Janssen Research & Development*

Study Protocol for Retrospective Observational Studies Using Secondary Data

Comparing the Estimated Risk of Hip Fracture Among Subjects Exposed to Tramadol as Compared to Subjects Exposed to Codeine

* Janssen Research & Development (Janssen R&D) is a global organization that operates through different legal entities in various countries. Therefore, the legal entity acting as the sponsor for studies of Janssen R&D may vary. The term "sponsor" is used throughout the protocol to represent these various legal entities.

Status:	Final
Date:	22 June 2020
Prepared By:	Janssen Research & Development, LLC

Table of Contents

1. List of Abbreviations

ASOs	administrative services only
C1	Comparator Cohort 1: Codeine
C2	Comparator Cohort 2: Codeine (exclude cough, antibiotic, cold and cough medications, and antihistamines in last 30 days)
CB	Covariate Balance
CCAE	IBM MarketScan® Commercial Database
CDM	Common Data Model
CI	confidence interval
CPRD	Clinical Practice Research Datalink
E	Equipoise
ER	Emergency Room
HR	hazard ratio
IP	Inpatient
IRB	Institutional Review Board
ITT	intent-to-treat
JMDC	Japan Medical Data Center
MDCD	IBM MarketScan® Multi-State Medicaid Database
MDCR	IBM MarketScan® Medicare Supplemental Database
MHRA	The Medicines and Healthcare products Regulatory Agency
NDC	National Drug Codes
NIHR	NHS National Institute for Health Research
O1	Outcome Cohort 1: Primary Hip Fracture (READ Codes for CPRD)
O2	Outcome Cohort 2: (Primary Hip Fracture ER/IP Dx with Hip Fracture procedures +/- 7 days) OR (Primary Hip Fracture procedures with Hip Fracture ER/IP +/- 7 days)
OHDSI	Observational Health Data Sciences and Informatics
OMOP	Observational Medical Outcomes Partnership
OPTUM_DOD	Optum© De-Identified Clinformatics® Data Mart Database – Date of Death
PPV	positive predictive value
PS	Propensity scores
RR	relative risk
SMD	standardized mean difference
T1	Target Cohort 1: Tramadol
T2	Target Cohort 2: Tramadol (exclude cough, antibiotic, cold and cough medications, and antihistamines in last 30 days)
TAR	time-at-risk
THIN	The Health Improvement Network
UK	United Kingdom
US	United States

2. Responsible Parties

2.1. Investigators and Authors

- Erica A Voss, MPH, Director, Observational Health Data Analytics[1,2,3]
- Rana Saberi Ali, MD, MPH, Director, Global Medical Safety, Janssen[1]
- Arun Singh , DO, MS, Director, Clinical Leader in Established Products[1]
- Gowtham Rao, MD, PhD, Senior Director, Observational Health Data Analytics[1,2]
- Peter R Rijnbeek, PhD, Associate Professor Health Data Science[2,3]
- Martijn J Schuemie, PhD, Senior Director, Observational Health Data Analytics[1,2]
- Daniel Fife, MD, PhD, Senior Director, Epidemiology[1]

Affiliations:

1. Janssen Research and Development, Titusville, NJ
2. OHDSI collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY
3. Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, Netherlands

2.2. Sponsor

Global Epidemiology, Janssen Research & Development, LLC.
1125 Trenton-Harbourton Rd, Titusville NJ 08560

3. Abstract

Hip fractures greatly impact an individual's quality of life and carry a high risk of death within 1 year. Tramadol is a commonly used weak opioid for treatment of pain. A recent study by Wei et al. found that risk for hip fractures was higher for new users of tramadol than for new users of codeine or NSAIDs. We were concerned of that study's design choices because of several limitations such as: A less-than-optimal propensity score adjustment strategy, the absence of negative controls, the failure to address possible differences in the initial doses of tramadol versus codeine, and the fact that the study was done in only one data source limited to one countries data. We propose to do a study to assess hip fracture incidence among users of tramadol versus codeine that will reassess the relationship and address the Wei et al. study limitations.

4. Amendments and Updates

None

5. Rationale and Background

Hip fractures are a major public health issue, particularly for older persons [1]. Hip fractures occur when a person breaks the bone between the pelvis and knee and these fractures are known as femoral-neck fractures or intertrochanteric or subtrochanteric fractures [2]. Hip fractures greatly impact an individual's quality of life with a high risk of death within 1 year [2]. Globally it is estimated that hip fractures affect 18% of women and 6% of men; globally hip fractures rank among the top 10 causes of disability [3, 4].

Tramadol is a commonly used weak opioid for the treatment of pain [5]. Tramadol is considered an analgesic alternative to strong opioids or the NSAIDS, since it is not expected to produce significant gastrointestinal bleeding or renal problems [4-6]. For these reasons and others, Tramadol is increasingly used worldwide for pain management [5].

Recently, Wei et al [5] reported in an observational study, the incidence of hip fracture among patients aged 50 to 90 years who were new users of tramadol compared to a propensity-score matched cohort of new users of codeine, and of several NSAIDs in The Health Improvement Network (THIN) between January 2000 and December 2016. The study found that the hazard ratio (HR) for hip fractures was higher for new users of tramadol compared to new users of codeine (the opioid comparator in the study), HR 1.28, (95% confidence interval [CI] 1.13 to 1.46). However, the study design contained limitations:

1. The propensity score is not precisely described. Superior methods such as large-scale propensity score fitting with LASSO regression were not used.
2. The study did not use negative controls or other methods to check for residual confounding.
3. The study did not document whether the extent of exposure to tramadol was similar to the extent of exposure to codeine either in terms of morphine equivalents per day or in terms of days' supply dispensed.
4. The study was done in a single data source so there's no assurance of the generalizability of the findings to other data sources – and if the findings may be attributable to the unique characteristics of the data source being studied.

We propose to do a study to assess hip fracture incidence among new users of tramadol versus codeine aged between 50 to 89 years that will address the above identified limitations of the original study by Wei et al.

6. Study Objective

Does exposure to tramadol have a different risk of experiencing hip fracture within 1 year, relative to codeine?

7. Research Methods

7.1. Study Design and Setting

This study will follow a retrospective, observational, comparative cohort design [8]. We define 'retrospective' to mean the study will be conducted using data already collected prior to the start of the study. We define 'observational' to mean there is no intervention or treatment assignment imposed by the study. We define 'cohort' to mean a set of subjects satisfying one or more inclusion criteria for a duration of time. We define 'comparative cohort design' to mean the formal comparison between two cohorts, a target cohort and comparator cohort, for the risk of an outcome during a defined time-period after cohort entry.

7.2. Data Sources

Six datasets were considered for performing this study. Clinical Practice Research Datalink (CPRD) was selected as this data source is similar to the one used in the Wei et al. paper. We additionally wanted to see if the study, if run across different data sources and diverse populations would yield similar results. For these other data sources, we considered five: IBM MarketScan® Commercial Database (CCAE), IBM MarketScan® Medicare Supplemental Database (MDCR), IBM MarketScan® Multi-State Medicaid Database (MDCD), Optum© De-Identified Clinformatics® Data Mart Database – Date of Death (OPTUM_DOD), and Japan Medical Data Center (JMDC). However, after reviewing the performance of our outcome definition in these datasets we only selected three datasets to move forward with: MDCR, MDCD, and OPTUM_DOD. The details of the performance evaluation are discussed further in Section 7.4.1. The four datasets selected are described in detail in Table 1.

All data sources have been standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [9], version 5.3. The OMOP CDM includes a standard representation of health care experiences (such as information related to drug utilization and condition occurrences), common vocabularies for coding clinical concepts, and enables consistent application of analysis across multiple disparate data [9]. The completed specification for the OMOP CDM is available at: https://github.com/OHDSI/CommonDataModel. Details about the model can be found at: https://ohdsi.github.io/CommonDataModel/. Documentation on the database transformations to the CDM can be found at: https://github.com/OHDSI/ETL-CDMBuilder/tree/master/man.

All analyses will be performed independently within each of these four data sources to produce a set of four results for each analysis. No subject-level data will be pooled across the data sources for any analysis, in part to preserve internal validity of the comparative analyses within each data source and avoid the potential risk of 'double-counting' cases for duplicate subjects.

TABLE 1 – Description of Data Sources	
IBM MarketScan® Medicare Supplemental Database (MDCR)	
Version ID	1104
Database Start Date	2000-01-01
Database End Date	2019-07-31
Database Description	IBM MarketScan® Medicare Supplemental Database (MDCR) represents health services of retirees in the United States with primary or Medicare supplemental coverage through privately insured fee-for-service, point-of-service, or capitated health plans. These data include adjudicated health insurance claims (e.g. inpatient, outpatient, and outpatient pharmacy). Additionally, it captures laboratory tests for a subset of the covered lives. The major data elements contained within this database are outpatient pharmacy dispensing claims (coded with National Drug Codes (NDC), inpatient and outpatient medical claims which provide procedure codes (coded in CPT-4, HCPCs, ICD-9-CM or ICD-10-PCS) and diagnosis codes (coded in ICD-9-CM or ICD-10-CM). The data also contain selected laboratory test results (those sent to a contracted thirds-party laboratory service provider) for a non-random sample of the population (coded with LOINC codes).
IBM MarketScan® Multi-State Medicaid Database (MDCD)	
Version ID	1105
Database Start Date	2006-01-01
Database End Date	2018-12-30
Database Description	IBM MarketScan® Multi-State Medicaid Database (MDCD) adjudicated US health insurance claims for Medicaid enrollees from multiple states and includes hospital discharge diagnoses, outpatient diagnoses and procedures, and outpatient pharmacy claims as well as ethnicity and Medicaid eligibility. Members maintain their same identifier even if they leave the system for a brief period however the dataset lacks lab data. The Medicaid dataset contains data from 10-12 states.
Optum© De-Identified Clinformatics® Data Mart Database – Date of Death (DOD) (OPTUM_DOD)	
Version ID	1107
Database Start Date	2000-05-01
Database End Date	2019-06-30
Database Description	Optum Clinformatics Extended DataMart is an adjudicated administrative health claims database for members with private health insurance, who are fully insured in commercial plans or in administrative services only (ASOs), Legacy Medicare Choice Lives (prior to January 2006), and Medicare Advantage (Medicare Advantage Prescription Drug coverage starting January 2006). The population is primarily representative of US commercial claims patients (0-65 years old) with some Medicare (65+ years old). Since few individuals are aged > 90 years those subjects are assigned a birthdate that would imply an age of 90 years. It includes data captured from administrative claims processed from inpatient and outpatient medical services and prescriptions as dispensed, as well as results for outpatient lab tests processed by large national lab vendors who participate in data exchange with Optum. This dataset also provides date of death (month and year only) for members with both medical

TABLE 1 – Description of Data Sources	
	and pharmacy coverage from the Social Security Death Master File (however after 2011 reporting frequency changed due to changes in reporting requirements) and location information for patients is at the US state level. Family identifiers are provided and utilized to infer mother-child linkages. Optum requests review of work prior to submitting for publication.
Clinical Practice Research Datalink (CPRD)	
Version ID	1102
Database Start Date	1988-01-01
Database End Date	2019-05-30
Database Description	The Clinical Practice Research Datalink (CPRD) is a governmental, not-for-profit research service, jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA), a part of the Department of Health, United Kingdom (UK). CPRD consists of data collected from UK primary care for all ages. This includes conditions, observations, measurements, and procedures that the general practitioner is made aware of in additional to any prescriptions as prescribed by the general practitioner. In addition to primary care, there are also linked secondary care records for a small number of people. The major data elements contained within this database are outpatient prescriptions given by the general practitioner (coded with Multilex codes) and outpatient clinical, referral, immunization or test events that the general practitioner knows about (coded in Read or ICD10 or LOINC codes). The database also contains the patients' year of births and any date of deaths. Use of this data set is subject to ISAC approval.

7.3. Study Populations

This section describes the target and comparator cohorts. Section 7.4 below describes the outcome cohorts. The codes used to define each cohort can be found in the Annex, “Concept Sets for Cohorts”. The cohorts were reviewed using CohortDiagnostics (https://github.com/OHDSI/CohortDiagnostics).

The target cohorts (tramadol - Section 7.3.1) and the comparator cohorts (codeine - Section 7.3.2) both contain two cohorts each described in further detail below. The first cohort both the target and comparator have is one that resembles what was described by Wei et al. We believe a priori that these cohorts may have confounding by indication due to codeine more often than tramadol being used to treat cough. We believe that our diagnostics will show this. The second cohort set are modified to make the target and comparator cohort more comparable. This is done by excluding subjects diagnosed with cough or cold in the 30 days prior to initial exposure to the target medications. Additionally, to make the second cohort set more comparable, we excluded subjects that were prescribed cold or cough medications, antibiotics, or antihistamines in the 30 days prior to initial exposure of the target medications.

7.3.1. Target Cohorts

We have two target cohorts that are exposed to tramadol. Target Cohort 1 is like what was described in the Wei et. al. paper (Table 2). Codeine is used to treat cough more often than is tramadol. To address this potential confounder, Target Cohort 2 has a modified cohort definition compared to Wei et al. with the intent to make the target and comparator cohorts more comparable (Table 3).

TABLE 2 - Target Cohort 1 (T1): Tramadol		
Inputs	**Concept Sets**	• Codeine • Tramadol • Hip Fracture Diagnosis (Fracture of neck of femur) • Hip Fracture Source Codes to Include • Hip Fracture Procedures (with revision codes) • Opioids • Opioid Abuse • Malignant Neoplasm Excluding Non-Melanoma Skin Cancer
	Initial Event Criteria	• Index is an exposure to tramadol • After (>) 1994.12.31 (1995 was the first full year tramadol was on the market in the UK) • Between 50-89 years of age at index • 365 days of observable time prior to index

	Initial Event Inclusion Criteria or Additional Qualifying Inclusion Criteria	• In the 365 days prior to or on the index (>=,<=)
		  o No evidence of opioid abuse
		  o No evidence of malignant neoplasm excluding non-melanoma skin cancer
		• In the 365 days prior to and not including the index (>=,<)
		  o No evidence of a hip fracture (diagnosis or procedure)
		  o No exposure to tramadol
		  o No exposure to opioids

	Exit Criteria	• End of continuous drug exposure (with 30-day persistence window, 0-day surveillance window)
		• Exposure to codeine
		• Death
		• End of continuous observation
		• The following will be additionally added as exit criteria outside of ATLAS:
		  o Reached age 89
		  o 365 days of follow-up

Notes	• The cohort design ensures that there are not subjects in both the target and comparator cohorts. • Unlike the Wei et al paper, we censored persons at age 89. The Wei et al paper limited their analysis when the subjects were age 90, however one of our data sets, OPTUM_DOD, censors ages at 90 meaning this age could represent 90+ year old. Therefore, we decided to censor at age 89.	

<table>
<tr><th colspan="3">TABLE 3 - Target Cohort 2 (T2): Tramadol (exclude cough, antibiotic, cold and cough medications, and antihistamines in last 30 days)</th></tr>
<tr><td rowspan="4">Inputs</td><td>Concept Sets</td><td>• Codeine
• Tramadol
• Hip Fracture Diagnosis (Fracture of neck of femur)
• Hip Fracture Source Codes to Include
• Hip Fracture Procedures (with revision codes)
• Opioids
• Opioid Abuse
• Malignant Neoplasm Excluding Non-Melanoma Skin Cancer
• Cough, Acute bronchospasm, Respiratory tract infection, Tracheobronchial disorder, Acute respiratory disease, Sinusitis
• Cough and Cold Preparations (excluding codeine)
• Antibacterials for Systemic Use
• Antihistamines for Systemic Use</td></tr>
<tr><td>Initial Event Criteria</td><td>• Index is an exposure to tramadol
• After (>) 1994.12.31 (1995 was the first full year tramadol was on the market in the UK)
• Between 50-89 years of age at index
• 365 days of observable time prior to index</td></tr>
<tr><td>Initial Event Inclusion Criteria or Additional Qualifying Inclusion Criteria</td><td>• In the 30 days prior to or on the index (>=,<=)
    o No evidence of cough, acute bronchospasm, respiratory tract infection, tracheobronchial disorder, acute respiratory disease, or sinusitis
    o No exposure to cough and cold preparations (excluding codeine)
    o No exposure to antibacterials for systemic use
    o No exposure to antihistamines for systemic use
• In the 365 days prior to or on the index (>=,<=)
    o No evidence of opioid abuse
    o No evidence of malignant neoplasm excluding non-melanoma skin cancer
• In the 365 days prior to and not including the index the index (>=,<)
    o No evidence of a hip fracture (diagnosis or procedure)
    o No exposure to tramadol
    o No exposure to opioids</td></tr>
<tr><td>Exit Criteria</td><td>• End of continuous drug exposure (with 30-day persistence window, 0-day surveillance window)
• Exposure to codeine
• Death
• End of continuous observation
• The following will be additionally added as exit criteria outside of ATLAS:
    o Reached age 89
    o 365 days of follow-up</td></tr>
<tr><td>Notes</td><td colspan="2">• The cohort design ensures that there are no subjects in both the target and comparator cohorts.
• Unlike the Wei et al paper, we censored persons at age 89. The Wei et al paper limited their analysis when the subjects were age 90, however one of our data sets, OPTUM_DOD, censors ages at 90 meaning this age could represent 90+ year old. Therefore, we decided to censor at age 89.
• When testing original target and comparator cohorts used within the Wei et al. paper we felt the addition exclusions made the cohorts more comparable.</td></tr>
</table>

7.3.2. Comparator Cohorts

We also have two comparator cohorts that are exposed to codeine. Comparator Cohort 1 is like what was described in the Wei et. al. paper (Table 4) and Comparator Cohort 2 has changes made to the Wei et al. cohort to make for a cohort more comparable to the target cohorts (Table 5).

TABLE 4 - Comparator Cohort 1 (C1): Codeine		
Inputs	**Concept Sets**	• Codeine • Tramadol • Hip Fracture Diagnosis (Fracture of neck of femur) • Hip Fracture Source Codes to Include • Hip Fracture Procedures (with revision codes) • Opioids • Opioid Abuse • Malignant Neoplasm Excluding Non-Melanoma Skin Cancer
	Initial Event Criteria	• Index is an exposure to codeine • After (>) 1994.12.31 (1995 was the first full year tramadol was on the market in the UK) • Between 50-89 years of age at index • 365 days of observable time prior to index

	Initial Event Inclusion Criteria or Additional Qualifying Inclusion Criteria	• In the 365 days prior to or on the index (>=,<=)
		    ○ No evidence of opioid abuse
		    ○ No evidence of malignant neoplasm excluding non-melanoma skin cancer
		• In the 365 days prior to and not including the index (>=,<)
		    ○ No evidence of a hip fracture (diagnosis or procedure)
		    ○ No exposure to codeine
		    ○ No exposure to opioids

	Exit Criteria	• End of continuous drug exposure (with 30-day persistence window, 0-day surveillance window)
		• Exposure to Tramadol
		• Death
		• End of continuous observation
		• The following will be additionally added as exit criteria outside of ATLAS:
		    ○ Reached age 89
		    ○ 365 days of follow-up

Notes	• The cohort design ensures that there are not subjects in both the target and comparator cohorts. • Unlike the Wei et al paper, we censored persons at age 89. The Wei et al paper limited their analysis when the subjects were age 90, however one of our data sets, OPTUM_DOD, censors ages at 90 meaning this age could represent 90+ year old. Therefore, we decided to censor at age 89.	

<table>
<tr><th colspan="3">TABLE 5 - Comparator Cohort 2 (C2): Codeine (exclude cough, antibiotic, cold and cough medications, and antihistamines in last 30 days)</th></tr>
<tr><td rowspan="4">Inputs</td><td>Concept Sets</td><td>Codeine
Tramadol
Hip Fracture Diagnosis (Fracture of neck of femur)
Hip Fracture Source Codes to Include
Hip Fracture Procedures (with revision codes)
Opioids
Opioid Abuse
Malignant Neoplasm Excluding Non-Melanoma Skin Cancer
Cough, Acute bronchospasm, Respiratory tract infection, Tracheobronchial disorder, Acute respiratory disease, Sinusitis
Cough and Cold Preparations (excluding codeine)
Antibacterials for Systemic Use
Antihistamines for Systemic Use</td></tr>
<tr><td>Initial Event Criteria</td><td>N/A</td></tr>
<tr><td>Initial Event Inclusion Criteria or Additional Qualifying Inclusion Criteria</td><td>Index is an exposure to codeine
After (>) 1994.12.31 (1995 was the first full year tramadol was on the market in the UK)
Between 50-89 years of age at index
365 days of observable time prior to index
In the 30 days prior to or on the index (>=,<=)
No evidence of cough, acute bronchospasm, respiratory tract infection, tracheobronchial disorder, acute respiratory disease, or sinusitis
No exposure to cough and cold preparations (excluding codeine)
No exposure to antibacterials for systemic use
No exposure to antihistamines for systemic use
In the 365 days prior to or on the index (>=,<=)
No evidence of opioid abuse
No evidence of malignant neoplasm excluding non-melanoma skin cancer
In the 365 days prior to and not including the index (>=,<)
No evidence of a hip fracture (diagnosis or procedure)
No exposure to codeine
No exposure to opioids</td></tr>
<tr><td>Exit Criteria</td><td>End of continuous drug exposure (with 30-day persistence window, 0-day surveillance window)
Exposure to Tramadol
Death
End of continuous observation
The following will be additionally added as exit criteria outside of ATLAS:
Reached age 89
365 days of follow-up</td></tr>
<tr><td>Notes</td><td colspan="2">The cohort design ensures that there are no subjects in both the target and comparator cohorts.
Unlike the Wei et al paper, we censored persons at age 89. The Wei et al paper limited their analysis when the subjects were age 90, however one of our data sets, OPTUM_DOD, censors ages at 90 meaning this age could represent 90+ year old. Therefore, we decided to censor at age 89.
When testing original target and comparator cohorts used within the Wei et al. paper we felt the addition exclusions made the cohorts more comparable.</td></tr>
</table>

7.4. Outcomes of Interest

7.4.1. Outcome Cohorts

Our outcome definition is hip fracture. However, because of the differences in the type of information available from the CPRD and the US claims databases, we have two algorithms; one for use in CPRD and one for use within the claims data (MDCR, MDCD, OPTUM_DOD). The first definition for use in CPRD, Outcome Cohort 1 described in Table 6, is a replication of what was done in the Wei et al paper (the specific codes used were found in the Berry et al. paper [10]). The second definition was developed using algorithms found in published literature [1, 11, 12] and more details are discussed about that definition in Table 7.

TABLE 6 - Outcome Cohort 1 (O1): Primary Hip Fracture (READ Codes for CPRD)		
Inputs	**Concept Sets**	• Hip Fracture (Read Codes)
	Initial Event Criteria	• Index is the first occurrence in a person's history either of: ○ Diagnosis of hip fracture ○ Procedure associated to hip fracture • Between 50-89 years of age at index • 365 days of observable time prior to index
	Initial Event Inclusion Criteria or Additional Qualifying Inclusion Criteria	• N/A
	Exit Criteria	• End of continuous observation
Notes	• Based on the Berry et al. publication. (Berry, S. D., et al. (2013). "Diuretic initiation and the acute risk of hip fracture." Osteoporosis International 24(2): 689-695.) • Unlike the Wei et al paper, we censored persons at age 89. The Wei et al paper limited their analysis when the subjects were age 90, however one of our data sets, OPTUM_DOD, censors ages at 90 meaning this age could represent 90+ year old. Therefore, we decided to censor at age 89.	

TABLE 7 - Outcome Cohort 2 (O2): (Primary Hip Fracture ER/IP Dx with Hip Fracture procedures +/- 7 days) OR (Primary Hip Fracture procedures with Hip Fracture ER/IP +/- 7 days)		
Inputs	**Concept Sets**	• Hip Fracture Diagnosis (Fracture of neck of femur) • Hip Fracture Procedures (with revision codes) • Hip Fracture Procedures (without revision codes) • Hip Fracture Source Codes to Exclude • Hip Fracture Source Codes to Include

	Initial Event Criteria	• Index is the first occurrence in a person's history either of:
		  ○ Emergency Room (ER)/Inpatient (IP) Visit diagnosis of hip fracture with a hip fracture procedure within +/- 7 days
		  ○ Hip fracture procedure with a hip fracture diagnosis within +/- 7 days
		• Between 50-89 years of age at index
		• 365 days of observable time prior to index

	Initial Event Inclusion Criteria or Additional Qualifying Inclusion Criteria	• Exclude prior hip fracture diagnosis • Exclude certain index source codes • Exclude prior hip fracture procedures
	Exit Criteria	End of continuous observation.

Notes	• This definition came from careful review of the following papers:
	  ○ Ray WA, Griffin MR, Fought RL, Adams ML. Identification of fractures from computerized Medicare files. J Clin Epidemiol. 1992 Jul;45(7):703-14. PubMed PMID: 1619449.
	  ○ Nair SS, Lenihan CR, Montez-Rath ME, Lowenberg DW, Chertow GM, Winkelmayer WC. Temporal trends in the incidence, treatment and outcomes of hip fracture after first kidney transplantation in the United States. Am J Transplant. 2014 Apr;14(4):943-951. doi: 10.1111/ajt.12652. Epub 2014 Feb 20. PubMed PMID: 24712332; PubMed Central PMCID: PMC4117735.
	  ○ Hudson M, Avina-Zubieta A, Lacaille D, Bernatsky S, Lix L, Jean S. The validity of administrative data to identify hip fractures is high--a systematic review. J Clin Epidemiol. 2013 Mar;66(3):278-85. doi: 10.1016/j.jclinepi.2012.10.004. Review. PubMed PMID: 23347851.
	• Hudson et al. is a systematic review of identification of hip fractures in administrative data. Table 2 provides details on the papers reviewed and discussed a sensitivity that ranged from 65 to 97 and a positive predictive value that ranged from 34 to 98. Ray et al. was one of the papers reviewed by Hudson that had a high sensitivity and high positive predictive value. Nair leveraged Ray's definition, however 22 years later. Using both the information from Ray's evaluated definition and Nair's most recently implementation of the algorithm we created our phenotype. This algorithm was reviewed using a tool called PheValuator and had a performance in line with what we found during this systematic review.
	• Unlike the Wei et al paper, we censored everything at age 89. The Wei et al paper limited their analysis to patients aged 50 to 90, however one of our data sets, OPTUM_DOD, censors ages at 90 meaning this age could represent 90+ year old. Therefore, we decided to censor at age 89.	

As discussed in Section 7.2, the choice of claims databases was limited by the performance of our Outcome Cohort 2 in those databases. Performance was measured by a tool call PheValuator which can estimate phenotype algorithm performance within a database [13]. Table 8 shows the performance of Outcome Cohort 2 in all five databases considered. MDCR, MDCD, and OPTUM_DOD are in line with the sensitivity, positive predictive value (PPV), and specificity reported in the systematic review of hip fracture definitions in claims databases [11]. The performance of the outcome in CCAE and JMDC was poor and thus those databases were excluded from the analysis. It was hypothesized that since our outcome definition is looking at subjects 50 years or older, CCAE may not perform as well as this population is primarily subjects aged 65 and less. Additionally, for JMDC, it was hypothesized that a different outcome definition would be required to perform well due to different coding practices and it was decided to not move forward with this database.

TABLE 8 – Performance of Outcome Cohort 2 - (Primary Hip Fracture ER/IP Dx with Hip Fracture procedures +/- 7 days) OR (Primary Hip Fracture procedures with Hip Fracture ER/IP +/- 7 days) Across the Claims Databases Considered for this analysis.			
DB	**SENSITIVITY**	**PPV**	**SPECIFICITY**
CCAE	0.466 (0.450 - 0.482)	0.582 (0.564 - 0.599)	0.999 (0.999 - 0.999)
MDCR	0.835 (0.832 - 0.838)	0.716 (0.712 - 0.719)	0.992 (0.992 - 0.992)
MDCD	0.668 (0.663 - 0.674)	0.761 (0.755 - 0.766)	0.996 (0.996 - 0.996)
OPTUM_DOD	0.624 (0.619 - 0.629)	0.862 (0.857 - 0.866)	0.998 (0.998 - 0.998)
JMDC	0.212 (0.179 - 0.248)	0.312 (0.266 - 0.362)	0.999 (0.999 - 0.999)
The extremely specific cohort (xSpec) used by PheValuator was 5 occurrence of a hip fracture diagnosis.			

7.4.2. Negative Control Cohorts

Negative controls are cohorts defined by a diagnosis or exposure known to not be associated with the target or comparator cohorts, such that we can assume the true relative risk between the two cohorts is 1. Such cohorts can be exposure cohorts or outcome cohorts. For further details, including the process used to select negative controls, see reference [14]. The 101 negative controls for the present study are found in the 'Negative Control List' in the Annex, each represent a diagnosis that is used to define the negative control cohort. For this analysis we will use outcome cohorts; like the outcome of hip fracture, we will look for first occurrence of each negative control concept and their descendant concepts.

The same analysis that will be performed for each pairwise comparison to assess the risk of hip fracture (see Sections 9.3.1. and 9.3.2) will also be performed to assess the risk of each negative control outcome. Because the negative control qualifying criteria support the a priori assertion of no effect, we assume the true relative risk (RR) for each negative control outcome is 1, and the difference between RR=1 and the observed effect estimate will be considered error, encompassing both random and potentially systematic. We will be able to calibrate the hazard ratio and confidence intervals on the basis of the empirical null distribution which consists of the estimates for the negative control outcomes [15, 16].

7.5. Exposures of Interest

Our exposures of interest are new users of tramadol and codeine. See Sections 7.3.1 and 7.3.2 for discussion on the target and comparator cohorts.

7.6. Other Variables of Interest (Demographic Characteristics, Effect Modifiers)

Propensity scores (PS) will be used as an analytic strategy to reduce potential confounding due to imbalance between the target and comparator cohorts in baseline covariates [17]. The PS is the probability of a subject being classified in the target cohort versus the comparator cohort, given a set of observed covariates (see Section 9.3.1).

The types of baseline covariates used to fit the propensity score model will be:

- Demographics
    - Gender
    - Age group (5-year bands)
    - Index year
    - Index month
- Time Bound Era Covariates
    - Condition group concepts both 365 days and 30 days on or prior to cohort index
    - Ingredients both 365 days and 30 days on or prior to cohort index
    - Drug groups both 365 days and 30 days on or prior to cohort index
- Time Bound Covariates
    - Procedure occurrence concept and any of its descendants both 365 days and 30 days on or prior to cohort index
    - The occurrence of a measurement concept and any of its descendants both 365 days and 30 days on or prior to cohort index
    - The occurrence of an observation concept and any of its descendants both 365 days and 30 days on or prior to cohort index
    - Device concept and any of its descendants both 365 days and 30 days on or prior to cohort index
    - Number of visits observed both 365 days and 30 days on or prior to cohort index (visits are spans of time a person continuously receives medical services from one or more providers typically classified into outpatient care, inpatient confinement, emergency room, and long-term care)
    - Number of visits by type (i.e. emergency room, inpatient, outpatient) observed both 365 days and 30 days on or prior to cohort index
- Index Score Covariates
    - CHA2DS2-VASc - using conditions all time on or prior to cohort index
    - Charlson Index - Romano adaptation, using conditions all time on or prior to cohort index

Specific drug exposure concepts that define the target and comparator cohorts will be excluded from the propensity score model fitting. This large-scale empirical adjustment strategy should address expected confounders, including demographics, outcome risk factors, comorbidities associated with mortality, and health service utilization behavior. The study will be subject to the limitation that some confounders may be unmeasured or inadequately represented in observational data, including weight, smoking status, and lifestyle behaviors.

7.7. Tools

This study will be designed using OHDSI tools [18] (specifically the Population-Level Estimation tools) and run with R [19].

8. Sample Size and Study Power

The sample size of the cohorts is reported in Table 9. These patient counts represent the initial population, prior to statistical adjustment, so provide an upper bound of exposure available for each analysis. For population-level effect estimation, where our aim is to produce an unbiased estimate of the average treatment effect, the precision we will achieve will vary by the incidence rate of each outcome. Because our focus is to estimate the magnitude of the effect, it is acceptable to be underpowered for the analyses, recognizing that this will manifest as wider confidence intervals that account for the random sampling error inherent to the analysis. Smaller sample size for specific comparisons may be associated with larger statistical uncertainty. Small samples may also limit the ability to fit adequate propensity models and thus limit our ability to control confounding. Note that we will not pool the raw data across the 4 databases for analysis.

There is no a priori hypothesis testing for this study, therefore there is no prespecified requirement of sample sizes for the comparative analyses. After all design specifications have been implemented for each pairwise comparison, the minimum detectable hazard ratio will be calculated. The calculation includes a targeted type I error rate (alpha) of 0.05 (2-sided) and a type II error rate (beta) of 0.20 (power=80%) and reports the minimum hazard ratio detectable given the final target and comparator patient count, outcome event count, and TAR [20].

TABLE 9 – Number of Subjects in Target and Comparator Cohorts Before Matching

Cohort	CPRD	MDCR	MDCD	OPTUM_DOD
Target Cohort 1: Tramadol	166,884	381,096	111,716	948,214
Target Cohort 2: Tramadol (exclude cough, antibiotic, cold and cough medications, and antihistamines in last 30 days)	139,015	283,036	72,113	684,406
Comparator Cohort 1: Codeine	1,116,400	551,519	69,065	1,215,785
Comparator Cohort 2: Codeine (exclude cough, antibiotic, cold and cough medications, and antihistamines in last 30 days) *	894,883	182,930	22,720	356,804
Outcome Cohort 1: Primary Hip Fracture (READ Codes for CPRD)	57,584	-	-	-
Outcome Cohort 2: (Primary Hip Fracture ER/IP Dx with Hip Fracture procedures +/- 7 days) OR (Primary Hip Fracture procedures with Hip Fracture ER/IP +/- 7 days)	10,286	138,466	33,759	148,889

CPRD = Clinical Practice Research Datalink, MDCR = IBM MarketScan® Medicare Supplemental Database, MDCD = IBM MarketScan® Multi-State Medicaid Database, DOD = Optum© De-Identified Clinformatics® Data Mart Database – Date of Death

* On the claims data sources (MDCR, MDCD, OPTUM_DOD) we see the implementation of the exclusion of cough, antibiotic, cold and cough medications, and antihistamines in the last 30 days on the codeine cohort has a different effect on the proportion removed as compared to the UK data source (CPRD). The exclusion only reduces the CPRD cohort by 20% while in the claims database the exclusions reduces the cohort by 70%.

9. Data Analysis Plan

The analysis has been specified in ATLAS [21], detail description of the are found in this section. The code for this study can be found: https://github.com/ohdsi-studies/TramadolVsCodeineForHipFracture

9.1. Calculation of Time-at-Risk

Two time-at-risk (TAR) definitions will be used for follow-up of outcome of interest in this study:

9.1.1. Primary TAR - TAR 1 – On-Treatment

The on-treatment TAR starts on index and will append 7 days to the last exposure date. The end of drug exposure defined with 30-day persistence window and 0-day surveillance window across drug exposures.

Persistence window is a period of tolerance that is allowed when constructing periods of persistence exposure. For example, a 30-day persistence window would allow for a gap between two prescriptions not exceeding 30 days over the number of days supplied or prescribed. Surveillance window represents the number of days added to the end of the persistence exposure to a drug as an addition period of surveillance prior to the cohort exit. Example, if you have a drug exposure that ends on January 1 and your surveillance window is 30 days, if another drug exposure for that same drug occurs during the surveillance window you will consider the exposure to continue without stop.

9.1.2. Sensitivity TAR - TAR 2 – Intent-to-Treat

The intent-to-treat (ITT) TAR starts on index until target or comparator ends observable time within the data.

9.2. Patient Characteristics Summary

9.2.1. Descriptive Characterizations

A descriptive characterization of subjects included in each exposure cohort. Continuous variables will be summarized using mean (± standard deviation) and median. Counts and proportions will be used to summarize categorical variables. Clinical characterization results will be reported in covariate balance tables for the target and comparator cohort in each pairwise comparison. Covariate balance between the comparison cohorts will be summarized by showing the proportions and mean values for all baseline covariates with the associated standardized mean difference computed for each covariate. Attrition tables will report the loss of subjects from the original target and comparator cohorts to the subpopulations that remain after all design considerations have been applied.

9.2.2. Description of Initial Dose

The original study did not document whether the extent of the exposure to tramadol was similar to the extent of exposure to codeine as estimated in morphine equivalents. We will characterize the initial dose to understand if there are differences. If there are differences we will address them within our analysis.

9.2.3. Incidence Analysis

The unadjusted incidence of hip fracture will be calculated for all exposure cohorts during two time-at-risk periods (see Section 9.1) to establish the base rate of event occurrence which will provide context to the subsequent population-level effect estimates. The number of persons, number of events during the time-at-risk period, the incidence proportion per 1,000 persons, and the incidence rate per 1,000 person-years will be computed for both outcomes, during both time-at-risk periods. The incidence analyses involve direct observation of the experience of subjects, which can provide context about the real-world patterns of event occurrence in different populations but cannot be used for causal inference, to draw comparative conclusions about the effects of any treatment or extrapolate to the general population.

9.3. Model Specification

In this study, we compare the target cohort with the comparator cohort for the hazards of outcome during the time-at-risk by applying a Cox proportional hazards model. Estimates of risk will be generated as the empirically calibrated hazard ratios (HR), 95% confidence intervals (CI), and p-values. The uncalibrated HR, CI, and p-value will also be reported. The number of persons, days amount of time-at-risk, and number of outcome events in each cohort in each pairwise comparison after PS adjustment will also be reported.

The time-to-event of outcome among subjects in the target and comparator cohorts is determined by calculating the number of days from the start of the time-at-risk window (index date), until the earliest event among 1) the first occurrence of the outcome or 2) the end of the time-at-risk window as defined by the cohort.

9.3.1. Propensity Score Model Specification

Propensity scores (PS) will be used as an analytic strategy to reduce potential confounding due to imbalance of subject characteristics at baseline between the target and comparator cohort in a pairwise comparison. The PS is the probability of a subject being classified in the target cohort versus the comparator cohort, given a set of observed covariates. The PS will be estimated for each subject using the predicted probability from a regularized logistic regression model, fit with a Laplace prior (LASSO) and the regularization hyperparameter selected by optimizing the likelihood in a 10-fold cross validation, using a starting variance of 0.01 and a tolerance of 2e-7. Covariates that occur in fewer than 0.1% of the combined target and comparator cohort in a pairwise comparison will be excluded prior to model fitting. Subjects will be matched on 1:1 ratio matching of target to comparator subjects. This approach will use a greedy matching algorithm by applying a caliper of 0.2 of the standard deviation on the logit scale of the PS distribution.

9.3.2. Outcome Model Specification

A Cox proportional hazards regression model will be used to model the time to the first outcome occurrence for the target group relative to the comparator group while accounting for the PS matching. Estimates of risk will be generated as the empirically calibrated hazard ratios (HR), 95% confidence intervals (CI), and p-values (see Section 7.4.2).The uncalibrated HR, CI, and p-value will also be reported. The number of persons, days amount of time-at-risk, and number of outcome events in each cohort in each pairwise comparison after PS adjustment will also be reported.

For each target-comparator-outcome-analysis combination, heterogeneity of the hazards ratios will be estimated, using I^2 as a metric [22]. If there is sufficient homogeneity across sources (I^2<40%) [23], database-specific estimates will be pooled through random effect meta-analysis using the Hartung-Knapp-Sidik-Jonkman inverse-variance method [24]. Pooled results will include p-values corrected for multiple testing using Hochberg's step-up procedure. Where observed heterogeneity across sources is greater than I^2≥40%, pooled estimates will not be generated.

9.4. Evidence Evaluation

For each population-level effect estimate generated by the study, i.e. each target-comparator-outcome-analysis-database combination, we will report diagnostics to assess its potential for bias and threats to its valid interpretation. The diagnostics include both propensity score distribution and covariate balance before and after propensity score matching.

9.4.1. Propensity Score Distribution

Once the PS model is fit for each pairwise comparison, the PS distribution for the target and comparator cohort will be plotted to evaluate the comparability, as a proxy for exchange ability, of the two cohorts before matching. The plot will be scaled to the preference score, which normalizes for initial cohort size imbalance. If the proportion of subjects in clinical equipoise, i.e. the patients with a preference score between 0.3 and 0.718, is less than 50%, then the estimate will not be reported.

9.4.2. Covariate balance before and after propensity score matching

Covariate balance will be evaluated by plotting the standardized mean difference (SMD) of each covariate before against the SMD after propensity score matching. After matching SMDs with values of <0.1 are asserted to indicate negligible group differences [19].

9.5. Analyses to Perform

Total number of population-level effect estimates that will be generated by this study can be found in Table 10. Note that the total number of estimates generated does not necessarily mean the number of estimates that will be reported. The number of estimates reported depends on diagnostic assessment (see Section 9.4).

TABLE 10 – Analysis to Perform		
Analysis	**Analysis Specifications**	**Number of Analysis**
Wei Replication	***Analysis 101*** Tramadol (T1) vs Codeine (C1) for Hip Fracture (O1) [2 TAR specifications * 1 databases (CPRD)]	2
	Analysis 102 Tramadol (T1) vs Codeine (C1) for Hip Fracture (O2) [2 TAR specifications * 3 databases (MDCR, MDCD, DOD)]	6
Wei Replication (assuming some model exclusions)	***Analysis 201*** Tramadol (T1) vs Codeine (C1) for Hip Fracture (O1) [2 TAR specifications * 1 databases (CPRD)]	2
	Analysis 202 Tramadol (T1) vs Codeine (C1) for Hip Fracture (O2) [2 TAR specifications * 3 databases (MDCR, MDCD, DOD)]	6
Replication using Best Practices	***Analysis 301*** Tramadol (T2) vs Codeine (C2) for Hip Fracture (O1) [2 TAR specifications * 1 databases (CPRD)]	2
	Analysis 302 Tramadol (T2) vs Codeine (C2) for Hip Fracture (O2) [2 TAR specifications * 3 databases (MDCR, MDCD, DOD)]	6
	Total Analysis	***24***

Small sample sizes of some exposure cohorts and subgroup exposure cohorts may limit the ability to generate population-level effect estimates for which valid inferences can be made. For example, small exposure cohort sample sizes may limit the ability of the PS adjustment strategy to achieve acceptable covariate balance in a pairwise comparison or in conjunction with outcome event occurrence may be underpowered to detect an estimate of a meaningful magnitude. Rather than deciding a priori to not make certain comparisons on this basis, this study will generate a full set of population-level effect estimation diagnostics, including empirical calibration, for all pre-specified pairwise comparisons; the estimates for target-comparator-outcome-analysis-databases combinations that acceptably pass all study diagnostics will be reported. Consistent application of pre-specified methods in high throughput observational studies may reduce results reproducibility problems observed when study design decisions are made on a study- or comparison-specific basis [25].

9.6. Output

Characteristics of the subjects will be provided as discussed in Section 9.2.

Covariate balance before and after matching will be summarized in tabular form by showing the mean value for all baseline covariates in the target and comparator cohort, with the associated standardized mean difference computed for each covariate.

Once the propensity score model is fit, we will plot the propensity score distribution of the target and comparator cohorts to evaluate the comparability of the two cohorts. The plot will be scaled to the preference score, normalizing for any imbalance in cohort size. The covariates selected within the propensity score model, with associated coefficients will also be reported.

A plot showing the propensity score distributions for both cohorts after stratification will be provided, with each quantile cut point shown as a vertical line. Covariate balance will be evaluated by plotting the standardized mean difference of each covariate before propensity score stratification against the standardized mean difference for each covariate after propensity score stratification.

An attrition diagram will be provided to detail the loss of subjects from the original target cohort and comparator cohort to the subpopulations that remain after all design considerations have been applied.

The final outcome model, a conditional Cox proportional hazards model, will be summarized by providing the hazards ratio and associated 95% confidence interval. The number of persons, amount of time-at-risk, and number of outcomes in each cohort will also be reported.

10. Evidence Evaluation Results

All evidence evaluation diagnostic results were generated and reviewed. Table 11 is a summary of the diagnostics produced.

TABLE 11 – Diagnostic Results			
Analysis	**Analysis Specifications**	**Number of Analysis**	**Evidence Diagnostics**
Wei Replication	***Analysis 101*** Tramadol (T1) vs Codeine (C1) for Hip Fracture (O1) [2 TAR specifications * 1 databases (CPRD)]	2	Not Applicable*
	Analysis 102 Tramadol (T1) vs Codeine (C1) for Hip Fracture (O2) [2 TAR specifications * 3 databases (MDCR, MDCD, DOD)]	6	Analysis 102 **
Wei Replication (assuming some model exclusions)	***Analysis 201*** Tramadol (T1) vs Codeine (C1) for Hip Fracture (O1) [2 TAR specifications * 1 databases (CPRD)]	2	Analysis 201
	Analysis 202 Tramadol (T1) vs Codeine (C1) for Hip Fracture (O2) [2 TAR specifications * 3 databases (MDCR, MDCD, DOD)]	6	Analysis 202
Replication using Best Practices	***Analysis 301*** Tramadol (T2) vs Codeine (C2) for Hip Fracture (O1) [2 TAR specifications * 1 databases (CPRD)]	2	Analysis 301
	Analysis 302 Tramadol (T2) vs Codeine (C2) for Hip Fracture (O2) [2 TAR specifications * 3 databases (MDCR, MDCD, DOD)]	6	Analysis 302

*These diagnostics were not produced as “acetaminophen” was so predictive of the comparator cohort the propensity score was perfectly predictive between the two groups.

**Only results for MDCR and OPTUM_DOD produced as results for MDCD were not produced as “acetaminophen” was so predictive of the comparator cohort the propensity score was perfectly predictive between the two groups.

TAR = Time at Risk,
CPRD = Clinical Practice Research Datalink, MDCR = IBM MarketScan® Medicare Supplemental Database, MDCD = IBM MarketScan® Multi-State Medicaid Database, DOD = Optum© De-Identified Clinformatics® Data Mart Database – Date of Death

We pre-specified that we will only report results that met two criteria: >= 50% of subjects in clinical equipoise and covariate balance as achieved (after matching SMDs with values of <0.1 are asserted to indicate negligible group differences). This was described in Section 10. Table 12 reviews all 24 comparison to see which results will be reported. Of the 24 analysis, we will report population-level effect estimates for 10. For example, in Analysis 201 both the OT and ITT analysis pass the diagnostics for equipoise and adequate covariate balance while in Analysis 202, for both the OT and ITT analysis in the claims database the analysis failed to meet equipoise. For the analysis that passed diagnostics for CPRD, no results for CPRD will be reported either internally or externally until ISAC approval is gained, 4 of the 10 results would be generated from CPRD.

TABLE 12 – Cohort Comparisons by database that achieve (pass) or do not achieve (fail) adequate equipoise (>50%) and adequate covariate balance after 1:1 propensity score matching. Population-level effect estimates from database-specific exposure cohort comparisons that achieve adequate equipoise and adequate covariate balance will be reported.

Analysis	Target	Com-parator	Out-come	TAR	CPRD		MDCR		MDCD		DOD		Number of Passing Analysis
					E	*CB*	*E*	*CB*	*E*	*CB*	*E*	*CB*	
101	T1	C1	O1	OT	X	X							0
				ITT	X	X							0
102	T1	C1	O2	OT			FAIL	PASS	X	X	FAIL	PASS	0
				ITT			FAIL	PASS	X	X	FAIL	PASS	0
201	T1	C1	O1	OT	Pass	Pass							1
				ITT	Pass	Pass							1
202	T1	C1	O2	OT			FAIL	PASS	FAIL	PASS	FAIL	PASS	0
				ITT			FAIL	PASS	FAIL	PASS	FAIL	PASS	0
301	T2	C2	O1	OT	Pass	Pass							1
				ITT	Pass	Pass							1
302	T2	C2	O2	OT			PASS	PASS	PASS	PASS	PASS	PASS	3
				ITT			PASS	PASS	PASS	PASS	PASS	PASS	3

E = Equipoise, CB = Covariate Balance, TAR = Time at Risk, OT = On Treatment Analysis, ITT = Intent to Treat,
T1 = Target Cohort 1: Tramadol, C1 = Comparator Cohort 1: Codeine ,
T2 = Target Cohort 2: Tramadol (exclude cough, antibiotic, cold and cough medications, and antihistamines in last 30 days), C2 = Comparator Cohort 2: Codeine (exclude cough, antibiotic, cold and cough medications, and antihistamines in last 30 days)
O1 = Outcome Cohort 1: Primary Hip Fracture (READ Codes for CPRD),
O2 = Outcome Cohort 2: (Primary Hip Fracture ER/IP Dx with Hip Fracture procedures +/- 7 days) OR (Primary Hip Fracture procedures with Hip Fracture ER/IP +/- 7 days)
CPRD = Clinical Practice Research Datalink, MDCR = IBM MarketScan® Medicare Supplemental Database, MDCD = IBM MarketScan® Multi-State Medicaid Database, DOD = Optum© De-Identified Clinformatics® Data Mart Database – Date of Death
X = No preference score overlap thus no matches found and no diagnostics produced

11. Strengths and Limitations

11.1. Strengths

- Cohort studies allow direct estimation of incidence rates following exposure of interest, and the new-user design can capture early events following treatment exposures while avoiding confounding from previous treatment effects. New use allows for a clear exposure index date.
- Propensity score matching allows balancing on many baseline potential confounders.
- Use of negative and positive control outcomes allows for evaluating the study design in terms of residual bias.

11.2. Limitations

Even with the improvements to the study design as proposed by Wei et al, there are still limitations with this work:

- Even though many potential confounders will be included in this study, there may be residual bias due to unmeasured or mis-specified confounders.
- Additionally, since we are not looking for patients with a specific indication there is a potential residual confounding by indication in the analysis.
- Indication and outcome misclassification are a concern in administrative data sources because sometimes diagnosis codes intended for reimbursement is not considered the gold-standard documentation of patient present clinical condition.
- Causality between drug exposure and any given event cannot be drawn for individual cases.

12. Protection of Human Subjects

The New England Institutional Review Board (IRB) has determined that studies conducted in IBM MarketScan CCAE, MDCR, and Optum Extended DOD are exempt from study-specific IRB review, as these studies do not qualify as human subjects research. The JMDC, the owner of the JMDC database, has certified that the data "is anonymously processed information so ethics approval is not necessary when you use it for the publications."

Confidentiality of subject records will be maintained always. All study reports will contain aggregate data only and will not identify individual subjects or physicians. At no time during the study will the sponsor receive subject identifying information except when it is required by regulations in case of reporting adverse events.

13. Safety Data Collection and Reporting

This study uses coded data that already exist in an electronic database. In this type of data source, the minimum criteria for reporting an adverse event (i.e., identifiable subject, identifiable reporter, a suspect product, and event) are not available, and adverse events are not reportable as individual case safety reports. The study results will be assessed for medically important results.

14. Plans for Disseminating and Communicating Study Results

No results for the analysis within CPRD will be shared either internally or externally without prior ISAC approval. We will not move forward with publication with the Optum data without prior notification to Optum.

The protocol will be registered at European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (EnCePP) after finalization. Results will be reported to the registration location within 12 months of completion. Additionally, results will be submitted for peer-reviewed publication.

15. List of Tables and Figures

16. Annex

16.1. Appendix 1 – Concept Sets for Cohorts

APPENDIX 1 – Table 1 – Concept Set Lookup	
Cohort	**Concept Sets**
Target Cohort 1 (T1): Tramadol	T1.zip
Target Cohort 2 (T2): Tramadol (exclude cough, antibiotic, cold and cough medications, and antihistamines in last 30 days)	T2.zip
Comparator Cohort 1 (C1): Codeine	C1.zip
Comparator Cohort 2 (C2): Codeine (exclude cough, antibiotic, cold and cough medications, and antihistamines in last 30 days)	C2.zip
Outcome Cohort 1 (O1): Primary Hip Fracture (READ Codes for CPRD)	O1.zip
Outcome Cohort 2 (O2): (Primary Hip Fracture ER/IP Dx with Hip Fracture procedures +/- 7 days) OR (Primary Hip Fracture procedures with Hip Fracture ER/IP +/- 7 days)	O2.zip

16.2. Appendix 2 – Negative Controls List

The following is a list of outcomes not believed to be caused by tramadol or codeine.

APPENDIX 2 – Table 1 – Negative Control List	
CONCEPT_ID	**CONCEPT_NAME**
46273370	Abnormal chest sounds
4093531	Absence of toe
4092879	Absent kidney
441968	Acetonuria
4081648	Acute vaginitis
4061157	Antenatal ultrasound scan abnormal
200528	Ascites
438409	Attention deficit hyperactivity disorder
434170	Atypical squamous cells of undetermined significance on cervical Papanicolaou smear
4045471	Autoimmune reaction mediated by cell-mediated immunity
132736	Bacteremia
438990	Benign neuroendocrine tumor
4195873	Breath smells unpleasant
4067069	Callosity
4228429	Carnitine deficiency
376116	Central scotoma
443570	Cervicovaginal cytology: Low grade squamous intraepithelial lesion
201613	Chronic nonalcoholic liver disease
4201390	Colostomy present
377888	Conductive hearing loss
4022071	Convalescence
380724	Corneal ghost vessels
42537730	Coronary artery graft present
436233	Delayed milestone
438759	Descemet's membrane fold
377910	Deviated nasal septum
438701	Disseminated malignancy of unknown primary
381877	Dysfunction of eustachian tube
192367	Dysplasia of cervix
433111	Effects of hunger
435170	Effects of thirst
4028689	Electrocerebral silence
200775	Endometrial hyperplasia
433527	Endometriosis
4170770	Epidermoid cyst
4167696	Estrogen receptor positive tumor

APPENDIX 2 – Table 1 – Negative Control List	
CONCEPT_ID	**CONCEPT_NAME**
374358	Excess skin of eyelid
4086512	Excess subcutaneous fat
4229403	Flat anterior chamber of eye
4166231	Genetic predisposition
438111	Hematologic neoplasm of uncertain behavior
439871	Hemospermia
372897	Homonymous hemianopia
435511	Hypercalcemia
133729	Hyperparathyroidism
4287416	Hyperphenylalaninemia
440129	Hypertrophy of nasal turbinates
4057743	Hyperuricuria
4029280	Hypervitaminosis B6
435522	Hypervitaminosis D
434004	Hypervolemia
440072	Hypogammaglobulinemia
435515	Hypo-osmolality and or hyponatremia
432596	Immune defect
374375	Impacted cerumen
4280828	Infectious disease carrier
440053	Infestation by insect
4168222	Intra-abdominal and pelvic swelling, mass and lump
440710	Intraretinal microvascular abnormality
196168	Irregular periods
4228331	Leukokeratosis nicotina palati
4027782	Lipid storage disease
435516	Lipoprotein deficiency disorder
4166126	Localized swelling, mass and lump, trunk
44784454	Localized visual field defect
433997	Lymphangioma
40482859	Malignant carcinoid tumor
440058	Malignant lymphoma of extranodal AND/OR solid organ site
436426	Malleus mobility reduced
72737	Microcalcifications of the breast
45757412	Mitochondrial metabolism defect
4298207	Mouth breathing
437543	Multiple cranial nerve palsy
375077	Neglect of one side of body
320073	Neutropenia

APPENDIX 2 – Table 1 – Negative Control List	
CONCEPT_ID	**CONCEPT_NAME**
40480893	Nonspecific tuberculin test reaction
444428	Nonvenomous insect bite without infection
439035	Otosclerosis
4153516	Patient immunocompromised
4141640	Perimenopausal disorder
22856	Polyglandular dysfunction
437369	Postmature infancy
36675035	Prematurity of infant
435028	Puerperal pyrexia of unknown origin
4308074	Pyogenic granuloma
372614	Retained magnetic foreign body in multiple sites
436828	Saliva abnormal
435088	Senility
4090205	Sequelae of tuberculosis
29056	Sialoadenitis
141825	Simple goiter
443082	Starvation
40636815	Supernumerary teeth
4182164	Temporomandibular joint crepitus
40485495	Thymoma
433244	Tooth loss
4201387	Tracheostomy present
4029731	Trimethylaminuria
196821	Urethral discharge
195603	Vulval and/or perineal noninflammatory disorders
440193	Wristdrop

17. References

1. Ray, W.A., et al., *Identification of fractures from computerized Medicare files.* J Clin Epidemiol, 1992. **45**(7): p. 703-14.
2. Bhandari, M. and M. Swiontkowski, *Management of Acute Hip Fracture.* N Engl J Med, 2017. **377**(21): p. 2053-2062.
3. Cooper, C., G. Campion, and L.J. Melton, 3rd, *Hip fractures in the elderly: a world-wide projection.* Osteoporos Int, 1992. **2**(6): p. 285-9.
4. Veronese, N. and S. Maggi, *Epidemiology and social costs of hip fracture.* Injury, 2018. **49**(8): p. 1458-1460.
5. Wei, J., et al., *Association of Tramadol Use With Risk of Hip Fracture.* J Bone Miner Res, 2020.
6. Gibson, T.P., *Pharmacokinetics, efficacy, and safety of analgesia with a focus on tramadol HCl.* Am J Med, 1996. **101**(1A): p. 47S-53S.
7. Miotto, K., et al., *Trends in Tramadol: Pharmacology, Metabolism, and Misuse.* Anesth Analg, 2017. **124**(1): p. 44-51.
8. Ryan, P.B., et al., *Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system.* Drug Saf, 2013. **36 Suppl 1**: p. S59-72.
9. Voss, E.A., et al., *Feasibility and utility of applications of the common data model to multiple, disparate observational health databases.* J Am Med Inform Assoc, 2015. **22**(3): p. 553-64.
10. Berry, S.D., et al., *Diuretic initiation and the acute risk of hip fracture.* Osteoporosis International, 2013. **24**(2): p. 689-695.
11. Hudson, M., et al., *The validity of administrative data to identify hip fractures is high--a systematic review.* J Clin Epidemiol, 2013. **66**(3): p. 278-85.
12. Nair, S.S., et al., *Temporal trends in the incidence, treatment and outcomes of hip fracture after first kidney transplantation in the United States.* Am J Transplant, 2014. **14**(4): p. 943-951.
13. Swerdel, J.N., G. Hripcsak, and P.B. Ryan, *PheValuator: Development and evaluation of a phenotype algorithm evaluator.* J Biomed Inform, 2019. **97**: p. 103258.
14. Voss, E.A., et al., *Accuracy of an automated knowledge base for identifying drug adverse reactions.* J Biomed Inform, 2017. **66**: p. 72-81.
15. Schuemie, M.J., et al., *Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data.* Proc Natl Acad Sci U S A, 2018. **115**(11): p. 2571-2577.
16. Schuemie, M.J., et al., *Interpreting observational studies: why empirical calibration is needed to correct p-values.* Stat Med, 2014. **33**(2): p. 209-18.
17. Tian, Y., M.J. Schuemie, and M.A. Suchard, *Evaluating large-scale propensity score performance through real-world and synthetic data experiments.* Int J Epidemiol, 2018. **47**(6): p. 2005-2014.
18. *The Book of OHDSI*. 2019-09-26; Available from: https://ohdsi.github.io/TheBookOfOhdsi/.
19. *R: A language and environment for statistical computing*. 2019, R Foundation for Statistical Computing: Vienna, Austria.
20. Schoenfeld, D.A., *Sample-size formula for the proportional-hazards regression model.* Biometrics, 1983. **39**(2): p. 499-503.
21. *ATLAS*. [web page] [cited 2020; Available from: https://github.com/OHDSI/Atlas/wiki.
22. Higgins, J.P., et al., *Measuring inconsistency in meta-analyses.* BMJ, 2003. **327**(7414): p. 557-60.
23. *Cochrane Handbook for Systematic Reviews of Interventions: 9.5.2 Identifying and measuring heterogeneity*. 2011, Cochrane Reviews.

24. IntHout, J., J.P. Ioannidis, and G.F. Borm, *The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method.* BMC Med Res Methodol, 2014. **14**: p. 25.
25. Schuemie, M.J., et al., *Improving reproducibility by using high-throughput observational studies with empirical calibration.* Philos Trans A Math Phys Eng Sci, 2018. **376**(2128).