1.77). Carman et al. reported a positive association between chronic opioid therapy for nonmalignant pain and MI incidence in a commercially insured cohort, with greater risk observed at higher doses [Incidence Rate Ratio (IRR) = 2.66, 95% CI (2.30-3.08)]. Within a group of patients who underwent coronary artery bypass (CABG) (Sadeghian et al., 2009), a relative risk of 0.34 [95% CI (0.02-2.32)], was reported for perioperative MI, among patients with preoperative opioid dependence. However, the extremely wide CI indicates the possibility of a sparse data bias.

In a case-control study of primary care patients with physician-diagnosed osteoarthritis (Baclic et al., 2020), Pontes et al. reported a positive association between the odds of MI and use of opioid analgesics for treatment of OA [OR=1.13, 95% CI (1.03-1.24)], with odds increasing as the mean monthly dose of opioids increased. Jobski et al. reported associations with recent (within 30 days of index MI) discontinuation of extended-release high-potency opioid (ER-HPO) therapy [OR=1.11, 95% CI (0.98-1.25)] and recent (within 30 days of index MI) switch [OR=1.38, 95% CI (1.02-1.86)] of ER-

HPO medication type (Jobski et al., 2017). Within a group of general practice patients with OA, Roberto et al. reported no statistically significant association with acetaminophen-codeine for treatment of OA pain [OR=1.22, 95% CI (0.92-1.63)] (Roberto et al., 2015). Li et al. 31 reported a positive association between MI and current opioid use [OR=1.28, 95% CI (1.19-1.37)], two-year cumulative prior use consisting of 11 to 50 prescriptions [OR=1.38, 95% CI (1.28-1.49)], and two-year cumulative prior use consisting of more than 50 prescriptions [OR=1.25, 95% CI (1.11-1.40)] (Li et al., 2013).

Findings from Marmor's cross-sectional study of serological evidence of methadone or opioid use at autopsy, and its relationship to coronary artery plaque, suggest a protective effect with respect to CAD [OR 0.43, 95% CI (0.20-0.94)] (Marmor et al., 2004). However, the study did not provide information on duration of opioid use or of methadone treatment of opioid addiction. Among hospital in-patients undergoing major elective orthopedic surgery, Menendez et al. reported a positive association between preoperative opioid abuse or dependence and in-hospital MI [OR=1.90, 95% CI (1.30-2.60)] (Menendez et al., 2015).

Conversely, data from the REGARDS study suggest no association overall between prescription opioid use for nonmalignant chronic pain and coronary heart disease (CHD) over the course of four to seven years of follow-up [HR=1.03, 95% CI (0.83-1.26)]. In an analysis stratified by gender, the authors report a modest increase in CHD risk among women [HR=1.38, 95% CI (1.05-1.82)], but a decrease in risk among men [HR 0.70, 95% CI (0.50-0.97)] with evidence of opioid use (Khodneva et al., 2016).

2.4.2 Heart failure

There were no studies of NOU and heart failure identified in our search.

2.4.3 Arrhythmia

Of the three studies that reported on the association of opioid use with arrhythmia, two used data from retrospective cohorts (Lentine et al., 2015; Sadeghian et al., 2009) and one was a cross-sectional study (Mirzaiepour, 2012). In a cohort of hospital patients who underwent CABG surgery, Sadeghian et al. reported a protective association between atrial fibrillation and opium addiction that was not statistically significant [OR=0.65, 95% CI (0.43-1.03)] (Sadeghian et al., 2009). In hospital patients admitted with acute MI, Mirzaiepour et al. reported a strong, positive association between post-MI arrhythmia and opium addiction [OR=21.9, 95% CI (9.58-50.0)] (Mirzaiepour, 2012). In a cohort of kidney transplant patients, Lentine et al. reported a positive association between ventricular arrhythmia and pre-transplant opioid use at a dose greater than 23.8 mg/kg morphine equivalents [HR=5.58, 95% CI (2.19-14.21)] (Lentine et al., 2015).

2.4.4 Stroke

Of the four studies reporting on the association of opioid use and ischemic or hemorrhagic stroke, two used data from a retrospective cohort (Khodneva et al., 2016; Omran et al., 2019) and two were nested case-control studies (Jobski et al., 2017; Li et al., 2013). Jobski et al. reported an association with recent (within 30 days of index MI) discontinuation of extended-release high-potency opioid therapy [OR=1.14, 95% CI (1.02-1.27)] (Jobski et al., 2017). No association was reported for current opioid use or

recent switch of opioid type. Omran et al. reported percentage of stroke among hospitalization changed 20.3% [95% CI 10.5% - 30.9%] from 2008 to 2015 with the combination of opioid abuse (Omran et al., 2019). Khodneva et al. reported no association between prescription opioid use for nonmalignant chronic pain and stroke over the course of median 42 of 5.2 (1.8) years of follow-up in the REGARDS study [HR=1.04, 95% CI (0.78-1.38)] (Khodneva et al., 2016). Lee et al. reported a positive association between morphine use for cancer-related pain and hemorrhagic stroke [OR=1.36, 95% CI (1.02-1.82)] but not ischemic stroke [OR=1.08, 95% CI (0.92-1.27)] (Lee et al., 2014). When restricting to prostate cancer patients only, the association with hemorrhagic stroke was higher [OR=4.24, 95% CI (1.03-17.4)], and a significant association with ischemic stroke was reported [OR=2.90, 95% CI (1.58-5.35)].

2.4.5 Infective endocarditis

Seven studies investigated the association between opioid use and infective endocarditis (IE). Five of these were trend analyses, of which four reported a temporal association between injection drug use (IDU) and IE (Gray et al., 2018; Hartman et al., 2016; Lewer et al., 2017; Meisner et al., 2019; Weir et al., 2019), and one reported a temporal association between mixed drug use and IE (Jain et al., 2008). Jain and colleagues reported an association between IDU and tricuspid valve IE (Jain et al., 2008). Keeshin and colleagues suggested that increases in hospital admissions for IE may provide an indirect surveillance marker for IDU within the surrounding community (Keeshin and Feinberg, 2016).

2.5 Discussion

There has been a growing interest in the possible cardiovascular effects of opioid drugs. Khodneva et al. described self-reported, baseline cardiovascular disease (CVD) in a cohort of community-dwelling adults consisting of 1,851 participants with prescription opioid use 26 and 27,174 nonusers (Khodneva et al., 2016). They found that coronary heart disease (22.8% vs. 17.4%), stroke (13.2% vs. 8.5%) and QTc prolongation (3.3% vs. 2.8%) were more commonly reported by participants with POU. Studies have investigated the link between methadone treatment for OUD and elongation of the QT interval/torsade de pointes (TdP), which can lead to cardiac arrhythmias and cardiac arrest (Barkin et al., 2006; Keller et al., 2016b). Solomon et al. reported different relative risk of cardiac events after the start of different opioid therapy(Solomon et al., 2010). Moreover, several studies have reported a small or moderate increase in the risk of myocardial infarction (MI) in persons with chronic exposure to opioids, due to abuse/dependence or long-term opioid therapy for chronic pain (Carman et al., 2011; Menendez et al., 2015; Sen et al., 2019). Conversely, it has been suggested that long-term opiate exposure may mitigate the severity of coronary artery disease (Marmor et al., 2004).

We set out to summarize systematically previous research on the association between NOU and five CVD outcomes. The amount and strength of the evidence varied across the outcomes. The most commonly studied outcomes were MI (10 reports) and IE (7 reports). Across studies included in this review, was generally consistent evidence of a positive association between community prevalence of injection drug use (with opioids being the most commonly injected type of drug) and community prevalence of IE, and

between nonacute opioid exposure (primarily for medical reasons) and MI incidence. The other four outcomes were less commonly studied (three reports each for CAD, arrhythmia and stroke; one for heart failure), and there was less consensus about their relationship with opioid use. Many of the studies, for all outcomes, lacked detailed information on the duration and dose of opioid exposure. Several studies have reported a temporal association between the prevalence of IDU and the prevalence of IE in a community, suggesting an increase in the prevalence of IE with increasing prevalence of IDU. The sharing of needles and other materials promotes the spread of microbial infections, with IE cases frequently resulting from Staphylococcal infection (Mylonakis and Calderwood, 2001) . This, together with the fact that prescription opioids and heroin are among the most commonly injected drugs (Cicero et al., 2017). Thus, increasing exposure to opioids in a population can lead to greater prevalence of IE, by increasing the prevalence of IDU within that population.

Coronary artery disease is the most common cause of MI, but is only directly observable by invasive procedures, such as cardiac catheterization or coronary angiogram, or at autopsy. This may explain why we identified 7 studies with MI as the endpoint, but only 3 with CAD. Only 4 studies described detailed assessment of dose and duration of opioid exposure, and all of these studies reported an association between opioid use and MI (Carman et al., 2011; Jobski et al., 2017; Li et al., 2013; Roberto et al., 2015).

Cardiac opioid receptors have been identified 48, but possible biological pathways between nonacute opioid use and MI or CAD are still not well understood. Li et al. (2013) speculated about possible relationships between opioids, hormones, including

testosterone, and coronary atheroma, but their study did not explore these connections (Li et al., 2013). Although some studies, such as that of Tanaka , attempt to address from a molecular perspective the role of endogenous and exogenous opioids and cardiac opioid receptors in limiting cardiac damage in acute MI patients (ischemic preconditioning, opioid-induced cardio protection) (Tanaka et al., 2014), our findings suggest that long-term opioid exposure is associated with an increase in the incidence of acute MI.

In a systematic review of opioid use and arrhythmia Behzadi reported that, some opioids, including methadone, tramadol and oxycodone, are associated with increased risk of long QT syndrome, which in turn may lead to dangerous arrhythmias including TdP (Behzadi et al., 2018). While arrhythmia was one of the cardiovascular conditions included in our review, we found only 3 studies of the relationship between NOU and cardiac arrhythmia that met our inclusion criteria. Our initial query returned a number of articles on opioids and the QT interval, which, upon review, turned out to lack an appropriate control group. As a result, those studies were excluded. Moreover, we excluded studies of arrhythmias associated with acute opioid exposure as in, for example, studies conducted among surgery patients or acute MI patients. Thus, although there is a body of evidence linking use of certain opioids with the long QT syndrome, we found little high-quality, epidemiological evidence examining directly the association of NOU with cardiac arrhythmias per se. This appears to be a gap in need of future attention.

We found no studies that addressed the association of NUO with the risk of stroke or heart failure in a general cohort that included a reliable assessment of dose and duration of opioid use. The identified studies had one or more serious limitations, including highly selective cohorts, a primary focus on short-term exposures, such as

recent use, change of medication, discontinuation of medication, or inadequate assessment of dose and duration. There remains a need for high-quality studies examining the relationship between nonacute opioid use and stroke and CHF.

Much of the research on opioids and CVD has focused on acute exposures related to surgical procedures or other treatment for acute medical conditions. Examples include opioids used for anesthesia during surgery or for post-operative analgesia, and morphine as part of treatment for acute MI. There has been relatively little high-quality research on nonacute opioid exposure and its relationship with cardiovascular conditions. A significant challenge for this type of research is the accurate assessment of the duration and intensity of opioid exposure over an extended period of time. For example, estimated that the period between the appearance of major risk factors for CAD – high serum cholesterol and high systolic blood pressure – and their effects on morality may be 10 years or more (Rose, 2005). Exposure to prescription opioids is well-documented in administrative claims databases, but members may be lost to follow-up if they change insurance plans. Moreover, exposure to nonmedical use of opioids is practically impossible to assess through secondary data sources.

2.6 Conclusion

In conclusion, this review of the literature on the association of nonacute opioid use with the risk of cardiovascular events provides summative evidence that such exposure poses a risk not only for cardiac disorders associated with infections due to needle re-use, such as infective endocarditis, but may also predispose persons to chronic cardiovascular disorders, including MI and arrhythmias. There is a dearth of high-quality

evidence on the relationship between NOU and CVD. Many of the identified studies lacked detailed information on the duration and intensity of opioid exposure and all were retrospectively conducted. This is understandable, as the challenges to accurate assessment of NOU are considerable. Innovative approaches to opioid exposure assessment will be required.

CHAPTER 3. USING NATURAL LANGUAGE PROCESSING TO IDENTIFY OPIOID USE DISORDER IN HOSPITAL AND EMERGENCY DEPARTMENT ELECTRONIC HEALTH RECORD DATA

3.1 Abstract

As opioid prescriptions have risen, there has also been a rise in opioid use disorder (OUD) and its related health issues and death, but epidemiologic surveillance is difficult. Electronic health records (EHR) are one potential source of surveillance data, but the nature of EHR data presents challenges [Howell, 2020]. The objective of this study was to ascertain prevalence of OUD using two methods to identify OUD: applying natural language processing (NLP) on unstructured clinical notes to identify OUD and using ICD-10-CM diagnostic codes to identify OUD. Data were drawn from EHR information on hospital and emergency department patient visits to a large regional academic medical center from 2017 to 2019. EHR corresponding to 50 patient visits from 2017 and 30,124 patient visits from 2018 were used to develop and evaluate an NLP-based algorithm and 29,212 patients visits from 2019 to testing the algorithm. International Classification of Disease, tenth Edition, Clinic Modification (ICD-10-CM) diagnostic codes were extracted for each visit. We developed five unique dictionaries and six parsing rules for physician mentions of opioid use disorder, and an NLP algorithm to identify these terms in unstructured clinical text. To confirm the validity of the NLP results, physician on manually reviewed randomly selected records. Prevalence of OUD was higher according to NLP classification vs. ICD-10-CM codes. Based only on information contained in ICD-10-CM codes, 1,811 patient visits were identified with an OUD diagnosis among the 29,212 visits (6.1%). In contrast, the NLP algorithm identified 1,902 (6.5%) visits with an OUD classification. We estimated the sensitivity and specificity of EHR-based OUD

classification at 81.8%%, 97.5% respectively, relative to manual record review. NLP-based algorithms can automate extraction and identify evidence of opioid use disorders from unstructured electronic healthcare records. Combining with ICD-10-CM codes, More OUD cases can be identified. It also will make the unstructured EHR notes useable for researchers to do epidemiological study of OUD.

3.2 Background

Electronic health records are a rich source of data that can be leveraged to inform strategies for measuring and addressing the ongoing opioid crisis in the United States (Smart et al., 2020; Wang et al., 2018). Accurate and timely identification of patients with opioid use disorder (OUD) is an important step in any such effort. International Classification of Diseases (ICD) codes are commonly used for this purpose due to their widespread use in medical record coding and their accessibility to researchers (Beam et al., 2021; Mezzich, 2002). The limitations of ICD codes, including low sensitivity and specificity for many conditions, have been well-documented (Hughes Garza et al., 2021; Kurbasic et al., 2008; O'Malley et al., 2005; Quan et al., 2008), although there has been little investigation of the extent to which this may pertain to the identification of OUD cases(Palumbo et al., 2020).

In addition to ICD codes, EHR's contain substantial information in the form of unstructured, narrative text entered by providers, nurse, lab technician or any other member of patient's healthcare team as notes in the course of treatment (Spasic and Nenadic, 2020). These clinical notes include information on patient symptoms, conditions, behaviors, healthcare advice and plans, and more (Wang et al., 2018). Generally, information in unstructured notes include demographics, medical encounters, developmental history,

obstetric history, medications and medical allergies, family history, social history, habits, immunization records (Gliklich et al., 2019).

Natural language processing (NLP) holds great promise as a tool for recovering valuable information from unstructured textual data in many domains (Pendergrass and Crawford, 2019). NLP is a branch of artificial intelligence that is concerned with computer understanding of human languages (Baclic et al., 2020). Information extraction is a subtask of NLP that is focused on the extraction of structured data from text (Ford et al., 2016; Meystre et al., 2008). Typically, information extraction involves splitting text into basic units called tokens, which are individual words, punctuation marks, etc. (Ford et al., 2016). Rule-based approaches to information extraction attempt to identify matches of pre-specified sequences of tokens (Nadkarni et al., 2011). Statistical approaches rely on probabilistic models, or on supervised learning methods applied to very large corpuses of text that have been labeled to indicate which instances do, and do not, contain the information of interest (Carrell et al., 2015). In supervised learning approaches, systems can be trained to recognize entities within text documents by seeing many correctly-labelled examples and "learning" features of the text that accurately predict the presence of those entities (Spasic and Nenadic, 2020). Although powerful, a limitation of this approach is the enormous effort required to create a sufficiently large, prelabeled corpus of examples (Nadkarni et al., 2011; Velupillai et al., 2018).

Several recent reviews have demonstrated the potential for using NLP to extract clinical information from EHR provider notes (Singleton et al., 2021). Wang et al. reported on a sample of 135 peer-reviewed articles focused on identifying diseases or conditions from clinical notes in EHR systems ("disease phenotyping") (Wang et al., 2018). The most

commonly studied diseases were cancer, venous thromboembolism, peripheral arterial disease, and diabetes mellitus (Afzal et al., 2017; Datta et al., 2019; Woller et al., 2021; Zheng et al., 2016). Many of the studies reported high sensitivity and specificity. Sheikhalishahi et al. identified 106 peer-reviewed articles focused on the application of NLP to free-text clinical notes related to chronic diseases, with the aim of transforming clinical text into structured clinical data (Sheikhalishahi et al., 2019). The most commonly studied chronic disease areas in this review were similar to those identified by Wang et al: diseases of the circulatory system, neoplasms, and diabetes mellitus. It is not immediately obvious why these particular conditions have been the most commonly studied.

Carrell and co-workers investigated the potential to apply NLP to EHR records to increase the identification of problem use of prescription opioids (POU) among patients undergoing chronic opioid therapy (Carrell et al., 2015). POU was defined as indications of addiction, abuse, misuse or overuse, and is thus more broadly defined and less specific than OUD. The study documented POU between 2006 and 2012 in a sample of 22,142 patients who received chronic opioid therapy, defined as at least a 70-day supply of prescription opioid medications dispensed in a calendar quarter, within a large health plan serving the state of Washington. They used a rule-based approach to identify "mentions" of POU in clinical notes from patient's EHR records, such as phrases of the form "opioid addiction", "dependence of methadone", or "no evidence of drug abuse". Candidate mentions were then manually validated by trained reviewers. POU prevalence increased under the NLP approach: POU prevalence was 10.1% based on ICD-9-CM codes alone, and 13.4% including patients identified by NLP.

Many patients with OUD who attend EDs and hospitals will not be undergoing chronic opioid treatment (COT), and so the effectiveness of NLP approaches in broader patient populations is unknown (Kaye et al., 2017). And, as noted, Carrell et al. used ICD-9-CM-coded patient records (Carrell et al., 2015). In 2015, the United States transitioned to ICD-10-CM for medical coding (Register, 2014). Thus, we aimed to conduct a study of the application of NLP to OUD ascertainment in a general patient population where ICD-10-CM codes are used. Drawing on inpatient and emergency department EHR records from the University of Kentucky Albert B. Chandler Hospital (spanning 2017-2019), we investigated the performance of NLP algorithms, relative to ICD-10-CM codes, in the identifications of OUD cases.

3.3 Methods

3.3.1 Study Population

Data were drawn from all adult (age 18 years or older) inpatient and ED visits occurring at the University of Kentucky HealthCare (UKHC) Albert B. Chandler Hospital between January 1st, 2017 and December 31st, 2019. Due to high prevalence of opioid use for the treatment of cancer-related pain (Wong and Cheung, 2020), we excluded visits for patients with active cancer (ICD-10-CM code: C00-C27, C30-C42, C43-C59, C60-C81, C7A.*, C7B.*, C81-C97, D37-D50) (Neoplasms, 2021). Additionally, we required that patient visits had at least one of the following five types of notes, which we considered most likely to include information pertaining to opioid use disorder: ED triage notes, ED general notes, History and Physical notes, Addiction Medicine Consult notes, and Discharge Summary notes. Table 3. 1 describes the application of inclusion and exclusion

criteria. Care delivery was documented through Sunrise Clinical Manager (SCM) for inpatient stays and emergency department (ED) visits. In addition to unstructured provider notes, structured EHR data documenting delivered care includes information on patient demographics, diagnoses, and problem lists. This study was approved by both the UK Institutional Review Board (IRB# 20548) and the UKHC Data Management Committee.

3.3.2 ICD-10 definition of OUD

The ICD-10-CM definition for OUD included the codes for opioid abuse (F11.10, F11.11, F11.12, F11.14, F11.18, F11.19), opioid dependence (F11.20, F11.21, F11.22, F11.23, F11.24, F11.25, F11.28, F11.29), and unspecified opioid use (F11.90, F11.92, F11.93, F11.94, F11.95, F11.98, F11.99). For each patient visit in the study sample, the encounter was classified as positive for OUD if any of these diagnosis codes were present, otherwise non-OUD.

3.3.3 NLP-based definition of OUD

3.3.3.1 Overview of algorithm development process

The NLP algorithm was developed in following phases—dictionary development (Phase 1), parsing rule and algorithm development (Phase 2), and final classification (Phase 3). First, we used information from extant literature to create dictionaries of OUD-related terms. The dictionaries were refined based on advice from an expert in medical toxicology and emergency medicine (co-author PDA), and manual review of 25 randomly selected patient visits occurring in 2017 with OUD identified by ICD-10-CM and 25 randomly selected patient visits occurring in 2017 without OUD identification by ICD-10-CM. The

developed dictionaries were used to create parsing rules to search evidence that the patient visit did or did not indicate that a mention of OUD was appropriate.

We used data from patient visits occurring in 2018 to develop the algorithm. We applied the initial version of the algorithm to the data to classify each encounter as OUD or non-OUD. Next, to evaluate the performance of algorithm, we selected a 1% random sample each from the visits classified as OUD and from the visits classified as non-OUD for review. An expert clinician (co-author PDA) independently reviewed the EHR records for these cases and classified them as OUD or non-OUD, without knowledge the algorithmic classification. The conditions that were taken as evidence of OUD when manually reviewing the cases were refined from a list reported in Carrell, and are listed in Table 3.1 (Carrell et al., 2015). Based on findings from the manual review, we updated the dictionaries and revised the protocol pipeline to optimize performance. The finalized algorithm was then applied to the data set consisting of visits occurring in 2019.

Table 3. 1 Conditions indicating OUD

Condition indicating OUD
• Admits to opioid use • Recent inpatient admission for detox • Referral for opioid addiction treatment at the First Bridge Clinic • Currently receiving methadone or suboxone treatment for opioid addiction • Loss of control of opioid, craving • Family member reported opioid addiction to clinician • Current or recent opioid overdose • Obtained opioids for multiple MDs surreptitiously • Opioid taper/wean due to problems (not due to expected pain improvement) • Unsuccessful taper attempt • Physician or patient wants immediate taper • Positive response to Narcan treatment

3.3.3.2 Dictionaries

Five dictionaries were constructed (Table 3.2). We note that all dictionary terms are lower case because we transformed the EHR text data into all lower case to facilitate analysis. Based on published literature (Carrell 2015) and expert knowledge (author PDA), we created dictionaries for opioid types (dictionary 1; e.g., “morphine”, “narcotic”, “oxycodone”) and terms suggestive of use disorder (dictionary 2; e.g., “abuse”, “dependence”, “use disorder”). Next, we queried our training database (2018 patient visits) to refine these dictionaries. For example, we discovered spelling errors (e.g., “depandance” for “dependence”), commonly used abbreviations (e.g., “Sub” for “Subutex”, “OD” for “overdose”, “Vico” for “Vicodin”), and other types of nonstandard text. We created a third dictionary consisting of terms (e.g., “denies”, “without”, “negative”) that are used in clinical notes to negate a mention of OUD. The fourth dictionary was created to capture specialized terms used at UKHC related to OUD, which were identified during review of a random selected sample of 200 clinical notes from the training dataset visits that had an OUD diagnostic code. Those terms were found primarily in the social history and related to the negation of any drug use, e.g. “IVDA/intranasal: Denies”. The fifth dictionary included the name of the UKHC opioid treatment clinic where patients may be referred to following discharge (First Bridge Clinic)

Table 3. 2 Dictionaries of opioid, use disorder, and negation terms, and additional specialized terms, which were combined via parsing rules to form OUD search phrases

Dictionary	Key words
1. Opioid term	fentanyl, heroin, hydromorphone, dilaudid, oxymorphone, opanum, opana, methadone, oxycodone, oxycotin, roxicodone, percocet, morphine, hydrocodone, vicodin, vico, lortab, codeine, meperidine, demerol, tramadol, ultram, meloxicam, kratom, carfentanil, buprenorphine, meperidine, narcotic, dihydrocodeine, levorphanol, naloxone, naltrexone, pentazocine, suboxone, subutex, sub, tapentadol, vivitrol, opiate, opioid, opium
2. Use disorder terms	abuse, abuses, abused, abusive, abusing, addict, addicts, addicting, addicted, addiction, dependence, dependant, dependance, dependency, misuse, misuses, misused, misusing, overdose, overdoses, overdoes, over dose, over dosed, od, over use, over used, overuse, use disorder, use-disorder, inject, injected, injects, injection, injecting, ivda, intravenous drug abuse, iv drug use, intravenous drug user, iv drug user, ivdu, intravenous drug abuse, iv drug abuse, iv drug abuse, iv drug abuser, withdrawal, withdraw, withdrew, withdrawling
3. Negation terms	absence, absent, deny, denies, denied, denying, do not, don't, donnot, exclude, excluded, excludes, excluding, lack, lacked, lacks, lacking, negative, negation, never, no, no evidence, did not have, no history, no hx, no sign, no signs, not observed, not present, without, without evidence, suspect, suspected
4. Specialized terms	See Supplement Table 2 for specialized term lists
5. Specific clinic	first bridge clinic, the bridge

3.3.3.3 Parsing rules

We developed six parsing rules representing combinations of dictionary terms (Table 3.3). The clinical notes had a consistent structure. Each note consisted of some number of sections that were separated into paragraphs, and each section started with a keyword follow by a colon, and then finished with a period. For example, a discharge summary note usually includes the following sections: reason for hospitalization, Significant findings, Procedures and treatment, Patient's discharge condition, Patient and family instructions, Attending physician's signature.

Notes were first parsed into sections. Sections were processed sequentially by first parsing the section into sentences, and then parsing each sentence into individual tokens – words, sub-words, abbreviations, punctuation marks, etc.

Parsing rule 1 consists of an opioid term (dictionary 1), followed by zero to three other valid tokens, followed by a use disorder term (dictionary 2); for example, “opioid use disorder”, “Oxycodone dependence”, or “heroin overdose” would return a classification of positive for OUD. Parsing rule 2 consists of a problem use term (dictionary 2), followed by zero to three other valid tokens, followed by an opioid term (dictionary 1). Examples of parsed terms that would return a classification of positive for OUD are “dependent on opioids”, “addicted to heroin”, and “abuse with opioids”. Parsing rule 3 is a negation rule that consists of a negation term (dictionary 3), followed by zero to three other valid tokens, followed an opioid term (dictionary 1), followed by a problem use term (dictionary 2). Examples of parsed terms that would return a classification of negative for OUD are “negative for opioid abuse” and “denied opioids addiction”. Parsing rule 4 is also a negation rule that consists of an opioid term (dictionary 1), followed by a problem use term

(dictionary 2), followed by zero to three other valid tokens, followed by a negation term (dictionary 3). Examples of parsed terms that would return a classification of negative for OUD include “opioid dependence is denied” and “opioid misuse is negative”. Parsing rule 5 is a searching rule included all the specialized terms in dictionary 4. Parsing rule 6 is specific to UKHC and includes the first bridge clinic in dictionary 5.

Table 3. 3 Parsing rules defining the combinations of dictionary terms used in then identification of OUD

Parsing rules	Rule contracture	Example
Rule 1	Opioid term + <= 3 tokens + use disorder term	Opioid use disorder Opiate dependence
Rule 2	Problem use term + <= 3 valid tokens + opioid terms	Addicted to suboxone
Rule 3	Negation term + <=3 valid tokens + opioid terms + use disorder term	Denies opioid addiction
Rule 4	Opioid term + use disorder term + <=3 valid tokens + negation term	Opioid dependence is denied
Rule 5	Specialized terms (use dictionary 4)	IVDA/intranasal: Denies
Rule 6	Specific clinic (use dictionary 5)	First Bridge Clinic

3.3.3.4 Implementation

The finalized algorithm was applied to the 2019 patient visit data. The NLP algorithm was coded using the Python programming language and the Natural Language Toolkit (NLTK and Spacy) modules (Bird, 2009; Honnibal, 2017); the logic and pipeline is shown in Figure 3.1. Each EHR note is read by Python codes and parsed into sections. The search rules were implemented as regular expression searches. The text in each section was then converted to lowercase and scanned for terms in dictionaries 1 and 5. If a match was found, the section was further separated into sentences. Each sentence was then

scanned for mentions of OUD matching one of the negation rules (parsing rules 3, 4 and 5). If there was a match, a negative OUD mention record was created, which included the matching text, the parsing rule to which it matched, and the document name. If no matching terms were found, the next section was processed. Otherwise, the algorithm continued scanning the same sentence for matches to parsing rules 1, 2 and 6 (positive OUD mentions). If a match was found, an OUD mention record was created and continue to search whole document. One or multiple OUD mentions can be identified from one document. If no match was found, the next sentence was processed. In this way, all mentions of OUD appearing in a clinical note associated with a particular visit are extracted and classified as either positive or negative mentions of OUD.

3.3.3.5 Classification.

To classify each patient visit as “OUD” or “non-OUD”, we applied the following logic: if all mentions of OUD for a visit were positive, we classified it as OUD. If a visit had no mentions of OUD, or if all identified OUD mentions were negative, the visit was classified as “non-OUD.” A small proportion of visits had both positive and negative mentions of OUD. We manually reviewed these cases and classified them as OUD or non-OUD.

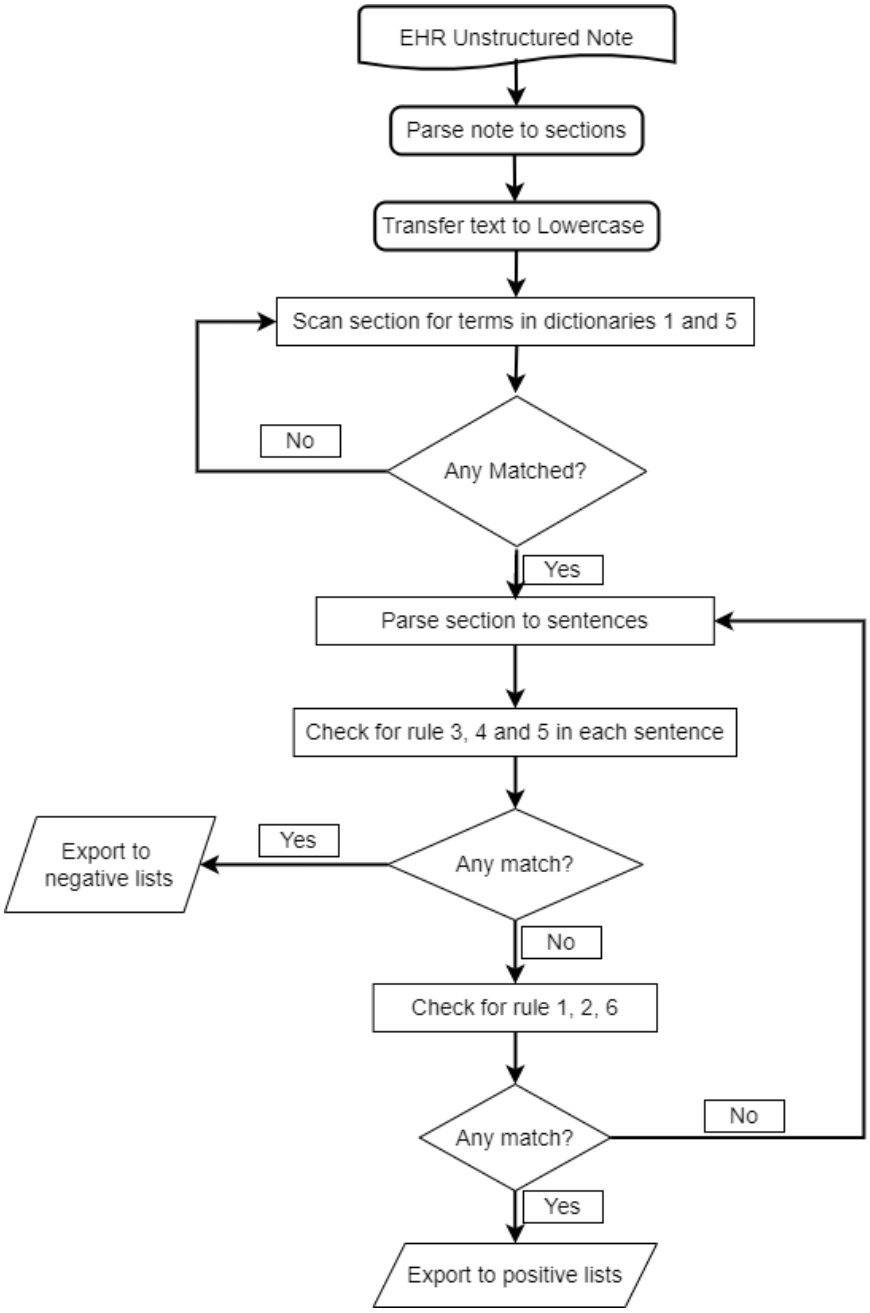


Figure 3. 1 Parsing and classification process for individual mentions of opioid use disorder in electronic health record unstructured clinical notes

3.3.3.6 Statistical methods

Sensitivity and specificity of the NLP algorithm were computed with reference to 300 manually-reviewed cases as the ground truth diagnosis. OUD cases were compared between the classifications by the NLP algorithm and by ICD-10 codes. Demographic characteristics of patients classified as positive for OUD by ICD-10 only, NLP only, and both ICD-10 and NLP were compared using means (Smart et al.) and proportions. All statistical analyses were carried out using SAS 9.4® M6 (SAS Inc, Cary, NC).

3.4 Results

We identified 29,212 hospital inpatient and Emergency Department visits occurring in 2019 (1,440 ED [4.9%] and 28,079 hospital [95.1%]) meeting inclusion criteria. Those visits generated 116,974 unstructured clinical notes, an average of 3.96 notes per patient visit. These notes comprised 59,780 discharge summaries (51.1% of all notes) (One visit may include several discharge summaries due to behavior discharge note was categorized as discharge summary notes), 22,080 History and Physical notes (18.9%), 18,679 ED General notes (16.0%), 14,927 ED Triage Notes (12.8%), and 1,508 Addiction Medicine Consult notes (1.3%).

About 67.0% of the study patients were between 35 and 74 years old, 20.4% were between 18 to 34 years old, and 12.9% were 75 years or older at the time of visit. The majority of patients' visits were among male patients (52.0%) and with patient race reported as White for 88.5% of the visits. Ten percent of patient visits were among Black patients (9.9%), and patients with reported other or unknown races accounted for 1.6%. Similarly. Hispanic, the majority of patients were reported as non-Hispanic ethnicity

(86.4%), and Hispanic ethnicity was reported for 2.8% patients, with the remainder having no ethnicity information reported. visits.

3.4.1 OUD case ascertainment by ICD-10-CM

We identified 1,811 visits having any ICD-10-CM code for OUD. Of these, 57 (3.1%) were ED visits and 1,754 (96.9%) were inpatient hospital stays. The numbers of visits having each ICD-10-CM code for OUD are shown in Supplement table 3.1. Based on ICD-10-CM coding, opioid dependence accounted for 48.9% of OUD diagnoses, opioid abuse for 42%, and unspecified opioid use for 9.1%.

3.4.2 OUD case ascertainment by NLP

The NLP algorithm identified 1,902 patient visits as having evidence of OUD in the unstructured clinical notes. Of these, 1,844 (97.0%) were identified from hospital visit data and 58 (3.0%) from ED visits. The NLP algorithm identified 24,822 total mentions of OUD across the 29,212 visits and the five selected note types. ED General Notes and History and Physical Notes contained a majority of OUD mentions, and these tended to be negative (for example, denial of opioid misuse). Nearly 74% of the positive mentions of OUD were identified in Discharge Summary Notes or Addiction Medicine Consult Notes, with another 18.3% identified in History and Physical Notes. Most of the negative mentions of OUD (94%) came from ED General Notes or History and Physical Notes. (Table 3.4).

Table 3. 4 Distribution of OUD mentions by note type

Note Group	Mentions of OUD (n=22,715)	Positive mentions of OUD[1] (n=6,186)	Negative mentions of OUD[2] (n=16,529)
ED General Notes	10,676 (47.0%)	446 (7.2%)	10,230 (61.9%)
History and Physical notes	6,949 (30.5%)	1,134 (18.3%)	5,360 (32.4%)
Discharge Summary Notes	3,948 (17.3%)	2,710 (43.8%)	788 (4.8%)
Addiction Medicine Consult Notes	1,972 (8.7%)	1,841 (29.8%)	131 (0.8%)
ED Triage Notes	75 (0.3%)	55 (0.9%)	20 (0.1%)
Total	22,715 (100%)	6,186 (100%)	16,529 (100%)

1 A positive mention of OUD is one which indicates the presence of OUD for the present visit
2 A negative mention of OUD is one which indicates the absence of OUD (e.g. "patient denies opioid abuse," refers to a historical condition (e.g. "history of opioid abuse"), etc.

3.4.3 Comparison of results

OUD cases identified by ICD-10-CM and NLP is summarized in Figure 3.2. The absolute number of visits with evidence of OUD identified by each method was similar, with NLP identifying 91 more cases. While there was substantial overlap in the identified cases (1,381 [59.2%]), overall 2,332 unique visits were identified. Of the total unique visits, 430 (18.4%) were identified only by ICD-10-CM codes, and 521 (22.3%) were identified only by NLP.

The prevalence of visits with evidence of an OUD diagnosis in this sample, ascertained using only ICD-10-CM codes, was 1,811/29,212 (6.1%). Including the additional 521 visits identified only by NLP, the estimated prevalence of OUD is 2,332/29,212 (7.9%), an increase of 29.5%.

Demographics characteristics of patient visits with a mention of OUD are presented in Table 3.5. Compared to ICD codes alone, NLP codes alone identified a greater proportion of males (54.7% vs. 49.1%), patients age 55 or older (29% vs. 17.7%), Black patients (10% vs. 5.1%) and Hispanic patients (1.3% vs. 0.5%), and married patients (23.2% vs. 17.2%).

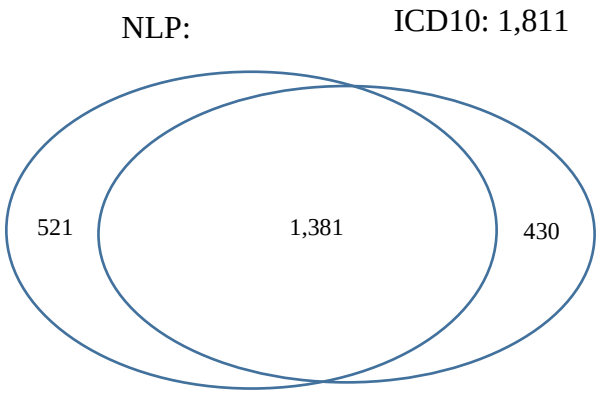


Figure 3. 2 Numbers of OUD visits identified by ICD-10-CM codes only, NLP algorithm only, and both ICD-10-CM codes and NLP algorithm; and total number of OUD visits identified by each method

Table 3. 5 Cases characteristics by ascertainment method

Characteristics	OUD-ICD only (n=430)	OUD-NLP only (n=521)	Common (n=1,381)
Gender			
Male	211 (49.1%)	285 (54.7%)	672 (48.7%)
Female	219 (50.9%)	236 (45.3%)	722 (52.3%)
Age			
18-34	127 (29.5%)	118 (22.6%)	507 (36.7%)
35-54	227 (52.8%)	252 (48.4%)	692 (50.1%)
55-74	71 (16.5%)	135 (25.9%)	169 (12.2%)
75+	5 (1.2%)	16 (3.1%)	13 (0.9 %)
Race			
African American	22 (5.1%)	52 (10.0%)	54 (3.9%)
European American	405 (94.2%)	461 (88.5%)	1318 (95.4%)
Other	3 (0.7%)	8 (1.5 %)	9 (0.7%)
Ethnicity			
Hispanic	2 (0.5%)	7 (1.3%)	6 (0.4%)
Non-Hispanic	379 (88.1%)	448 (86.0%)	1211 (87.7%)
Other	49 (11.4%)	66 (12.7%)	164 (11.9%)

3.5 Discussion

We measured the prevalence of opioid use disorder in UKHC hospital and emergency department patients using ICD-10-CM codes and an NLP rule-based algorithm. Our study identified 1,902 OUD (6.4%) visits out of 29,212 total visits by NLP algorithm, while a search of ICD-10-CM codes identified 1,811(6.1%) OUD cases from the same population. Combining ICD- and NLP-identified OUD visits gives 2,332 (7.9%) OUD cases in our sample.

Similar methods have been applied to a wide range of disease conditions (Afzal et al., 2016; Datta et al., 2019; Woller et al., 2021; Zheng et al., 2016). Here we mention only a few examples. Tian et al (2015) reported 0.94 sensitivity (Tian et al., 2017), 0.96 specificity and 0.73 PPV for identifying DVT in narrative radiology reports by NLP. Afzal

et al (2016) reported sensitivity 0.96, specificity 0.98 and PPV 0.92 for identifying PAD in clinical EHR notes using HER (Afzal et al., 2016). Wright et al (2013) reported areas under the curve (AUC) of 0.956 and 0.947 at two separate hospitals for identifying diabetes in EHR progress notes (Wright et al., 2013).

Application to OUD has been less common. Carrell et al. (2015) used a rule-based NLP approach to identify 1,875 (8.5%) patients with problem prescription opioid use (POU) in a sample of 21,795 patients who were receiving chronic opioid therapy (Carrell et al., 2015). An ICD-9-CM code search identified 2,240 (10.1%) patients with POU from the same sample. Our study found a similar result for OUD in a broader population of patients that was not restricted to those receiving chronic opioid therapy.

The discrepancy between ICD- and NLP-identified OUD in our study has several possible explanations. First, we did not search all clinical notes. Instead, we selected only the 5 types of notes that we believed were most likely to contain mentions of OUD. This could account for some cases in which the ICD code search identified a case of OUD but the NLP algorithm did not. Second, certain terms used to refer to OUD in text notes are ambiguous. For example, upon manual review, we identified cases in which the ICD code search identified a case of OUD and the NLP algorithm failed to identify a mention of OUD. However, the note text contained a reference to "polysubstance abuse." Third, Other reason may cause by ICD10 missed coded by coders, some cases specified that the patient "denied drug use" and no other opioid use information was mentioned in any notes, but there was an ICD10 code for this case.

There were several significant differences in the demographic characteristics of patients with OUD identified by NLP and ICD. The group of patients identified only by

NLP included slightly higher percentages of males, older patients, black patients and Hispanic patients. It is not immediately clear why these differences should exist.

3.6 Limitations

In this study, we used a rule-based NLP algorithm to analyze electronic notes from patient visits to identify OUD cases, and we compared the results to cases identified using a search of ICD-10-CM codes. Our sample was limited to hospital and emergency department visits in a single healthcare setting. The findings cannot be generalized to all healthcare settings. We limited the text mining search to five types of notes that were considered most likely to include mentions of OUD. In particular we did not include psychiatry notes that include behavioral and mental health information. In our rule-based algorithm, the development of keyword dictionaries and parsing rules relied on literature reviews and expert opinion. Although we included over 1,000 entries that healthcare workers might use to describe OUD in text notes, it is still possible that terms were missed by our dictionaries and parsing rules. In particular, OUD mentions that contain abbreviations or spelling errors may type error result in false negatives. Finally, negation rules are imperfect, and may result in misclassification of cases. For example, one of our negation rules allowed 3 or fewer tokens between the negation term and the opioid use term. This rule would fail to correctly negate the following, mention of OUD: “Patient denies fevers or chills, vomiting or diarrhea, and IV drug abuse”.

3.7 Conclusion

The findings in this study concluded that it is feasible to identify patients with OUD by rule-based NLP algorithm from unstructured clinical notes. It suggested that rule-based NLP algorithm applied to identify potential OUD cases have the potential to improve surveillance of opioid use disorder compare to the traditional methods that only rely on ICD-10-CM codes.

CHAPTER 4. EFFECT OF EXPOSURE MISCLASSIFICATION ON THE ASSOCIATION BETWEEN OPIOID USE DISORDER AND CARDIOVASCULAR DISEASE RISK

4.1 Abstract

In recent years, concerns have emerged about possible cardiovascular effects of opioid use. One common approach to studying this relationship involves administrative hospital records, using International Classification of Diseases (ICD) codes for opioid use disorder (OUD) as a proxy for opioid exposure. This type of study is common due to the convenience, low cost and ready availability of administrative hospital data. However, reliance on ICD codes only can lead to potential misclassification of OUD status and bias effect estimates examining the association of OUD with the risk of cardiovascular disease. In a previous study of 29,519 inpatient hospital and emergency department visits, we identified patient visits where the patient was classified as having OUD using ICD-10-CM codes (F11.1*, F11.2*, F11.9*) and separately using a rule-based natural language processing (NLP) algorithm from unstructured clinical notes. Using this sample, we conducted a cross-sectional study to estimate the association between OUD and six cardiovascular diseases (CVD) and conditions. Prevalence of OUD was ascertained from ICD codes alone or ICD codes plus NLP algorithm applied to clinical notes. Multivariable Poisson regression models, with a robust variance estimator, were used to estimate prevalence rate ratios to quantify the association between OUD and CVD prevalence. Our sample included 22,501 adult inpatients. Prevalence of OUD was identified from ICD-10-CM codes for 1,478 patients. Another 391 patients could only be identified by an NLP algorithm applied to unstructured clinical notes. Changes in prevalence rate ratio estimates, when patients with OUD that was only identifiable by NLP analysis of clinical notes were

reclassified from non-OUD to OUD, were modest. All of the changes were less than 10% different from the estimates that were based on OUD assessment using ICD-10-CM codes alone. We observed modest effects of misclassification of OUD status on estimates of the association between OUD and CVD. However, weak but statistically significant associations could result from misclassification of OUD status based on ICD-10-CM codes, when in fact no association exists. Estimates from administrative data should be interpreted with caution.

4.2 Background

Exposure to opioid drugs among United States (U.S.) residents has increased exponentially over the past 30 years (Hedegaard et al., 2018). This includes the use of prescription opioids for medical purposes as directed by a physician – such as for treatment of cancer-related or noncancer chronic pain – as well as non-medically indicated use of both prescription and illicit opioids (Cochran et al., 2015; Kaye et al., 2017; Khodneva et al., 2016). Between 1997 and 2007, opioid prescribing in the U.S. increased from 100 to nearly 700 morphine milligram equivalents per capita (Paulozzi et al., 2011). This increase in opioid availability has been accompanied by steep increases in fatal and nonfatal overdoses as well as increases in the prevalence of opioid use disorder (OUD) (Haight et al., 2018; Martins et al., 2017).

In recent years, concerns have emerged about possible cardiovascular effects of opioid use (Khodneva et al., 2016). The risk of infective endocarditis associated with injection drug (including opioid) use has been well-documented (Mihm et al., 2020; Sinner et al., 2021). In a review of the relationship between opioid use and cardiac

arrhythmia, Behdazi et al (2018) reported that methadone posed a high risk of QT interval prolongation and arrhythmogenicity, even at low doses (Behzadi et al., 2018).

Previously we published a review of research on nonacute opioid use – which we defined to include chronic opioid therapy and opioid use disorder (OUD) – and CVD prevalence (Singleton et al., 2021). We summarized the literature on the association between NOU and 5 classes of cardiovascular disease, including infective endocarditis (IE), coronary heart disease, acute myocardial infarction (MI), congestive heart failure, cardiac arrythmia, and stroke. There was generally consistent evidence of a positive association with the risk of IE and MI. There was less consensus about the relationship between NOU and the risk of heart failure, cardiac arrhythmia, and stroke.

Methodologically, several of extant studies relied on administrative hospital records (Dakour-Aridi et al., 2019; Gupta et al., 2016; Menendez et al., 2015), using a diagnosis of opioid use disorder (OUD) as a proxy for opioid exposure. Both OUD and cardiovascular conditions can be identified from electronic health records using International Classification of Diseases (ICD) codes. The limitations of ICD codes, including low sensitivity and specificity, have been well-documented for many conditions (Hughes Garza et al., 2021; Kurbasic et al., 2008; O'Malley et al., 2005; Quan et al., 2008), although there has been little investigation of the extent to which this may be true for OUD. If ICD codes have low sensitivity to identify OUD when it is present, then many patients with OUD could be misclassified as not having OUD. This could result in biased estimates of the association between OUD and any other conditions of interest, including CVD.

The objective of this study was to assess the degree to which incomplete identification of OUD based on ICD codes could bias the estimated association between

OUD prevalence and the risk of CVD. Study data were drawn from a unique data set consisting of hospital and emergency encounters from the electronic health record (EHR) of a single academic medical center. OUD was classified into two sets, one was only by ICD-10-CM diagnostic codes, and the other was by either ICD-10-CM codes or by NLP algorithm on unstructured clinical notes. Using this data set, we estimated univariable ("crude") and multivariable ("adjusted") associations between OUD and 6 types of CVD, with these two sets of OUD and discussed the effect of misclassification of OUD on the association of OUD and CVD.

4.3 Methods

4.3.1 Data Source and study sample

University of Kentucky HealthCare (UKHC) documents were delivered through Sunrise Clinical Manager (SCM) EHR for inpatient visits before June 2021. Structured data documenting delivered care include, but are not limited to, information on patient enrollment, clinical encounters, demographic characteristics, diagnoses, procedures, problems, medications, and laboratory orders and results. Clinical notes can be obtained from EHR as unstructured data elements. This study was approved by both the UK Institutional Review Board (IRB #20548) and the UKHC Data Management Committee.

The sample consisted of hospital inpatient and emergency department visits to UKHC for adults age 18 years of age or older, occurring between January 1, 2019 and December 31, 2019. We excluded patients with active cancer due to their high level of opioid use, which does not necessarily reflect OUD (Alzeer et al., 2018). We summarized

the visit-oriented sample (potentially multiple visit records per patient) into a patient-oriented sample (one record per patient, N=22,501) by scanning over all visits for the exposure (OUD) and outcome (CVD) of interest in 2019.

4.3.2 Exposure

The exposure was OUD, ascertained by two methods: OUD identified by ICD codes alone, and OUD identified by either ICD codes or NLP applied to clinical notes. The ICD-10-CM codes used to identify OUD were F11.1* for opioid abuse, F11.2* for opioid dependence and F11.9 for unspecified opioid use.

From our previous study, we identified 1,811 patients visits with an ICD-10-CM code for OUD, and 1,902 patients visits with OUD mentioned in clinical notes. Of the 1,902 visits with OUD identified by NLP, 591 (21%) were not identifiable by ICD-10-CM codes (i.e., there was no ICD-10-CM code documenting OUD). In this study, we looked at the OUD status in patient level instead of patient visit level. There are four conditions for which we had to convert observations from the patient visit level to the patient level (Table 4.1) For example, if a patient visit was classified as having OUD only by ICD-10-CM code, then this patient was classified as OUD by ICD; if a patient visit was classified as OUD only by NLP, then this patient was classified as OUD by NLP; if a patient visit was classified as OUD by both ICD and NLP, then this patient was classified as common; if a patient had multiple visits, and one visit was classified as OUD by ICD and the others classified as OUD by NLP, then this patient was classified as common. Proportions of OUD patients classified according to information obtained from ICD-10-CM or NLP are shown in Figure 4.1.

Table 4. 1 Convert patient visit level to patient level

Patient visit ID	OUD status for visit	Patient ID	OUD status for patient
000000001-0001	OUD_NLP	000000001	OUD_NLP
000000002-0001	OUD_NLP	000000002	OUD_NLP
000000003-0001	Common	000000003	Common
000000004-0001	OUD_NLP	000000004	Common
000000004-0002	OUD_ICD	000000004	Common

In this sample of patients (N=22,501), we identified 1,478 patients with an ICD-10-CM code for OUD, and 1,467 patients with OUD mentioned in clinical notes. Of the 1,467 visits with OUD identified by NLP, 391 (21%) were not identifiable by ICD-10-CM codes (i.e., there was no ICD-10-CM code documenting OUD). Numbers of OUD patients classified according to information from ICD-10-CM or NLP are shown in Figure 4.1.

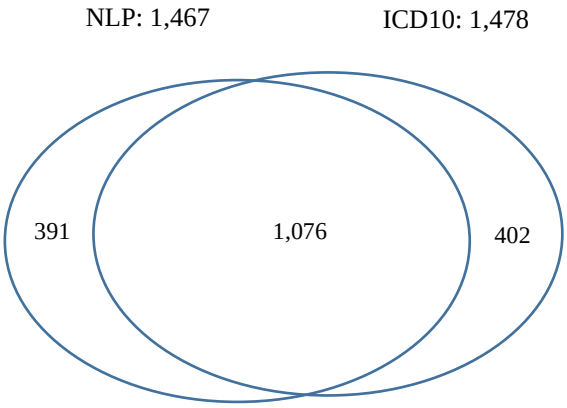


Figure 4. 1 Numbers of OUD Patients identified by ICD-10-CM codes only, NLP algorithm only, and both ICD-10-CM codes and NLP algorithm; and total number of OUD patients identified by each method

4.3.3 Outcome: CVD conditions

Prevalence of cardiovascular conditions was assessed by scanning all diagnosis codes associated with the hospital visits occurring in 2019. ICD-10-CM codes for cardiovascular conditions are listed in Supplemental Table 4.1. The types of CVD conditions identified included cardiac arrhythmia, heart failure, acute myocardial

infarction (MI), stroke, ischemic heart disease, and infective endocarditis. These conditions were selected due to their high prevalence and public health impact. There is research suggesting that ICD codes for MI, arrhythmias and stroke have good sensitivity and specificity (Birman-Deych et al., 2005 ; Frolova et al., 2015; Jensen et al., 2012; Kivimaki et al., 2017; Metcalfe et al., 2013). Significant misclassification of heart failure is to be expected when using ICD codes (Rosamond et al., 2012). In this study, we did not take potential misclassification of CVD outcomes into account.

4.3.4 Covariates

Possible confounders of the association between OUD and CVD, assuming a causal relationship exists, were identified based on the review of extant literature and construction of a directed acyclic graph (DAG; Figure 4.2). Based on the literature review and the minimal sufficient adjustment set identified by the DAG (Hernan and Robins, 2020), we included the following demographic variables: patient age at first visit in 2019, gender, race/ethnicity; clinical comorbidity variables: hypertension, diabetes, and the total number of inpatient visits in 2019 as a proxy for healthcare utilization. Prevalence of hypertension and diabetes at the time of hospitalization was identified from ICD-10-CM Codes: E11.9 for diabetes and I10 for Hypertension.

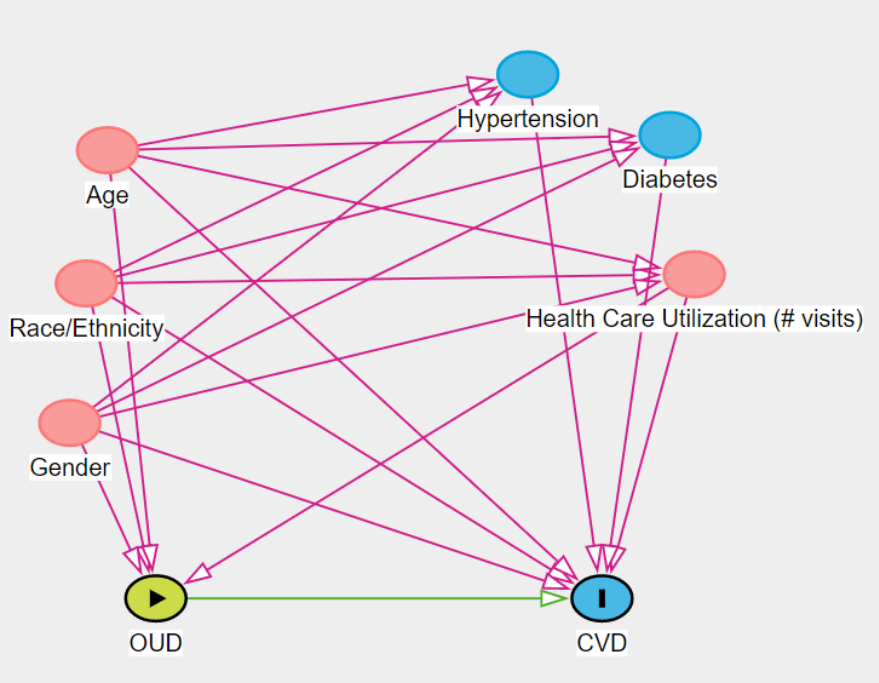


Figure 4. 2 Directed acyclic graph (DAG) of the association of opioid use disorder (OUD) with the prevalence of cardiovascular disease (CVD)

Note: Red lines indicate non-causal pathways

4.3.5 Study design and Statistical Analysis

Characteristics of patients, stratified by the method of OUD ascertainment (ICD or NLP), were summarized using means and frequencies.

We conducted two versions of a cross-sectional study to estimate the association between OUD and the prevalence of CVD conditions. Poisson regression models, with a robust variance estimator, were used to estimate prevalence rate ratios (Barros and Hirakata, 2003) quantifying the association between OUD and CVD prevalence. In one set of models, OUD was identified only using ICD-10-CM codes (all patient visits without an ICD-10-CM code for OUD were coded as non-OUD). In the second set of models, OUD

was identified from ICD-10-CM codes and through NLP of clinical notes (here, all patient visits without any ICD code for OUD or any evidence of OUD from the clinical notes were coded as non-OUD). Estimates from the two sets of models were compared to assess the potential bias due to misclassification resulting from under-ascertainment of OUD from ICD codes. Multivariable models included the covariates listed above and the model was specified as:

Model 1: $\text{Log}(Pr^*CVD) = \beta_0 + \beta_1 * \text{OUD\_ICD} + \beta_2 * \text{Age\_group} + \beta_3 * \text{Gender} + \beta_4 * \text{Race} + \beta_5 * \text{Ethnicity} + \beta_6 * \text{Diabetes} + \beta_7 * \text{Hypertension} + \beta_8 * \text{Visit\_Count}$

Model 2: $\text{Log}(Pr^*CVD) = \beta_0 + \beta_1 * \text{OUD\_ICD/NLP} + \beta_2 * \text{Age\_group} + \beta_3 * \text{Gender} + \beta_4 * \text{Race} + \beta_5 * \text{Ethnicity} + \beta_6 * \text{Diabetes} + \beta_7 * \text{Hypertension} + \beta_8 * \text{Visit\_Count}$

Note: Pr* is the probability of a patient having CVD, conditional on the values of the independent variables.

The statistical significance level for this study was fixed at 0.05. All analyses were conducted using SAS version 9.4 TS (SAS Institute, Cary, NC).

4.4 Results

Our sample included 22,501 adult patients. A diagnosis of OUD, based on ICD-10-CM codes was available for 1,478 patients. Another 391 patients could only be identified as having OUD from an NLP algorithm applied to unstructured clinical notes. Table 4.2 summarized different types of CVD identified in different OUD or non-OUD groups. Table 4.3 summarizes the characteristics of these patients. As compared to patients with OUD who could be identified by ICD codes, patients with OUD that could only be identified by NLP they were older (46.3 years vs. 40.9 years), higher percentage male (55% vs. 47.8%)

and African American (9.2% to 4.9%), and had higher prevalence of diabetes (12.3% vs. 8.2%) and hypertension (40.9% vs. 32.5%).

Table 4. 2 CVD identification in different OUD groups

	OUD by ICD (n=1,478)	OUD by NLP only (n=391)	Non-OUD (n=20,632)	Total (n=22,501)
Overall CVD	667 (6.2)	184 (1.7)	9909 (92.1)	10,760 (100%)
Cardiac arrhythmia	546 (7.4)	132(1.8)	6734 (90.9)	7,412 (100%)
Acute myocardial infarction	50 (4.3)	16 (1.4)	1095 (94.3)	1,161(100%)
Stroke	113 (4.1)	27 (1.0)	2598 (94.9)	2,738 (100%)
Heart failure	138 (3.0)	62 (1.3)	4433 (95.7)	4,633 (100%)
Ischemic heart disease	162 (4.9)	61 (1.9)	3071 (93.2)	3,294 (100%)
Infective endocarditis	13 (65.0)	1 (5.0)	6 (30.0)	20 (100%)

Table 4. 3 Characteristics of all patients and patients by OUD status as ascertained by ICD-10-CM or NLP

	All Patients (n=22,501)	All OUD* (n=1,869)	OUD_ICD (n=1,478)	OUD_NLP only (n=391)
Age, mean	53.2 (18.8)	42.1 (13.1)	40.9 (12.3)	46.3 (14.8)
Male, Sex (%)	10677 (47.5%)	922 (49.3%)	707 (47.8%)	215 (55.0%)
Race, (%)				
European American	20136 (89.5%)	1760 (94.2%)	1405 (95.1%)	355 (89.4%)
African American	2365 (10.5%)	109 (5.8%)	73 (4.9%)	36 (9.2%)
Ethnicity, (%)				
Non-Hispanic	18830 (83.7%)	1593 (85.2%)	1265 (85.6%)	328 (83.9%)
Others	3671 (16.31)	276 (14.8%)	213 (14.4%)	63 (16.1%)
Diabetes	3570 (15.9%)	169 (9.0%)	121 (8.2%)	48 (12.3%)
Hypertension	9608 (42.7%)	641 (34.3%)	481 (32.5%)	160 (40.9%)
Visit Counts	1.3 (0.9)	1.6 (1.6)	1.6 (1.6)	1.7 (1.7%)

Note: All OUD is ascertained either by ICD-10-CM or NLP

Figure 4.3 presents the differences in the proportion of patients with OUD discoverable only by NLP analysis of clinical notes, for patients with and without each type of CVD. We compared the prevalence of CVD or non-CVD that was identified from patients with OUD identified by NLP. The differences were greatest for stroke, ischemic heart disease and acute MI. In other words, there is some amount of differential misclassification of OUD status for these three types of CVD.

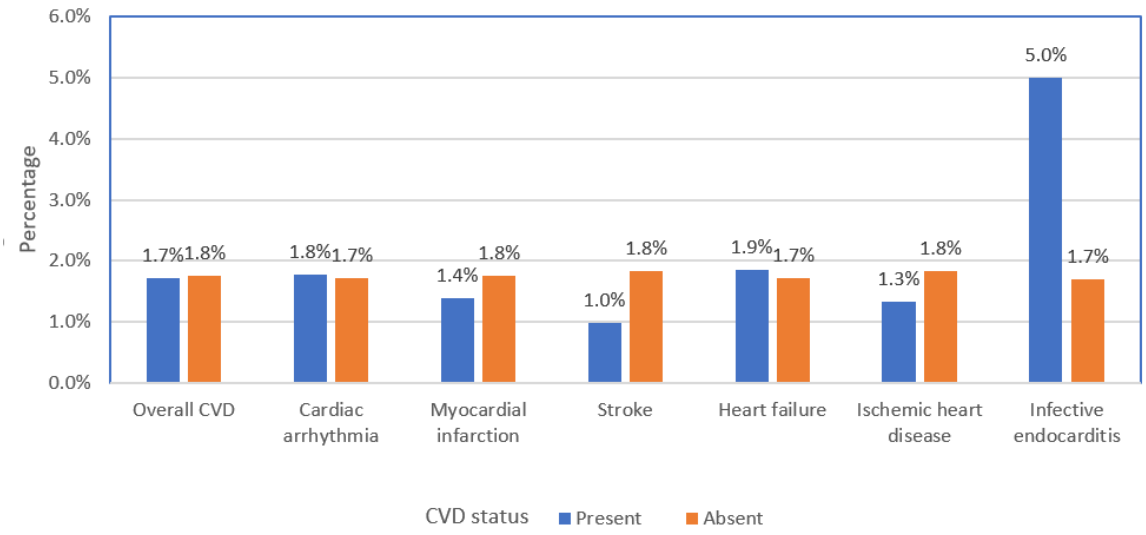


Figure 4. 3 Proportion of CVD conditions and non-CVD conditions where OUD was ascertained only by NLP

Table 4.4 presents crude prevalence rate ratio estimates for CVD when OUD was assessed by ICD-10-CM codes alone or by ICD-10-CM codes combined with NLP. The change in the prevalence ratio point estimates when assessment of OUD was expanded to include cases identified by NLP was generally less than 5%. The point estimate for ischemic heart disease increased by 13.6% and the point estimate for heart failure increased by 8%.

Table 4. 4 Crude prevalence rate ratios for cardiovascular disease by opioid use disorder (OUD) ascertainment method

Dependent variable	OUD by ICD-10-CM	OUD by ICD-10-CM + NLP
Any cardiovascular disease	0.94 (0.89-1.00)	0.95 (0.90-1.00)
Cardiac Arrhythmia	1.13 (1.06-1.21)	1.11 (1.04-1.18)
Acute myocardial infarction	0.64 (0.48-0.85)	0.67 (0.52-0.85)
Stroke	0.61 (0.51-0.73)	0.59 (0.51-0.70)
Heart failure	0.74 (0.63-0.85)	0.80 (0.71-0.91)
Ischemic heart disease	0.44 (0.37-0.51)	0.50 (0.44-0.57)
Infectious endocarditis	26.4 (10.5-66.1)	25.7 (9.9-66.9)

Table 4.5 presents the adjusted prevalence rate ratio estimates when age, gender, race, ethnicity, diabetes, hypertension, and number of inpatient visits in 2019 are taken into account. Compared to the crude estimates, the adjusted estimates for all types of CVD changed. For any cardiovascular disease and heart failure, the changes accrued in both direction and magnitude. This was due primarily to the imbalances in patient characteristics for patients with or without OUD. As table 4.3 shows, patients with OUD were on average much younger than xxx, and age is a strong risk factor for CVD. However, the differences in xxx by OUD ascertainment method between the sets of adjusted estimates were fairly small. The largest difference was for stroke: a decrease of 8.3% (from 0.96 to 0.88). The estimate for ischemic heart disease increased by 6.3% (0.80 to 0.85). The adjusted estimates for the other CVD types changed by less than 5% given ascertainment by NLP. This suggests that the effect of misclassification of OUD based on using ICD-10-CM codes alone was small.

Table 4. 5 Adjusted prevalence rate ratios for cardiovascular disease when OUD is assessed by ICD-10-CM codes alone and by ICD-10-CM codes and NLP together

Dependent variable	OUD by ICD-10-CM	OUD by ICD-10-CM + NLP
Any cardiovascular disease	1.26 (1.19-1.34)	1.22 (1.16-1.29)
Cardiac Arrhythmia	1.41 (1.31-1.52)	1.34 (1.25-1.43)
Acute myocardial infarction	0.94 (0.7-1.26)	0.92 (0.71-1.18)
Stroke	0.96 (0.8-1.15)	0.88 (0.75-1.04)
Heart failure	1.15 (0.98-1.35)	1.20 (1.05-1.36)
Ischemic heart disease	0.80 (0.69-0.94)	0.85 (0.74-0.96)
Infectious endocarditis	18.7 (6.95-50.3)	18.78 (6.8-51.82)

4.5 Discussion

We investigated the degree to which under-ascertainment of OUD by ICD-10-CM codes could bias the estimated association between OUD and six types of CVD, in a sample of patients identified from the EHR system of an academic medical center. Changes in prevalence ratio estimates, when patients with OUD that was only identifiable by NLP analysis of clinical notes were reclassified from non-OUD to OUD, were modest. All of the changes were less than 10% different from the estimates that were based on OUD assessment using ICD-10-CM codes alone.

When misclassification bias is nondifferential, there is a tendency for effect estimates to be biased towards the null value (Whitcomb and Naimi, 2020). When differential, the result is less predictable (Li and VanderWeele, 2020; Sorahan and Gilthorpe, 1994). In this study, misclassification of OUD based on ICD-10-CM codes was nondifferential for stroke, acute MI, IHD and IEF or IE, although IE was uncommon in our sample, therefore, the robustness of this estimate is questionable. The impact of OUD ascertainment method on the estimated prevalence ratio for IE was minimal.

The adjusted prevalence rate ratios for heart failure illustrate the potential for more complete ascertainment of OUD to affect conclusions. In Table 4 we reported that when OUD cases identified only by NLP were reclassified, the point estimate for heart failure increased from 1.15 to 1.20, and the limits of the 95% confidence interval increased from (0.98-1.35) to (1.05, 1.36). The confidence intervals also narrowed slightly in Model 2 for all CVD outcomes due the increasing sample size after the OUD cases identified by NLP were included.

In our multivariable models, we adjusted for several potential confounders of the association of OUD with the risk of CVD, that were identified based on published data. These variables were incorporated into a directed acyclic graph (DAG) to identify a minimally sufficient adjustment set. Ideally, in addition to the specified covariates, the DAG should have included smoking and alcohol use as those behaviors are known to associated with the prevalence of OUD and CVD. However, information on smoking and alcohol is not reliably captured in EHR data, so we did not include it.

Logistic regression is often used to estimate the odds ratio for a binary outcome in cross-sectional studies. For common outcomes such as CVD, the odds ratio can be misleading as a measure of the prevalence ratio. Poisson regression provides a direct estimate of the prevalence rate ratio. Applying Poisson regression to a binomial outcome will tend to overestimate standard errors for parameter estimates. However, this can be addressed using robust variance estimation, which is the approach we took in this study (Barros et al. 2003).

NLP applied to clinical notes can be used to assess the extent and impact of misclassification bias in research using ICD-coded medical record data. Unstructured

clinical notes can be transformed into structured information for analysis. However, these methods do require a considerable investment of resources, including skilled analysts and medical experts who can manually review cases to assess algorithm performance.

4.6 Conclusion

The effects of misclassification of OUD status on estimates of the association between OUD and CVD prevalence were very modest in this study. Weak, but statistically significant, associations could result from misclassification of OUD status based on ICD-10-CM codes, when in fact no association exists. Such estimates should be interpreted with caution.

CHAPTER 5. CONCLUSION

5.1 Summary

As the prevalence of long-term exposure to opioid drugs has increased considerably since 1990's, concerns have arisen about the potential cardiovascular effects of opioid use. This is a challenging topic to study for several reasons, the major one being the difficulty of accurately assessing an individual's exposure to opioid drugs. For example, in healthcare data, patient symptoms, conditions, behaviors and compliance with health care team advices may be documented in unstructured text notes and/or problem lists but not diagnosis codes. Additionally, some caregivers may be reluctant to assign diagnosis codes for opioid use disorder because of associated stigma and the relative ease with which this information is detectable through automated surveillance. The primary aims of this dissertation research were to demonstrate the utility of natural language processing (NLP) to identify cases of opioid use disorder (OUD) in electronic health records (EHR) that cannot be identified by ICD-10-CM codes; and to investigate the effect of misclassification of OUD by ICD-10-CM codes on estimates of the association between OUD and CVD. For context, we conducted a scoping review of the epidemiological literature on non-acute opioid use (Kivimaki et al.) and CVD. Data from the University of Kentucky's Hospital and Emergency Department inpatient EHR system for 2017 to 2019 were used to conduct two studies: "Using natural language processing to identify opioid use disorder from EHR notes," (2) "Effect of exposure misclassification on the association between opioid use disorder and cardiovascular disease." The remainder of this chapter summarizes the major findings from these studies.

Chapter Two was a scoping review of the epidemiological literature on nonacute opioid use and CVD. Twenty-three original articles from the PubMed database were identified either by key term search or Mesh term search. This review summarized the current evidence about the association between NOU and five classes of CVD, including infective endocarditis, coronary heart disease, congestive heart failure, cardiac arrythmia, and stroke. This review provided evidence that NOU poses a risk not only for cardiac disorders associated with infections due to needle re-use, such as infective endocarditis, but may also predispose persons to chronic cardiovascular disorders, including MI and arrhythmias. This review also demonstrated the dearth of high-quality evidence on the relationship between NOU and CVD. Many of the identified studies lacked detailed information on the duration and intensity of opioid exposure and all were retrospectively conducted. This is understandable, as the challenges to accurate assessment of NOU are considerable. Innovative approaches to opioid exposure assessment will be required. This review was published in the Journal of American Heart Association (Singleton et al., 2021).

In Chapter Three we developed a Natural Language Processing (NLP) pipeline for identifying OUD cases from unstructured clinic notes which no ICD-10-CM codes were assigned but with strong evidence of OUD in the unstructured clinical notes. We selected 5 types of clinical notes for inclusion in the study: ED triage notes, ED general notes, History and Physical notes, Addiction Medicine Consult notes, and Discharge Summary notes. With expert input and literature article reviewing, we developed five dictionaries: opioid terms, opioid use disorder terms, negation term, special terms of drug use and a special clinic related to opioid dependence; and 6 parsing rules from the five dictionaries. Three rules opedrationalized searching for positive mentions of OUD, and three rules

pertained to the negation of OUD. Using the Python Natural Language Toolkit, NLTK and Spacy, we developed an algorithm to carry out rule searches for OUD phrases in the unstructured clinical notes. All patient visits were classified as OUD or non-OUD based on information obtained from ICD-10-CM codes and by NLP algorithm. Overall, we identified 2,332 patients visit as OUD by ICD-10-CM codes or by NLP algorithm applied to clinical notes. 1902 (6.4%) OUD were identified by NLP and 1,811 (6.1%) identified by ICD-10-CM codes. 1,381 OUD were identified by both ICD-10-CM and NLP algorithm. 521 patient visits were only identified by NLP algorithm, which was the "hidden" OUD cases that were identified by NLP algorithm from the unstructured clinical notes but were missing structured ICD-10-CM. Applying the NLP algorithm, we identified 29% more of OUD cases compared with the traditional method that only relies on ICD-10-CM diagnosis codes.

In Chapter Three, we identified an extra group of OUD hospitalizations (521) using the NLP algorithm in addition to OUD classified by ICD-10-CM codes. In Chapter Four we carried out an experiment to assess the nature of the misclassification of OUD that occurs when it is identified using ICD-10-CM codes alone, and the effect of that misclassification on estimation of the association between OUD and CVD. We conducted two cross-sectional studies with OUD as the exposure and CVD as the outcome. We used multivariable Poisson regression models to estimate the prevalence odds ratio for CVD among patients with OUD, while adjusting for possible confounders. The only difference between the two studies was the method for ascertaining the patient's OUD status: by ICD-10-CM codes along, or by a combination of ICD-10-CM codes and NLP analysis of clinical notes. Changes in prevalence rate ratio estimates, when patients with OUD, that was only

identifiable by NLP analysis of clinical notes, were reclassified from non-OUD to OUD, were modest. All of the changes were less than 10% different from the estimates that were based on OUD assessment using ICD-10-CM codes alone. We concluded that weak associations between OUD and CVD based on ICD-coded administrative hospital data should be interpreted cautiously.

5.2 Strengths and Limitations

A major strength of this dissertation is that the data were drawn from a large, clinical cohort with access to all the physician notes, and ICD-10-CM codes. Our team included a clinical physician who helped to interpret the results and clarify questions about clinical processes. As we know that NLP provides a powerful tool for text mining of unstructured notes to produce structured information for research. Using the open-source programming language, Python, we developed a rule-based NLP algorithm to search the unstructured clinical notes for mentions of OUD terms and identified an extra group of OUD cases compared to traditional method that only rely on ICD-10-CM codes.

Using a rule-based NLP algorithm, we identified 2,332 patient visits with evidence of OUD in unstructured clinical notes. In 2015, Carrell et al. used a rule-based NLP approach to identify 1,875 (8.5%) patients with problem prescription opioid use (POU) in a sample of 21,795 patients who were receiving chronic opioid therapy. An ICD-9-CM code search identified 2,240 (10.1%) patients with POU from the same sample. Our work extends the findings of ICD-9-CM to ICD-10-CM codes, and extends those receiving chronic opioid therapy to a broader population of patients not restricted to persons receiving chronic opioid therapy.

After we classified patients as having OUD by both ICD-10-CM and by the NLP algorithm applied to unstructured notes from 2019, we also identified all the patients in 2019 with six type of cardiovascular disease conditions using ICD-10-CM codes: cardiac arrhythmia, heart failure, acute myocardial infarction (MI), stroke, ischemic heart disease, and infective endocarditis. We conducted a cross sectional study to estimate the association between OUD and six cardiovascular diseases (CVD) and conditions. Prevalence of OUD was ascertained from ICD-10-CM codes alone or ICD-10-CM codes plus NLP algorithm applied to unstructured notes. Multivariable Poisson regression models, with a robust variance estimator, were used to estimate prevalence rate ratios to quantify the association between OUD and CVD prevalence. We also investigated the misclassification of OUD status on estimates of the association between OUD and CVD.

A limitation of our study is that the patient sample derives from a single hospital over a period of two years, which is not able to generalized to population level. Moreover, we used a single year of data (2019) for the study of association of OUD and CVD. This limits the ability to generalize the findings.

Another limitation is that we included only 5 types of clinical notes. These notes were selected based on expert opinion and literature review that they were the most likely to include mentions of OUD. However, the total number of clinical notes available was much larger. This could account for some cases in which the ICD code search identified a case of OUD but the NLP algorithm did not.

In addition, we identified the outcome (CVD conditions) using ICD-10-CM codes alone, introducing potential misclassification of the outcome. Finally, in our cross-sectional design in Chapter Four, CVD and OUD were both assessed over a period of one

year. Our design could be improved by assessing both OUD and CVD using both ICD codes and NLP in a multi-year, longitudinal sample.

5.3 Future research

To fully understand the association of CVD and OUD using EHR data, a future study would identify both OUD and CVD from both ICD codes and full sample of clinical notes, covering multiple sites and years. Also, confounding by smoking, alcohol, and other variables should be considered. Duration and dose of opioid use should be included in the study. This kind of study will require access to the data and coordination across facilities, and a longer time period because of the long latency period for CVD.

Supplemental Table 2. 1 Specific opioids used in the 23 reviewed studies (cf Table 2.2)

References	Opioid name
I	Codeine, codeine, dihydrocodeine, dihydrocodone, fentanyl, hydrocodone, hydromorphone, levorphanol, meperidine, methadone, morphine, oxycodone, oxymorphone, propoxyphene, tramadol,
II	Heroin, oxycodone, oxymorphone
III	Heroin and methadone
IV	Morphine, oxycodone, oxycodone-naloxone, hydromorphone, tapentadol, fentanyl, Buprenorphine
V	Heroin
VI	Diphenoxylate, hydrocodone, hydromorphone, meperidine, morphine sulfate, oxycodone, pentazocine, tramadol, fentanyl, codeine
VII	Any prescribed opioid: buprenorphine, morphine, meperidine, tramadol, codeine, dihydrocodeine, propoxyphene, meptazinol
VIII	Hydrocodone bitartrate, codeine phosphate, oxycodone hydrochloride, propoxyphene hydrochloride, tramadol hydrochloride
IX	Anileridine, codeine phosphate, hydromorphone HCL, morphine HCL, meperidine HCL, oxycodone HCL, codeine sulfate, codeine phosphate, acetaminophen-caffeine-codeine, acetaminophen-codeine phosphate, fentanyl transdermal, acetaminophen-codeine, acetylcalicylic acid-codeine-caffeine, acetylsalicylic acid-codeine, oxycodone-HCL-acetaminophen, oxycodone HCL-acetylsalicylic acid, morphine HCL, morphine sulfate

Supplemental Table 3. 1 OUD distribution of ICD-10-CM codes

ICD-10-CM Code for OUD	Description	Counts (%)
F11.1	Opioid abuse	
F11.10	Uncomplicated	692 (38.2%)
F11.11	In remission	58 (3.2%)
F11.12, F11.14, F11.18, F11.19	With complications	11 (0.6%)
F11.2	Opioid dependence	
F11.20	Uncomplicated	632 (35.0%)
F11.21	In remission	24 (1.3%)
F11.22, F11.23, F11.24, F11.25, F11.28, F11.29	With complications	229 (12.6%)
F11.9	Opioid use	165 (9.1%)
Total		1811 (100%)

Supplemental Table 3. 2 Specialized term lists

Specialized term lists
denies alcohol/illicits/tobac
denies drug
denies drug abuse
denies etoh, illicit drugs
denies history of alcohol, tobacco or illegal drug use
denies illicit drug use
denies illict or iv drug use
denies iv drug use
denies smoking, alcohol, illicit drug use
denies smoking, drinking, or using drugs
denies tobacco or illicit drug use
denies tobacco, alcohol, drug use
denies tobacco/etoh/illicit drug use
drug abuse:denies
drugs: denies
illicit drug use: denies
illicits:denies
ivda/intranasal: denies
negative for current tobacco, alcohtextcpol, or recreational drug use
recreational drugs: denies
no ivda
no intravenous drug abuse
no iv drug use
no intravenous drug user
no iv drug user
no ivdu
no intravenous drug abuse
no iv drug abuse
no iv drug abuse
no iv drug abuser

Supplemental Table 4. 1 ICD-10-CM codes for CVD conditions

Outcome	ICD-10-CM codes
Atrial Fibrillation	DX I48.0, I48.1, 148.11, 148.19, I48.2, 148.20, 148.21, I48.91 (ONLY first or second DX on the claim)
Acute Myocardial Infarction	DX I21.01, I21.02, I21.09, I21.11, I21.19, I21.21, I21.29, I21.3, I21.4, I21.9, I21.A1, I21.A9, I22.0, I22.1, I22.2, I22.8, I22.9 (ONLY first or second DX on the claim)
Stroke/Transient Ischemic Attack	DX G45.0, G45.1, G45.2, G45.8, G45.9, G46.0, G46.1, G46.2, G46.3, G46.4, G46.5, G46.6, G46.7, G46.8, G97.31, G97.32, I60.00, I60.01, I60.02, I60.10, I60.11, I60.12, I60.20, I60.21, I60.22, I60.30, I60.31, I60.32, I60.4, I60.50, I60.51, I60.52, I60.6, I60.7, I60.8, I60.9, I61.0, I61.1, I61.2, I61.3, I61.4, I61.5, I61.6, I61.8, I61.9, I63.00, I63.02, I63.011, I63.012, I63.013, I63.019, I63.02, I63.031, I63.032, I63.039, I63.09, I63.10, I63.111, I63.112, I63.119, I63.12, I63.131, I63.132, I63.139, I63.19, I63.20, I63.211, I63.212, I63.213, I63.219, I63.22, I63.231, I63.232, I63.233, I63.239, I63.29, I63.30, I63.311, I63.312, I63.313, I63.319, I63.321, I63.322, I63.323, I63.329, I63.331, I63.332, I63.333, I63.339, I63.341, I63.342, I63.343, I63.349, I63.39, I63.40, I63.411, I63.412, I63.413, I63.419, I63.421, I63.422, I63.423, I63.429, I63.431, I63.432, I63.433, I63.439, I63.441, I63.442, I63.443, I63.449, I63.49, I63.50, I63.511, I63.512, I63.513, I63.519, I63.521, I63.522, I63.523, I63.529, I63.531, I63.532, I63.533, I63.539, I63.541, I63.542, I63.543, I63.549, I63.59, I63.6, I63.8, I63.9, I66.01, I66.02, I66.03, I66.09, I66.11, I66.12, I66.13, I66.19, I66.21, I66.22, I66.23, I66.29, I66.3, I66.8, I66.9, I67.841, I67.848, I67.89, I97.810, I97.811, I97.820, I97.821 (any DX on the claim) EXCLUSION: If any of the qualifying claims have any of the following codes in any DX position then EXCLUDE: S01.90XA, S02.0XXA, S02.0XXB, S02.10XA, S02.10XB, S02.101A, S02.101B, S02.102A, S02.102B, S02.109A, S02.109B, S02.11GA, S02.11GB,

	S02.11HA, S02.11HB, S02.110A, S02.111A, S02.112A, S02.113A, S02.110B, S02.111B, S02.112B, S02.113B, S02.118A, S02.118B, S02.119A, S02.119B, S02.121A, S02.121B, S02.121D, S02.121G, S02.121K, S02.121S, S02.122A, S02.122B, S02.122D, S02.122G, S02.122K, S02.122S, S02.129A, S02.129B, S02.129D, S02.129G, S02.129K, S02.129S, S02.19XA, S02.19XB, S02.2XXA, S02.2XXB, S02.3XXA, S02.30XA, S02.3XXB, S02.30XB, S02.31XA, S02.31XB, S02.32XA, S02.32XB, S02.40AA, S02.40AB, S02.40BA, S02.40BB, S02.40CA, S02.40CB, S02.40DA, S02.40DB, S02.40EA, S02.40EB, S02.40FA, S02.40FB, S02.400A, S02.400B, S02.401A, S02.401B, S02.402A, S02.402B, S02.411A, S02.411B, S02.412A, S02.412B, S02.413A, S02.413B, S02.42XA, S02.42XB, S02.600A, S02.600B, S02.601A, S02.601B, S02.602A, S02.602B, S02.609A, S02.609B, S02.61XA, S02.610A, S02.610B, S02.611A, S02.611B, S02.612A, S02.612B, S02.62XA, S02.620A, S02.62XB, S02.620B, S02.621A, S02.621B, S02.622A, S02.622B, S02.63XA, S02.630A, S02.63XB, S02.630B, S02.631A, S02.631B, S02.632A, S02.632B, S02.64XA, S02.640A, S02.64XB, S02.640B, S02.641A, S02.641B, S02.642A, S02.642B, S02.65XA, S02.650A, S02.65XB, S02.650B, S02.651A, S02.651B, S02.652A, S02.652B, S02.66XA, S02.66XB, S02.67XA, S02.670A, S02.670B, S02.671A, S02.671B, S02.672A, S02.672B, S02.69XA, S02.61XB, S02.62XA, S02.63XA, S02.64XA, S02.65XA, S02.66XA, S02.67XB, S02.69XB, S02.8XXA, S02.80XA, S02.8XXB, S02.80XB, S02.81XA, S02.81XB, S02.82XA, S02.82XB
Heart failure	DX I09.81, I11.0, I13.0, I13.2, I50.1, I50.20, I50.21, I50.22, I50.23, I50.30, I50.31, I50.32, I50.33, I50.40, I50.41, I50.42, I50.43, I50.810, I50.811, I50.812, I50.813, I50.814, I50.82, I50.83, I50.84, I50.89, I50.9
Ischemic heart disease	DX I20.0, I20.1, I20.8, I20.9, I21.01, I21.02, I21.09, I21.11, I21.19, I21.21, I21.29, I21.3, I21.4, I21.A1, I21.A9, I22.0, I22.1, I22.2, I22.8, I22.9, I23.0, I23.1, I23.2, I23.3,

	I23.4, I23.5, I23.6, I23.7, I23.8, I24.0, I24.1, I24.8, I24.9, I25.10, I25.110, I25.111, I25.118, I25.119, I25.2, I25.3, I25.41, I25.42, I25.5, I25.6, I25.700, I25.701, I25.708, I25.709, I25.710, I25.711, I25.718, I25.719, I25.720, I25.721, I25.728, I25.729, I25.730, I25.731, I25.738, I25.739, I25.750, I25.751, I25.758, I25.759, I25.760, I25.761, I25.768, I25.769, I25.790, I25.791, I25.798, I25.799, I25.810, I25.811, I25.812, I25.82, I25.83, I25.84, I25.89, I25.9
Infective endocarditis	I33.0, I33.9

Supplemental Table 4. 2 final model for any CVD outcome, model 1, OUD_ICD

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-1.9049	
OUD vs. Non-OUD	0.2326	1.26 (1.18-1.34)
35-54 vs. 18-34	0.6269	1.87 (1.74-2.00)
>= 55 vs. 18-34	1.1826	3.26 (3.06-3.47)
Sex: Male vs. Female	0.1789	1.19 (1.16-1.22)
Race: White vs. Other	0.041	1.04 (0.99-1.08)
Non-Hispanic vs. others	-0.0778	0.92 (0.89-0.95)
Diabetes	0.1947	1.21 (1.18-1.24)
Hypertension	0.0965	1.10 (1.07-1.13)
Visit counts	0.0959	1.10 (1.08-1.11)

* PRR: Prevalence rate ratio

Supplemental Table 4. 3 final model for any CVD outcome, model 2, OUD_(ICD+NLP)

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-1.9053	
OUD vs. Non-OUD	0.1998	1.22 (1.15-1.28)
35-54 vs. 18-34	0.6249	1.86 (1.74-1.99)
>= 55 vs. 18-34	1.1812	3.25 (3.05-3.47)
Sex: Male vs. Female	0.1788	1.19 (1.16-1.22)
Race: White vs. Other	0.0421	1.04 (0.99-1.09)
Non-Hispanic vs. others	-0.0771	0.92 (0.89-0.95)
Diabetes	0.195	1.21 (1.18-1.24)
Hypertension	0.0961	1.10 (1.07-1.13)
Visit counts	0.0951	1.09 (1.08-1.11)

* PRR: Prevalence rate ratio

Supplemental Table 4. 4 final model for Cardiac arrhythmia, model 1, OUD_ICD

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-2.1851	0.11 (0.10-0.12)
OUD vs. Non-OUD	0.3434	1.40 (1.30-1.52)
35-54 vs. 18-34	0.4605	1.58 (1.46-1.71)
>= 55 vs. 18-34 (2)	0.9982	2.71 (2.52-2.92)
Sex: Male vs. Female	0.2343	1.26 (1.21-1.31)
Race: White vs. Other	0.0782	1.08 (1.01-1.15)
Non-Hispanic vs. others	-0.0879	0.91 (0.87-0.96)
Diabetes	0.1353	1.14 (1.09-1.19)
Hypertension	0.0333	1.03 (0.99-1.07)
Visit counts	0.1328	1.14 (1.11-1.16)

* PRR: Prevalence rate ratio

Supplemental Table 4. 5 final model for Cardiac arrhythmia, model 2, OUD_(ICD+NLP)

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-2.186	0.11 (0.10-0.12)
OUD vs. Non-OUD	0.289	1.33 (1.24-1.43)
35-54 vs. 18-34	0.4576	1.58 (1.46-1.71)
>= 55 vs. 18-34 (2)	0.9949	2.70 (2.51-2.91)
Sex: Male vs. Female	0.2345	1.26 (1.21-1.31)
Race: White vs. Other	0.0802	1.08 (1.01-1.15)
Non-Hispanic vs. others	-0.0864	0.91 (0.87-0.96)
Diabetes	0.1354	1.14 (1.09-1.19)
Hypertension	0.0325	1.03 (0.99-1.07)
Visit counts	0.1324	1.14 (1.11-1.16)

* PRR: Prevalence rate ratio

Supplemental Table 4. 6 final model for MI, model 1, OUD_ICD

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-5.1376	
OUD vs. Non-OUD	-0.0619	0.93 (0.70-1.25)
35-54 vs. 18-34	1.6739	5.33 (3.63-7.82)
>= 55 vs. 18-34 (2)	2.3966	10.9 (7.56-15.9)
Sex: Male vs. Female	0.4273	1.53 (1.36-1.71)
Race: White vs. Other	-0.0364	0.96 (0.80-1.16)
Non-Hispanic vs. others	-0.3474	0.70 (0.61-0.81)
Diabetes	0.456	1.57 (1.38-1.79)
Hypertension	-0.0825	0.92 (0.81-1.03)
Visit counts	0.1477	1.15 (1.12-1.19)

* PRR: Prevalence rate ratio

Supplemental Table 4. 7 final model for MI, model 2, OUD_(ICD+NLP)

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-5.1355	
OUD vs. Non-OUD	-0.0885	0.91 (0.70-1.18)
35-54 vs. 18-34	1.6751	5.33 (3.63-7.84)
>= 55 vs. 18-34 (2)	2.3936	10.9 (7.53-15.9)
Sex: Male vs. Female	0.4274	1.53 (1.36-1.71)
Race: White vs. Other	-0.0355	0.96 (0.80-1.16)
Non-Hispanic vs. others	-0.3475	0.70 (0.61-0.81)
Diabetes	0.4554	1.57 (1.38-1.79)
Hypertension	-0.0822	0.92 (0.81-1.03)
Visit counts	0.1488	1.16 (1.13-1.19)

* PRR: Prevalence rate ratio

Supplemental Table 4. 8 final model for stroke, model 1, OUD_ICD

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-3.4245	
OUD vs. Non-OUD	-0.0388	0.96 (0.80-1.15)
35-54 vs. 18-34	0.8091	2.24 (1.86-2.70)
>= 55 vs. 18-34 (2)	1.4382	4.21 (3.52-5.03)
Sex: Male vs. Female	0.0116	1.01 (0.94-1.08)
Race: White vs. Other	0.1103	1.11 (0.98-1.26)
Non-Hispanic vs. others	-0.3786	0.68 (0.63-0.74)
Diabetes	0.2453	1.27 (1.18-1.38)
Hypertension	0.6048	1.83 (1.68-1.98)
Visit counts	0.0193	1.01 (0.98-1.05)

* PRR: Prevalence rate ratio

Supplemental Table 4. 9 final model for stroke, model 2, OUD_(ICD+NLP)

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-3.4189	
OUD vs. Non-OUD	-0.1286	0.87 (0.74-1.03)
35-54 vs. 18-34	0.8098	2.24 (1.86-2.70)
>= 55 vs. 18-34 (2)	1.429	4.17 (3.49-4.98)
Sex: Male vs Female	0.0123	1.01 (0.94-1.08)
Race: White vs. Other	0.113	1.11 (0.98-1.26)
Non-Hispanic vs others	-0.3786	0.68 (0.63-0.74)
Diabetes	0.2436	1.27 (1.17-1.38)
Hypertension	0.6051	1.83 (1.68-1.98)
Visit counts	0.0222	1.02 (0.98-1.05)

* PRR: Prevalence rate ratio

Supplemental Table 4. 10 final model for heart failure, model 1, OUD_ICD

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-3.7558	
OUD vs. Non-OUD	0.143	1.15 (0.98-1.35)
35-54 vs. 18-34	1.5364	4.64 (3.79-5.68)
>= 55 vs. 18-34	2.5906	13.3 (11.0-16.1)
Sex: Male vs. Female	0.1972	1.21 (1.14-1.29)
Race: White vs. Other	-0.2089	0.81 (0.73-0.89)
Non-Hispanic vs. other	-0.1148	0.89 (0.82-0.96)
Diabetes	0.6685	1.95 (1.82-2.08)
Hypertension	-1.1844	0.30 (0.28-0.32)
Visit counts	0.165	1.17 (1.14-1.21)

* PRR: Prevalence rate ratio

Supplemental Table 4. 11 final model for heart failure, model 2, OUD_(ICD+NLP)

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-3.7604	
OUD vs. Non-OUD	0.1783	1.19 (1.04-1.36)
35-54 vs. 18-34	1.5338	4.63 (3.79-5.66)
>= 55 vs. 18-34	2.5955	13.4 (11.0-16.2)
Sex: Male vs. Female	0.197	1.21 (1.14-1.29)
Race: White vs. Other	-0.2097	0.81 (0.73-0.88)
Non-Hispanic vs. others	-0.115	0.89 (0.82-0.96)
Diabetes	0.6697	1.95 (1.83-2.08)
Hypertension	-1.1852	0.30 (0.28-0.32)
Visit counts	0.1634	1.17 (1.14-1.21)

* PRR: Prevalence rate ratio

Supplemental Table 4. 12 final model for ISCHEMICHD, model 1, OUD_ICD

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-5.6587	
OUD vs. Non-OUD	-0.2211	0.80 (0.68-0.93)
35-54 vs. 18-34	2.7935	16.3 (11.1-23.9)
>= 55 vs. 18-34	3.8904	48.9 (33.4-71.4)
Sex: Male vs. Female	0.4225	1.52 (1.45-1.60)
Race: White vs. Other	0.2366	1.26 (1.15-1.39)
Non-Hispanic vs. others	-0.0165	0.98 (0.92-1.05)
Diabetes	0.4342	1.54 (1.46-1.62)
Hypertension	0.0158	1.01 (0.96-1.06)
Visit counts	0.1041	1.10 (1.08-1.12)

* PRR: Prevalence rate ratio

Supplemental Table 4. 13 final model for ISCHEMICHD, model 2, OUD_(ICD+NLP)

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-5.6592	
OUD vs. Non-OUD	-0.1661	0.84 (0.74-0.96)
35-54 vs. 18-34	2.7951	16.3 (11.1-24.0)
>= 55 vs. 18-34	3.893	49.0 (33.5-71.6)
Sex: Male vs Female	0.4222	1.52 (1.45-1.60)
Race: White vs. Other	0.2355	1.26 (1.15-1.39)
Non-Hispanic vs others	-0.0172	0.98 (0.92-1.04)
Diabetes	0.434	1.54 (1.46-1.62)
Hypertension	0.016	1.01 (0.96-1.06)
Visit counts	0.1043	1.10 (1.08-1.13)

* PRR: Prevalence rate ratio

Supplemental Table 4. 14 final model for IE, model 1, OUD_ICD

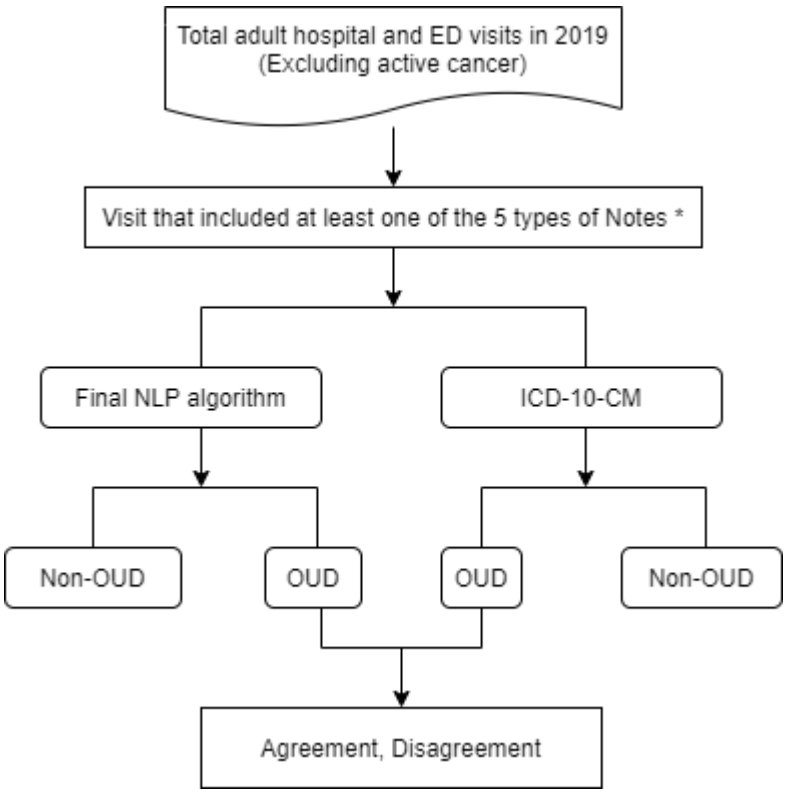
	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-6.5242	
OUD vs. Non-OUD	2.9287	18.7 (6.95-50.3)
35-54 vs. 18-34	0.0396	1.04 (0.41-2.62)
>= 55 vs. 18-34	-1.4034	0.24 (0.05-1.20)
Sex: Male vs Female	0.0311	1.03 (0.43-2.43)
Non-Hispanic vs others	-1.1903	0.30 (0.12-0.72)
Diabetes	-0.6684	0.51 (0.05-4.66)
Hypertension	0.5439	1.72 (0.70-4.19)
Visit counts	-0.2464	0.78 (0.34-1.77)

* PRR: Prevalence rate ratio

Supplemental Table 4. 15 final model for IE, model 2, OUD_(ICD+NLP)

	Coefficient Estimate	Adjusted PRR* (95% CI)
Intercept	-6.5919	
OUD vs. Non-OUD	2.9326	18.7 (6.80-51.8)
35-54 vs. 18-34 (Keller et al.)	0.0328	1.03 (0.40-2.63)
>= 55 vs. 18-34	-1.4034	0.24 (0.05-1.14)
Sex: Male vs. Female	-0.0232	0.97 (0.40-2.33)
Non-Hispanic vs. others	-1.1904	0.30 (0.12-0.73)
Diabetes	-0.6501	0.52 (0.05-4.68)
Hypertension	0.5323	1.70 (0.70-4.12)
Visit counts	-0.2736	0.76 (0.33-1.72)

* PRR: Prevalence rate ratio



* Discharge Summary, History and Physical, ED General, ED Triage, Addiction Medicine Consult

Supplemental Figure 2. 1 Diagram of applying NLP algorithm to testing data set

(This page is optional. Delete this page and the following page break if you don't need it).

[APPENDIX 3. EVALUATION GUIDE]

REFERENCES

Afzal, N., S. Sohn, S. Abram, H. Liu, F.J. Kullo, and A.M. Arruda-Olson. 2016. Identifying peripheral arterial disease cases using natural language processing of clinical notes.

Afzal, N., S. Sohn, S. Abram, C.G. Scott, R. Chaudhry, H. Liu, I.J. Kullo, and A.M. Arruda-Olson. 2017. Mining peripheral arterial disease cases from narrative clinical notes using natural language processing. *J Vasc Surg*. 65:1753-1761.

Aghadavoudi, O., N. Eizadi-Mood, and M.R. Najarzadegan. 2015. Comparing cardiovascular factors in opium abusers and non-users candidate for coronary artery bypass graft surgery. *Adv Biomed Res*. 4.

Alzeer, A., J. Patel, B. Dixon, M. Bair, and J. Jones. 2018. A Comparison Between Two Approaches to Identify Opioid Use Problems: ICD-9 vs. Text-Mining Approach. *In* 2018 IEEE International Conference on Healthcare Informatics (ICHI). 455-456.

Baclic, O., M. Tunis, K. Young, C. Doan, H. Swerdfeger, and J. Schonfeld. 2020. Challenges and opportunities for public health made possible by advances in natural language processing. *Can Commun Dis Rep*. 46:161-168.

Barkin, R., S. Barkin, and D. Barkin. 2006. <Barkin 2006 Propoxyphene A critical review of a weak opioid analgesic.pdf>.

Barros, A.J., and V. Hirakata. 2003. Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology*. 3.

Bates, M.C., F. Annie, A. Jha, and F. Kerns. 2019. Increasing incidence of IV-drug use associated endocarditis in southern West Virginia and potential economic impact. *Clin Cardiol*. 42:432-437.

Beam, K.S., M. Lee, K. Hirst, A. Beam, and R.B. Parad. 2021. Specificity of International Classification of Diseases codes for bronchopulmonary dysplasia: an investigation using electronic health record data and a large insurance database. *J Perinatol*. 41:764-771.

Beck, D.L. 2019. Cardiology and the drug abuse crisis: Points of intersection. *Cardiology Magazine*.

Behzadi, M., S. Joukar, and A. Beik. 2018. Opioids and Cardiac Arrhythmia: A Literature Review. *Med Princ Pract*. 27:401-414.

Bernard, S.A., P.R. Chelminski, T.J. Ives, and S.I. Ranapurwala. 2018. Management of Pain in the United States-A Brief History and Implications for the Opioid Epidemic. *Health Serv Insights*. 11:1178632918819440.

Bird, S., Klein, E., & Loper, E. . 2009. Natural language processing with Python: analyzing text with the natural language toolkit. *O' Reilly Media, Inc.*

Birman-Deych, E., A.D. Waterman, Y. Yan, D.S. Nilasena, M.J. Radford, and B.F. Gage. 2005 Accuracy of ICD-9-CM Codes for Identifying Cardiovascular and Stroke Risk Factors. *Medical Care*. 43:480-485.

Carman, W.J., S. Su, S.F. Cook, J.I. Wurzelmann, and A. McAfee. 2011. Coronary heart disease outcomes among chronic opioid and cyclooxygenase-2 users compared with a general population cohort. *Pharmacoepidemiol Drug Saf*. 20:754-762.

Carrell, D.S., D. Cronkite, R.E. Palmer, K. Saunders, D.E. Gross, E.T. Masters, T.R. Hylan, and M. Von Korff. 2015. Using natural language processing to identify problem usage of prescription opioids. *Int J Med Inform*. 84:1057-1064.

Cicero, T.J., M.S. Ellis, and Z.A. Kasper. 2017. Increased use of heroin as an initiating opioid of abuse. *Addict Behav*. 74:63-66.

Cochran, G., B. Woo, W.H. Lo-Ciganic, A.J. Gordon, J.M. Donohue, and W.F. Gellad. 2015. Defining Nonmedical Use of Prescription Opioids Within Health Care Claims: A Systematic Review. *Subst Abus*. 36:192-202.

Dakour-Aridi, H., M. Arora, B. Nejim, S. Locham, and M.B. Malas. 2019. Association between Drug Use and In-hospital Outcomes after Infrainguinal Bypass for Peripheral Arterial Occlusive Disease. *Ann Vasc Surg*. 58:122-133 e124.

Datta, S., E.V. Bernstam, and K. Roberts. 2019. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. *J Biomed Inform*. 100:103301.

Dowell, D., H. T., and R. Chou. 2016. CDC Guideline for Prescribing Opioids for Chronic Pain — United States, 2016.

Ford, E., J.A. Carroll, H.E. Smith, D. Scott, and J.A. Cassell. 2016. Extracting information from the text of electronic medical records to improve case detection: a systematic review. *J Am Med Inform Assoc*. 23:1007-1015.

Frolova, N., J.A. Bakal, F.A. McAlister, B.H. Rowe, H. Quan, P. Kaul, and J.A. Ezekowitz. 2015. Assessing the use of international classification of diseases-10th revision codes from the emergency department for the identification of acute heart failure. *JACC Heart Fail*. 3:386-391.

Geidrimiene, D., and K. King. 2017. Burden of Cardiovascular Disease (CVD) on Economic Cost. Comparison of Outcomes in US and Europe. *Circulation: Cardiovascular Quality and Outcomes*.

Gliklich, R.E., M.B. Leavy, and N.A. Dreyer. 2019. Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes A User's Guide, 3rd Edition, Addendum 2. *L&M Policy Research*. DOI: https://doi.org/10.23970/AHRQEPCREGISTRIES3ADDENDUM2.

Gray, M.E., E.T. Rogawski McQuade, W.M. Scheld, and R.A. Dillingham. 2018. Rising rates of injection drug use associated infective endocarditis in Virginia with missed opportunities for addiction treatment referral: a retrospective cohort study. *BMC Infect Dis*. 18:532.

Gupta, T., M. Mujib, P. Agarwal, P. Prakash, A. Garg, N. Sharma, W. Aronow, and C. Nabors. 2016. <Gupta 2016 - Association Between Opioid Abuse Dependence and Outcomes in Hospitalized Heart Failure Patients.pdf>.

Haight, S., J. Ko, V. Tong, M. Bohm, and W. Callaghan. 2018. <Haight 2018 Opioid use disorder documented at delivery hospitalization 199-2014.pdf>.

Hartman, L., E. Barnes, L. Bachmann, K. Schafer, J. Lovato, and D.C. Files. 2016. Opiate Injection-associated Infective Endocarditis in the Southeastern United States. *Am J Med Sci*. 352:603-608.

Hedegaard, H., A. Miniño, and M. Warner. 2018. <Hedegaard - 2018 Drug overdose death in the united state, 1999-2018.pdf>.

Hernan, M.A., and J.M. Robins. 2020. Causao inference what if.

Honnibal, M.M., I. 2017. spaCy: Industrial-strength Natural Language Processing in Python.

Hughes Garza, H., K.E. Piper, A.N. Barczyk, A. Perez, and K.A. Lawson. 2021. Accuracy of ICD-10-CM coding for physical child abuse in a paediatric level I trauma centre. *Inj Prev*. 27:i71-i74.

Jain, V., M.H. Yang, G. Kovacicova-Lezcano, L.S. Juhle, A.F. Bolger, and L.G. Winston. 2008. Infective endocarditis in an urban medical center: association of individual drugs with valvular involvement. *J Infect*. 57:132-138.

Jensen, P.N., K. Johnson, J. Floyd, S.R. Heckbert, R. Carnahan, and S. Dublin. 2012. A systematic review of validated methods for identifying atrial fibrillation using administrative data. *Pharmacoepidemiol Drug Saf*. 21 Suppl 1:141-147.

Jobski, K., B. Kollhorst, E. Garbe, and T. Schink. 2017. The Risk of Ischemic Cardio- and Cerebrovascular Events Associated with Oxycodone-Naloxone and Other Extended-Release High-Potency Opioids: A Nested Case-Control Study. *Drug Saf*. 40:505-515.

Kaye, A.D., M.R. Jones, A.M. Kaye, J.G. Ripoll, V. Galan, B.D. Beakley, F. Calixto, J.L. Bolden, R.D. Urman, and L. Manchikanti. 2017. Prescription Opioid Abuse in Chronic Pain An updated review of opioid abuse predictors and strategies to curb opioid abuse Part1. *Pain Physician*. 20.

Keeshin, S.W., and J. Feinberg. 2016. Endocarditis as a Marker for New Epidemics of Injection Drug Use. *Am J Med Sci*. 352:609-614.

Keller, G., P. Alvarez, M. Ponte, W. Belloso, C. Bagnes, Sparanochia7, M. Gonzalez1, Etchegoyen1, and G.D. Girolamo. 2016a. <Keller 2016 Drug induced QTc interval prolongation.pdf>.

Keller, G., P. Alvarez, M. Ponte, W. Belloso, C. Bagnes, C. Sparanochia, C. Gonzalez, M. Etchegoyen, R. Diez, and G. Girolamo. 2016b. <Keller 2016 - Drug-Induced QTc Interval Prolongation A Multicenter Study to Detect Drugs and Clinical Factors Involved in Every Day Practice.pdf>.

Khodneva, Y., P. Muntner, S. Kertesz, B. Kissela, and M.M. Safford. 2016. Prescription Opioid Use and Risk of Coronary Heart Disease, Stroke, and Cardiovascular Death Among Adults from a Prospective Cohort (REGARDS Study). *Pain Med*. 17:444-455.

Kivimaki, M., G.D. Batty, A. Singh-Manoux, A. Britton, E.J. Brunner, and M.J. Shipley. 2017. Validity of Cardiovascular Disease Event Ascertainment Using Linkage to UK Hospital Records. *Epidemiology*. 28:735-739.

Kurbasic, I., H. Pandza, I. Masic, S. Huseinagic, S. Tandir, F. Alicajic, and S. Toromanovic. 2008. The advantages and limitations of international classification of diseases, injuries and causes of death from aspect of existing health care system of bosnia and herzegovina. *Acta Inform Med*. 16:159-161.

Larney, S.e.a. 2020 All-Cause and Cause-Specific Mortality Among People Using Extramedical Opioids: A Systematic Review and Meta-analysis. *JAMA Psychiatry*

77.

Lee, C.W., C.H. Muo, J.A. Liang, F.C. Sung, and C.H. Kao. 2013. Association of intensive morphine treatment and increased stroke incidence in prostate cancer

patients: a population-based nested case-control study. *Jpn J Clin Oncol*. 43:776-781.
Lee, S.C., W. Klein-Schwartz, S. Doyon, and C. Welsh. 2014. Comparison of toxicity associated with nonmedical use of benzodiazepines with buprenorphine or methadone. *Drug Alcohol Depend*. 138:118-123.
Lentine, K.L., N.N. Lam, H. Xiao, J.E. Tuttle-Newhall, D. Axelrod, D.C. Brennan, V.R. Dharnidharka, H. Yuan, M. Nazzal, J. Zheng, and M.A. Schnitzler. 2015. Associations of pre-transplant prescription narcotic use with clinical complications after kidney transplantation. *Am J Nephrol*. 41:165-176.
Lewer, D., M. Harris, and V. Hope. 2017. Opiate Injection-Associated Skin, Soft Tissue, and Vascular Infections, England, UK, 1997-2016. *Emerg Infect Dis*. 23:1400-1403.
Li, L., S. Setoguchi, H. Cabral, and S. Jick. 2013. Opioid use for noncancer pain and risk of myocardial infarction amongst adults. *J Intern Med*. 273:511-526.
Li, Y., and T.J. VanderWeele. 2020. Direct Effects under Differential Misclassification in Outcomes, Exposures, and Mediators. *Journal of Causal Inference*. 8:286-299.
Maloney, E., Degenhardt, L., Darke, S., Mattick, R. P. & Nelson, E. 2007. Suicidal behaviour and associated risk factors among opioid-dependent individuals: a case-control study. *Addiction*. 102:1933–1941.
Marmor, M., A. Penn, K. Widmer, R.I. Levin, and R. Maslansky. 2004. Coronary artery disease and opioid use. *Am J Cardiol*. 93:1295-1297.
Martins, S.S., L.E. Segura, J. Santaella-Tenorio, A. Perlmutter, M.C. Fenton, M. Cerda, K.M. Keyes, L.A. Ghandour, C.L. Storr, and D.S. Hasin. 2017. Prescription opioid use disorder and heroin use among 12-34 year-olds in the United States from 2002 to 2014. *Addict Behav*. 65:236-241.
Max, M.B. 1990. Improving outcomes of analgesic treatment: is education enough? *Ann Intern Med.* . 113.
Maxwell, J.C. 2011. The prescription drug epidemic in the United States: A perfect storm. D. *Drug and Alcohol Review*. 30.
Meisner, J.A., J. Anesi, X. Chen, and D. Grande. 2019. Changes in infective endocarditis admissions in Pennsylvania during the opioid epidemic. *Clin Infect Dis*.
Meldrum, M.L. 2016. The Ongoing Opioid Prescription Epidemic: Historical Context. *Am J Public Health*. 106:1365-1366.
Menendez, M.E., D. Ring, and B.T. Bateman. 2015. Preoperative Opioid Misuse is Associated With Increased Morbidity and Mortality After Elective Orthopaedic Surgery. *Clin Orthop Relat Res*. 473:2402-2412.
Mensah, G.A., and D.W. Brown. 2007. An overview of cardiovascular disease burden in the United States. *Health Aff (Millwood)*. 26:38-48.
Metcalfe, A., A. Neudam, S. Forde, M. Liu, S. Drosler, H. Quan, and N. Jette. 2013. Case definitions for acute myocardial infarction in administrative databases and their impact on in-hospital mortality rates. *Health Serv Res*. 48:290-318.
Meystre, S.M., G.K. Savova, K.C. Kipper-Schuler, and J.F. Hurdle. 2008. Extracting Information from Textual Documents in the electronic health record A review of recent research. *IMIA Yearbook of Medical Informatics*. 47 Suppl 1.
Mezzich, J.E. 2002. International Surveys on the Use of ICD-10 and Related Diagnostic Systems. *Psychopathology*. 35.

Mihm, A.E., M.C. Cash, S.A. Nisly, and K.A. Davis. 2020. Increased Awareness Needed for Inpatient Substance Use Disorder Treatment in Drug Use-Associated Infective Endocarditis. *J Gen Intern Med*. 35:2228-2230.

Mirzaiepour, F.D., M.; Forood, A.; Najafipour, H.; Shokoohi, M. 2012. <Mirzaiepour 2012 - The effect of opium addiction on arrhythmia following acute myocardial infarction.pdf>.

Mylonakis, E., and S. Calderwood. 2001. <Mylonakis 2001 - Infective endocarditis in adulsts.pdf>. *The New England Journal of Medicine*. 345.

Nadkarni, P.M., L. Ohno-Machado, and W.W. Chapman. 2011. Natural language processing: an introduction. *J Am Med Inform Assoc*. 18:544-551.

National Academies of Sciences, E., and Medicine. 2020. Opportunities to Improve Opioid Use Disorder and Infectious Disease Services: Integrating Responses to a Dual Epidemic. The National Academies Press, Washington, DC. 216 pp.

Neoplasms. 2021. C00-D49: https://www.icd10data.com/ICD10CM/Codes/C00-D49.

O'Malley, K.J., K.F. Cook, M.D. Price, K.R. Wildes, J.F. Hurdle, and C.M. Ashton. 2005. Measuring diagnoses: ICD code accuracy. *Health Serv Res*. 40:1620-1639.

Omran, S., A. Chatterjee, M.L. Chen, M.P. Lerario, A.E. Merkler, and H. Kamel. 2019. National Trends in Hospitalizations for Stroke Associated With Infective Endocarditis and Opioid Use Between 1993 and 2015. *Stroke*. 50:577-582.

Palumbo, S.A., K.M. Adamson, S. Krishnamurthy, S. Manoharan, D. Beiler, A. Seiwell, C. Young, R. Metpally, R.C. Crist, G.A. Doyle, T.N. Ferraro, M. Li, W.H. Berrettini, J.D. Robishaw, and V. Troiani. 2020. Assessment of Probable Opioid Use Disorder Using Electronic Health Record Documentation. *JAMA Netw Open*. 3:e2015909.

Paulozzi, L.J., R.H. Weisler, and A.A. Patkar. 2011. A national epidemic of unintentional prescription opioid overdose deaths: how physicians can help control it. *J Clin Psychiatry*. 72:589-592.

Pendergrass, S.A., and D.C. Crawford. 2019. Using Electronic Health Records To Generate Phenotypes For Research. *Curr Protoc Hum Genet*. 100:e80.

Phillips, D.M. 2000. JCAHO pain management standards are unveiled. Joint Commission on Accreditation of Healthcare Organizations. *JAMA*. 284.

Pierce, M., Bird, S. M., Hickman, M. & Millar, T. 2015. National record linkage study of mortality for a large cohort of opioid users ascertained by drug treatment or criminal justice sources in England. *Drug Alcohol Depend*. 146:146, 117–123

Portenoy, R.K., and K.M. Foley. 1986. Chronic use of opioid analgesics in non-malignant pain: report of 38 cases. . *Pain*. 25.

Porter, J., and H. Jick. 1980. Addiction rare in patients treated with narcotics. *N Engl J Med*. 302.

Quan, H., B. Li, L.D. Saunders, G.A. Parsons, C.I. Nilsson, A. Alibhai, W.A. Ghali, and I. Investigators. 2008. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res*. 43:1424-1441.

Register, F. 2014. Administrative Simplification: Change to the Compliance Date for the International Classification of Diseases, 10th Revision (ICD-10-CM and ICD-10-PCS) Medical Data Code Sets.

Roberto, G., M. Simonetti, C. Piccinni, P. Lora Aprile, I. Cricelli, A. Fanelli, C. Cricelli, and F. Lapi. 2015. Risk of Acute Cerebrovascular and Cardiovascular Events Among Users of Acetaminophen or an Acetaminophen-Codeine Combination in a Cohort of Patients with Osteoarthritis: A Nested Case-Control Study. *Pharmacotherapy*. 35:899-909.

Rosamond, W.D., P.P. Chang, C. Baggett, A. Johnson, A.G. Bertoni, E. Shahar, A. Deswal, G. Heiss, and L.E. Chambless. 2012. Classification of heart failure in the atherosclerosis risk in communities (ARIC) study: a comparison of diagnostic criteria. *Circ Heart Fail*. 5:152-159.

Rose, G. 2005. Incubation period of coronary heart disease. 1982. *Int J Epidemiol*. 34:242-244.

Sadeghian, S., A. Karimi, S. Dowlastshahi, S. Ahmadi, S. Davoodi, M. Marzban, N. Movahedi, K. Abbasi, M. Tazik, and M. Fathollahi. 2009. <Sadeghian 2009 - The association of opium dependence and postoperative complications following coronary artery bypass graft surgery a propensity-matched study.pdf>.

Schultz, J.E., and G.J. Gross. 2001. Opioids and cardioprotection. *Pharmacol Ther*. 89.

Sen, A., I. Vardaxis, B.H. Lindqvist, B.M. Brumpton, L.B. Strand, I.J. Bakken, L.J. Vatten, P.R. Romundstad, R. Ljung, K.J. Mukamal, and I. Janszky. 2019. Systematic assessment of prescribed medications and short-term risk of myocardial infarction - a pharmacopeia-wide association study from Norway and Sweden. *Sci Rep*. 9:8257.

Sheikhalishahi, S., R. Miotto, J.T. Dudley, A. Lavelli, F. Rinaldi, and V. Osmani. 2019. Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. *JMIR Med Inform*. 7:e12239.

Singh, J.A., and J.D. Cleveland. 2020. National U.S. time-trends in opioid use disorder hospitalizations and associated healthcare utilization and mortality. *PLoS One*. 15:e0229174.

Singleton, J.H., E.L. Abner, P.D. Akpunonu, and A.M. Kucharska-Newton. 2021. Association of Nonacute Opioid Use and Cardiovascular Diseases: A Scoping Review of the Literature. *J Am Heart Assoc*. 10:e021260.

Sinner, G.J., R. Annabathula, V. Karolina, and T. Alnabelsi. 2021. Infective endocarditis in pregnancy from 2009 to 2019: the consequences of injection drug use. *Infectious Diseases*. 53.

Smart, R., C.A. Kase, E.A. Taylor, S. Lumsden, S.R. Smith, and B.D. Stein. 2020. Strengths and weaknesses of existing data sources to support research to address the opioids crisis. *Prev Med Rep*. 17:101015.

Solomon, D., J. Rassen, R. Glynn, K. Garneau, R. Levin, and S. Schneeweiss. 2010. <Solomon 2010 - The comparative safety of opioids for nonmalignant pain in older adult 2010.pdf>.

Sorahan, T., and M.S. Gilthorpe. 1994. Non-differential misclassification of exposure always leads to an underestimate of risk: an incorrect conclusion. *Occupational and Environmental Medicine* 51.

Spasic, I., and G. Nenadic. 2020. Clinical Text Data in Machine Learning: Systematic Review. *JMIR Med Inform*. 8:e17984.

Strang, J., N.D. Volkow, L. Degenhardt, M. Hickman, K. Johnson, G.F. Koob, B.D.L. Marshall, M. Tyndall, and S.L. Walsh. 2020. Opioid use disorder. *Nat Rev Dis Primers*. 6:3.

Suryaprasad, A.G. 2014. Emerging epidemic of hepatitis C virus infections among young nonurban persons who inject drugs in the United States,. *Clin. Infect. Dis*. 59:1411–1419.

Tanaka, K., R.K. Judy, and L. Matthias. 2014. Opioid-induced Cardioprotection. *Current Pharmaceutical Design*. 20.

Tian, Z., S. Sun, T. Eguale, and C.M. Rochefort. 2017. <Tian 2017 Automated Extraction of VTE Events From Narrative.pdf>. *Medical Care*. 55.

Velupillai, S., H. Suominen, M. Liakata, A. Roberts, A.D. Shah, K. Morley, D. Osborn, J. Hayes, R. Stewart, J. Downs, W. Chapman, and R. Dutta. 2018. Using clinical Natural Language Processing for health outcomes research: Overview and actionable suggestions for future advances. *J Biomed Inform*. 88:11-19.

Volkow, N.D. 2016. Opioids in pregnancy. *BMJ*. 352.

Vozoris, N.T., X. Wang, P.C. Austin, D.S. Lee, A.L. Stephenson, D.E. O'Donnell, S.S. Gill, and P.A. Rochon. 2017. Adverse cardiac events associated with incident opioid drug use among older adults with COPD. *Eur J Clin Pharmacol*. 73:1287-1295.

Wang, Y., L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng, S. Mehrabi, S. Sohn, and H. Liu. 2018. Clinical information extraction applications: A literature review. *J Biomed Inform*. 77:34-49.

Weir, M.A., J. Slater, R. Jandoc, S. Koivu, A.X. Garg, and M. Silverman. 2019. The risk of infective endocarditis among people who inject drugs: a retrospective, population-based time series analysis. *CMAJ*. 191:E93-E99.

Whitcomb, B.W., and A.I. Naimi. 2020. Things Don't Always Go as Expected: The Example of Nondifferential Misclassification of Exposure-Bias and Error. *Am J Epidemiol*. 189:365-368.

Woller, B., A. Daw, V. Aston, J. Lloyd, G. Snow, S.M. Stevens, S.C. Woller, P. Jones, and J. Bledsoe. 2021. Natural Language Processing Performance for the Identification of Venous Thromboembolism in an Integrated Healthcare System. *Clin Appl Thromb Hemost*. 27:10760296211013108.

Wong, S.S., and C.W. Cheung. 2020. Optimization of opioid utility in cancer pain populations. *Ann Palliat Med*. 9:558-570.

Wright, A., A.B. McCoy, S. Henkin, A. Kale, and D.F. Sittig. 2013. Use of a support vector machine for categorizing free-text notes: assessment of accuracy across two institutions. *J Am Med Inform Assoc*. 20:887-890.

Zagaria, M.E. 2018. Cardiovascular considerations with prescription opioids and chronic pain. *US Pharm*. 43.

Zheng, L., Y. Wang, S. Hao, A.Y. Shin, B. Jin, A.D. Ngo, M.S. Jackson-Browne, D.J. Feller, T. Fu, K. Zhang, X. Zhou, C. Zhu, D. Dai, Y. Yu, G. Zheng, Y.M. Li, D.B. McElhinney, D.S. Culver, S.T. Alfreds, F. Stearns, K.G. Sylvester, E. Widen, and X.B. Ling. 2016. Web-based Real-Time Case Finding for the Population Health Management of Patients With Diabetes Mellitus: A Prospective Validation of the Natural Language Processing-Based Algorithm With Statewide Electronic Medical Records. *JMIR Med Inform*. 4:e37.

Zibbell, J.E. 2015. Increases in hepatitis C virus infection related to injection drug use among persons aged ≤30 years – Kentucky, Tennessee, Virginia, and West Virginia 2006–2012. *MMWR Morb. Mortal. Wkly Rep.* 64:453–458.

VITA

Education

PhD Epidemiology and Biostatistics, University of Kentucky, August 2014 – December 2021.

Dissertation: Assessing misclassification bias in the study of opioids and cardiovascular disease: Application of natural language processing (NLP) to electronic health records

MS Biostatistics, University of Kentucky, August 2009 – December 2011

Thesis: Statistical analysis of metabolite concentrations in heart tissue from four groups of mouse models in response to Adriamycin

PhD. in Analytical Chemistry, Fudan University, China, September 2001 – December 2004

Dissertation: Proteomics Study on Atherosclerosis Models and Related Methodology

Work Experience

Enterprise Data Specialist March 2017 – Present
University of Kentucky HealthCare IT Business Intelligence

Statistician August 2014 – April 2017
Kentucky Injury Prevention and Research Center, University of Kentucky

Post-Doc September 2005 – December 2008
Dept. of Anatomy and Neuroscience, University of Kentucky

Selected Publications

Jade S., Erin A., Peter A., Anna K., Association of Nonacute Opioid Use and Cardiovascular Diseases: A Scoping Review of the Literature, Journal of the American Heart Association. 2021;10: e021260

Akundi RS, Huang, Z., et al., Increased mitochondrial calcium sensitivity and abnormal expression of innate immunity genes precede dopaminergic defects in Pink1-deficient mice. PLoS One. 2011 Jan 13;6(1)

Huang, Z., Yang, P., Almofti, M., Yu, Y., et al., Comparative analysis of the proteome in rat coronary atherosclerotic lesions, Life Science, 2004, 75:3103-3115.

Yu, Y., Huang, Z., Yang, P. et al., Proteomic studies of macrophage-derived foam cell from human U937 cell line using 2D gel electrophoresis and tandem mass spectrometry J. Cardiovasc Pharmacol., 6 (42), 2003:1