Evaluation of methods for synthesizing interaction effect estimates

Martijn Schuemie, Yong Chen

2023-08-22

Table of Contents

# 1 Introduction

There are several challenges when estimating causal effects across a distributed (federated) research network:

1. Sharing of individual patient data (IPD) is not allowed.
2. The nature of the data gathering process leads to censored observation periods that require a time-to-event analysis rather than simpler incidence rate estimation.
3. Due to the observational nature of the data and the consequent potential for confounding, some correction for baseline differences between exposure groups always proves necessary, typically via stratifying, matching, or weighting by a propensity score or disease risk score.
4. Since different data sites represent different patient populations, inter-site heterogeneity often arises.
5. Even though many of these health care databases contain the records of large numbers of patients, co-occurrences of even moderately rare exposures and/or outcomes often prove to be sparse to non-existent.

In our prior work (Schuemie et al. 2022) we proposed to solve this problem by sharing likelihood profiles between sites; Instead of sharing the point estimates and standard errors (which imply the likelihood follows a normal distribution), sites would share the shape of the likelihood curve. Our result shows that, when counts are low (as often happens as mentioned under 5), our approach provides unbiased estimates. In contrast, using a normal approximation can lead to substantial bias.

Here we aim to extend our work to situations where there is more than one parameter of interest. Specifically, we are interested in the synthesis of evidence on interactions between two exposures. Our prior work applied to single parameters of interest (e.g. the main effect), but often we want to know both the main effect and some interaction effect (e.g. is the causal effect larger in some subgroup of interest?).

The research described here aims to completed extensive simulation studies by testing several approaches to likelihood profiling for effect interactions in a real worl setting.

# 2 Example interaction effect study

To evaluate the use of likelihood profiles when synthesizing evidence on interaction effects, we will use the following example: non-selective nonsteroidal anti-inflammatory drugs (NSAIDs) are known to increase the risk of gastrointestinal (GI) bleeding, and this risk is known to increase when the patient is also on an oral anticoagulant such as warfarin. (Choi et al. 2010)

We therefore formulate an example comparative cohort study as follows:

* **Target**: New users of diclofenac (a non-selective NSAID)
* **Comparator**: New users of celecoxib (a selective COX-2 inhibitor NSAID)
* **Outcome**: GI bleed
* **Interaction term**: Warfarin exposure at the time of NSAID initiation
* **Time at risk**: On-treatment: starting on the day of treatment initiation, ending when treatment is stopped (for at least 30 days).
* **Model**: Either Cox or Poisson regression

Cohort definitions of the target, comparator, and outcome are provided in Appendix A. The outcome cohort definition was taken from the OHDSI Phenotype Library.

Warfarin exposure is defined as a drug era overlapping with the target or comparator index date, having ingredient concept ID 1310149 (Warfarin).

We will require at least 365 days of continuous observation prior to the index date.

Large-scale propensity scores (PS) will be fitted using the standard set of features, including all drugs, conditions, procedures, measurements, and observations in the year prior to index date, as well as demographics. We will use the PS to stratify the population in 10 equally-sized strata.

The cohort and comparator cohorts will be restricted to at most 100,000 subjects. If these cohorts are larger, a 100,000 random sample will be taken.

A set of 34 negative control outcomes, outcomes believed to be caused by neither the target nor the comparator, has been defined. (See Appendix B). We assume that these controls are negative both for the main effect and the interaction effect.

# 3 Evaluation of evidence synthesis methods

The example study will be executed across the databases described in the ‘Data sources’ section. Various approaches for likelihood profiling will be applied to produce summary estimates for the main effect and the interaction effect, both for the outcome of interest (GI bleed) as well as the negative control outcomes.

The summary estimates for GI bleed will be compared with the gold standard, produced by pooling the data (stratified by data source).

The negative control summary estimates will be used to estimate residual systematic error.

# 4 Data sources

The DatabaseDiagnostics package was used to select those databases that appear to have the elements needed for the example estimation questions:

## 4.1 IBM Health MarketScan® Commercial Claims and Encounters Database

The IBM(R) MarketScan(R) Commercial Database (CCAE) includes health insurance claims across the continuum of care (e.g. inpatient, outpatient, outpatient pharmacy, carve-out behavioral healthcare) as well as enrollment data from large employers and health plans across the United States who provide private healthcare coverage for more than 155 million employees, their spouses, and dependents. This administrative claims database includes a variety of fee- for-service, preferred provider organizations, and capitated health plans.

## 4.2 Optum’s Clinformatics® Extended Data Mart – Date of Death (DOD)

Optum‘s Clinformatics(R) Data Mart is derived from a database of administrative health claims for members of large commercial and Medicare Advantage health plans.The database includes approximately 17-19 million annual covered lives, for a total of over 65 million unique lives over a 12 year period (1/2007 through 12/2019). Clinformatics(R) Data Mart is statistically de-identified under the Expert Determination method consistent with HIPAA and managed according to Optum(R) customer data use agreements. CDM administrative claims submitted for payment by providers and pharmacies are verified, adjudicated and de-identified prior to inclusion. This data, including patient-level enrollment information, is derived from claims submitted for all medical and pharmacy health care services with information related to healthcare costs and resource utilization. The population is geographically diverse, spanning all 50 states. Optum Clinformatics(R) Data Mart Data of Death (Optum DOD) also provides date of death (month and year only) for members with both medical 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.

## 4.3 Optum EHR

Optum‘s longitudinal EHR repository is derived from dozens of healthcare provider organizations in the United States, that include more than 700 Hospitals and 7000 Clinics; treating more than 102 million patients receiving care in the United States. The data is certified as de-identified by an independent statistical expert following HIPAA statistical de-identification rules, and managed according to Optum(R) customer data use agreements. Clinical, claims and other medical administrative data is obtained from both Inpatient and Ambulatory electronic health records (EHRs), practice management systems and numerous other internal systems. Information is processed, normalized, and standardized across the continuum of care from both acute inpatient stays and outpatient visits. Optum(R) data elements include: demographics, medications prescribed and administered, immunizations, allergies, lab results (including microbiology), vital signs and other observable measurements, clinical and inpatient stay administrative data and coded diagnoses and procedures. In addition, Optum(R) uses natural language processing (NLP) computing technology to transform critical facts from physician notes into usable datasets. The NLP data provides detailed information regarding signs and symptoms, family history, disease related scores (i.e. RAPID3 for RA, or CHADS2 for stroke risk), genetic testing, medication changes, and physician rationale behind prescribing decisions that might never be recorded in the EHR.

## 4.4 PharMetrics Plus

Data is from 2006 - 2022 and comprises of fully adjudicated medical and pharmacy claims. It contains a longitudinal view of inpatient and outpatient services, prescription and office/outpatient administered drugs, costs and enrollment information. With IQVIA Adjudicated Health Plan Claims, an enrolled patient can be tracked across all sites of care: hospital, specialist, emergency room, pharmacy, primary care, and more.<U+FFFD>

## 4.5 IBM Health MarketScan® Medicare Supplemental and Coordination of Benefits Database

The IBM(R) MarketScan(R) Medicare Supplemental Database (MDCR) represents the health services of approximately 10 million retirees in the United States with Medicare supplemental coverage through employer-sponsored plans. This database contains primarily fee-for-service plans and includes health insurance claims across the continuum of care (e.g. inpatient, outpatient and outpatient pharmacy).

## 4.6 IBM Health MarketScan® Multi-State Medicaid Database

The IBM(R) MarketScan(R) Multi-State Medicaid Database (MDCD) reflects the healthcare service use of individuals covered by Medicaid programs in numerous geographically dispersed states. The database contains the pooled healthcare experience of Medicaid enrollees, covered under fee-for-service and managed care plans. It includes records of inpatient services, inpatient admissions, outpatient services, and prescription drug claims, as well as information on long-term care. Data on eligibility and service and provider type are also included. In addition to standard demographic variables such as age and gender, the database includes variables such as federal aid category (income based, disability, Temporary Assistance for Needy Families) and race.

## 4.7 Japan Medical Data Center (JMDC)

JMDC database is a payer based database that has collected claims, ledger of the insured people and health checkup results from more than 250 payers. It covers workers and their dependents aged under 74. It is longitudinal and the largest one as commercially available database in Japan with more than 13 million enrollments. All medical history of the insured people are available and patient reported outcome research can be done through payers on-demand basis. Those aged 66 or older are less representative as compared with whole population in the nation. When estimated among the people who are younger than 66 years old, the proportion of children younger than 18 years old in JMDC is approximately the same as the proportion in the whole nation. Claims data are derived from monthly claims issued by clinics, hospitals and community pharmacies. The number of claims issued and added to JMDC database is about 6,000,000 per month. The size of JMDC population is about 6% of the whole nation.

# 5 References

Choi, K. H., A. J. Kim, I. J. Son, K. H. Kim, K. B. Kim, H. Ahn, and E. B. Lee. 2010. “Risk factors of drug interaction between warfarin and nonsteroidal anti-inflammatory drugs in practical setting.” *J Korean Med Sci* 25 (3): 337–41.

Schuemie, M. J., Y. Chen, D. Madigan, and M. A. Suchard. 2022. “Combining cox regressions across a heterogeneous distributed research network facing small and zero counts.” *Stat Methods Med Res* 31 (3): 438–50.

# 6 Appendix A: Cohort Definitions

## 6.1 Target

### 6.1.1 Cohort Entry Events

People with continuous observation of 365 days before event may enter the cohort when observing any of the following:

1. drug exposure of ‘Celecoxib’ for the first time in the person’s history.

Limit cohort entry events to the earliest event per person.

### 6.1.2 Inclusion Criteria

#### 6.1.2.1 1. Prior osteoarthritis of knee

Entry events having at least 1 condition occurrence of ‘Osteoarthritis of knee’, starting between 365 days before and 0 days after cohort entry start date.

### 6.1.3 Cohort Exit

The cohort end date will be based on a continuous exposure to ‘Celecoxib’: allowing 30 days between exposures, adding 0 days after exposure ends, and using days supply and exposure end date for exposure duration.

### 6.1.4 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

### 6.1.5 Celecoxib

| Concept ID | Concept Name | Code | Vocabulary | Excluded | Descendants | Mapped |
| --- | --- | --- | --- | --- | --- | --- |
| 1118084 | celecoxib | 140587 | RxNorm | NO | YES | NO |

### 6.1.6 Osteoarthritis of knee

| Concept ID | Concept Name | Code | Vocabulary | Excluded | Descendants | Mapped |
| --- | --- | --- | --- | --- | --- | --- |
| 4079750 | Osteoarthritis of knee | 239873007 | SNOMED | NO | YES | NO |

## 6.2 Comparator

### 6.2.1 Cohort Entry Events

People with continuous observation of 365 days before event may enter the cohort when observing any of the following:

1. drug exposure of ‘Diclofenac’ for the first time in the person’s history.

Limit cohort entry events to the earliest event per person.

### 6.2.2 Inclusion Criteria

#### 6.2.2.1 1. Prior osteoarthritis of knee

Entry events having at least 1 condition occurrence of ‘Osteoarthritis of knee’, starting between 365 days before and 0 days after cohort entry start date.

### 6.2.3 Cohort Exit

The cohort end date will be based on a continuous exposure to ‘Diclofenac’: allowing 30 days between exposures, adding 0 days after exposure ends, and using days supply and exposure end date for exposure duration.

### 6.2.4 Cohort Eras

Entry events will be combined into cohort eras if they are within 0 days of each other.

### 6.2.5 Diclofenac

| Concept ID | Concept Name | Code | Vocabulary | Excluded | Descendants | Mapped |
| --- | --- | --- | --- | --- | --- | --- |
| 1124300 | diclofenac | 3355 | RxNorm | NO | YES | NO |

### 6.2.6 Osteoarthritis of knee

| Concept ID | Concept Name | Code | Vocabulary | Excluded | Descendants | Mapped |
| --- | --- | --- | --- | --- | --- | --- |
| 4079750 | Osteoarthritis of knee | 239873007 | SNOMED | NO | YES | NO |

## 6.3 Outcome

### 6.3.1 Cohort Entry Events

People may enter the cohort when observing any of the following:

1. condition occurrences of ‘Gastrointestinal hemorrhage GI bleeding’.

Restrict entry events to having at least 1 visit occurrence of ‘Inpatient or ER visit’, starting anytime on or before cohort entry start date and ending between 0 days before and all days after cohort entry start date.

### 6.3.2 Cohort Exit

The cohort end date will be offset from index event’s start date plus 7 days.

### 6.3.3 Cohort Eras

Entry events will be combined into cohort eras if they are within 30 days of each other.

### 6.3.4 Inpatient or ER visit

| Concept ID | Concept Name | Code | Vocabulary | Excluded | Descendants | Mapped |
| --- | --- | --- | --- | --- | --- | --- |
| 262 | Emergency Room and Inpatient Visit | ERIP | Visit | NO | YES | NO |
| 9203 | Emergency Room Visit | ER | Visit | NO | YES | NO |
| 9201 | Inpatient Visit | IP | Visit | NO | YES | NO |

### 6.3.5 Gastrointestinal hemorrhage GI bleeding

| Concept ID | Concept Name | Code | Vocabulary | Excluded | Descendants | Mapped |
| --- | --- | --- | --- | --- | --- | --- |
| 4138962 | Acute duodenal ulcer without hemorrhage AND without perforation | 32490005 | SNOMED | YES | YES | NO |
| 4195231 | Acute gastric ulcer without hemorrhage AND without perforation | 67964002 | SNOMED | YES | YES | NO |
| 4147683 | Acute gastrojejunal ulcer without hemorrhage AND without perforation | 30514008 | SNOMED | NO | NO | NO |
| 4163865 | Acute peptic ulcer without hemorrhage AND without perforation | 45485004 | SNOMED | YES | YES | NO |
| 195584 | Acute peptic ulcer without hemorrhage AND without perforation but with obstruction | 58085004 | SNOMED | YES | YES | NO |
| 40482690 | Angiodysplasia of duodenum | 442267002 | SNOMED | NO | YES | NO |
| 28779 | Bleeding esophageal varices | 17709002 | SNOMED | NO | YES | NO |
| 4222896 | Chronic duodenal ulcer without hemorrhage AND without perforation | 40214005 | SNOMED | YES | YES | NO |
| 4296611 | Chronic gastric ulcer without hemorrhage AND without perforation | 76796008 | SNOMED | YES | YES | NO |
| 200769 | Chronic gastric ulcer without hemorrhage, without perforation AND without obstruction | 1567007 | SNOMED | YES | YES | NO |
| 4177387 | Chronic gastrojejunal ulcer without hemorrhage AND without perforation | 4269005 | SNOMED | YES | YES | NO |
| 434400 | Chronic gastrojejunal ulcer without hemorrhage AND without perforation but with obstruction | 56579005 | SNOMED | YES | YES | NO |
| 438795 | Chronic gastrojejunal ulcer without hemorrhage, without perforation AND without obstruction | 41626001 | SNOMED | YES | YES | NO |
| 4204555 | Chronic peptic ulcer without hemorrhage AND without perforation | 5492000 | SNOMED | YES | YES | NO |
| 24973 | Chronic peptic ulcer without hemorrhage AND without perforation but with obstruction | 12384004 | SNOMED | YES | YES | NO |
| 23808 | Chronic peptic ulcer without hemorrhage, without perforation AND without obstruction | 60400003 | SNOMED | YES | YES | NO |
| 2002608 | Control of hemorrhage and suture of ulcer of stomach or duodenum | 44.4 | ICD9Proc | NO | YES | NO |
| 198798 | Dieulafoy’s vascular malformation | 109558001 | SNOMED | NO | YES | NO |
| 4198381 | Ulcer of duodenum | 51868009 | SNOMED | NO | YES | NO |
| 4209746 | Duodenal ulcer without hemorrhage AND without perforation | 56776001 | SNOMED | YES | YES | NO |
| 4112183 | Esophageal varices with bleeding, associated with another disorder | 195475003 | SNOMED | NO | YES | NO |
| 2108900 | Esophagogastroduodenoscopy, flexible, transoral; with control of bleeding, any method | 43255 | CPT4 | NO | YES | NO |
| 2108878 | Esophagoscopy, flexible, transoral; with control of bleeding, any method | 43227 | CPT4 | NO | YES | NO |
| 4265600 | Gastric ulcer | 397825006 | SNOMED | NO | YES | NO |
| 4248429 | Gastric ulcer without hemorrhage AND without perforation | 73481001 | SNOMED | YES | YES | NO |
| 192671 | Gastrointestinal hemorrhage | 74474003 | SNOMED | NO | YES | NO |
| 4101104 | Gastrojejunal ulcer without hemorrhage AND without perforation | 2783007 | SNOMED | YES | YES | NO |
| 443530 | Hematochezia | 405729008 | SNOMED | YES | YES | NO |
| 197925 | Hemorrhage of rectum and anus | 266464001 | SNOMED | YES | YES | NO |
| 4027663 | Peptic ulcer | 13200003 | SNOMED | NO | YES | NO |
| 4291028 | Peptic ulcer without hemorrhage AND without perforation | 37442009 | SNOMED | YES | YES | NO |

# 7 Appendix B: Negative controls

| outcomeId | outcomeName |
| --- | --- |
| 199074 | Acute pancreatitis |
| 433753 | Alcohol abuse |
| 257007 | Allergic rhinitis |
| 443800 | Amenorrhea |
| 436665 | Bipolar disorder |
| 4084966 | Candida infection of genital region |
| 29735 | Candidiasis of mouth |
| 314658 | Cardiomegaly |
| 380094 | Carpal tunnel syndrome |
| 255573 | Chronic obstructive lung disease |
| 443617 | Conduct disorder |
| 192367 | Dysplasia of cervix |
| 134718 | Hirsutism |
| 433811 | Hydronephrosis |
| 438134 | Hypersomnia |
| 24609 | Hypoglycemia |
| 437784 | Infectious mononucleosis |
| 139099 | Ingrowing nail |
| 80004 | Injury of hand |
| 440358 | Lipoma (clinical) |
| 316084 | Lymphadenitis |
| 436940 | Metabolic syndrome X |
| 319843 | Mitral valve disorder |
| 440374 | Obsessive-compulsive disorder |
| 193739 | Ovarian failure |
| 375292 | Perforation of tympanic membrane |
| 321596 | Peripheral venous insufficiency |
| 436676 | Posttraumatic stress disorder |
| 194997 | Prostatitis |
| 73754 | Restless legs |
| 141932 | Senile hyperkeratosis |
| 374366 | Sensorineural hearing loss |
| 313459 | Sleep apnea |
| 197236 | Uterine leiomyoma |