# How Often: Large Scale Incidence Rate Calculation of Health Outcomes for Drugs Nested by Indication

**Hsin Yi Chen, BS[1], Christopher Knoll, BS[2], Elise Ruan, MD MPH[1], Adam Black, BA[3], Sarah Seager, BA[4], Patrick Ryan, PhD[2], George Hripcsak, MD MS[1]**

**[1]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA**

**[2]Janssen Research & Development, Titusville, NJ, USA**

**[3]Odysseus Data Services, Cambridge, MA, USA**

**[4]IQVIA, Real World Solutions, Cambridge, MA, USA**

## Background

Although drug product labels list potential adverse reactions observed in clinical trials and post-marketing surveillance, little is known about the real-world incidence of these side effects. While it is not causal, incidence rates can still support clinical decision making: if an event is extremely rare, then even if a causal relationship does exist, it may not be as clinically relevant than other more prevalent events.

Prior work with OHDSI has demonstrated the clinical value of real-world incidence rates through large-scale characterization of drug adverse events[1] and contextualizing COVID-19 vaccine-associated adverse effects based on background incidence rates[2,3]. Prior OHDSI has also shown that incidence estimates are sensitive to range of factors, including age, sex, calendar time, and indexing events[4]. However, the impact of drug indication on incidence rates remains unexplored. Given that some drugs can be indicated for multiple different target diseases, it is possible that the incidence of different health outcomes could differ by the indication, but the extent of this variation is unknown.

In this study, we used real-world data to calculate the incidence rates of various health outcomes of interest across 13 drug classes, stratified by indication. We then examined the amount of heterogeneity that can be attributed to drug indication, age, sex, and database.

## Methods

The study is an observational cohort study based on routinely collected health care data mapped to the OMOP CDM. The analysis was performed on 13 databases, including sources with different data source provenance (administrative claims data, electronic health record data), origin (the US, Belgium, Italy, Australia, and France), and representing different populations (privately insured employed patients in IBM MarketScan Commercial Claims (CCAE) vs patients with limited income in IBM MarketScan Multi-state Medicaid (MDCD)).

The analysis was conducted in October 2023 on the 13 databases. Exposure cohorts were based on the first occurrence of a drug exposure in a patient's longitudinal record using RxNorm terms for marketed drug ingredients to define instances of drug exposure. We excluded patients with less than one year of observation prior to the index date to better identify these patients who have previously been diagnosed with the outcome of interest, defined in the OHDSI Phenotype Library[5].

We selected 12 classes of drugs: 1) beta blockers, 2) cephalosporins, 3) calcium channel blockers, 4) DPP-4 inhibitors, 5) fluoroquinolones, 6) GLP-1 antagonists, 7) IL-23 inhibitors, 8) JAK inhibitors, 9) SGLT2 inhibitors, 10) thiazide diuretics, 11) trimethoprim, and 12) TNF-alpha inhibitors. For each drug, we

additionally stratified our analysis based on their indications. We looked at a total of 73 different health outcomes of interest, such as 3 and 4-point major adverse cardiovascular event (MACE) outcomes, abdominal pain, acute kidney injury, edema, and stroke. For each drug indication and outcome pair, we calculated the incidence rate of the outcome, defined as the ratio of the number of cases to the total time at risk for all patients in the target cohort. We defined our time-at-risk as starting one day after the index day and ending 365 days after the index date (regardless of cohort end date), and additionally stratified each target cohort based on age (separated into deciles) and sex.

We conducted a random effect meta-analysis of incidence rates across databases. Additionally, for drug classes with more than one indication, we conducted a variance components analysis to quantify the relative heterogeneity from indication, age, gender, and database.

All analyses were performed using code developed for the OHDSI Methods library. The code for this study can be found at https://ohdsi.github.io/CohortIncidence/ and https://github.com/ohdsi-studies/HowOften. Meta analysis was conducted using the R metafor package.

**Results**

We calculated a total of 77,631 incidence rates, which are available publicly at https://results.ohdsi.org/.

Eight drug classes had more than one identified indication: beta blockers, cephalosporin, fluoroquinolone, GLP-1 antagonists, JAK-inhibitors, SGLT-2 inhibitors, trimethoprim, and TNF-alpha inhibitors. All these drug classes had two identified indications, except for beta blockers, which had three indications, and TNF-alpha inhibitors, which had five identified indications.

As an example of the analysis we conducted, for beta blockers, the five outcomes of interest which had the largest variance components (indicating the outcomes where there was the largest amount of heterogeneity that can be attributed to the drug indication) were thrombosis with cytopenia, polymorphic ventricular tachycardia, progressive multifocal leukoencephalopathy, drug rash with eosinophilia and systemic symptoms, and coronary vessel revascularization. In general, incidence rates for these outcomes were lower for those who were prescribed a beta blocker for hypertension as opposed to acute myocardial infarction or left heart failure. Of note, acute myocardial infarction was technically among the outcomes with large variance components, but because acute myocardial infarction is an indication for beta blocker, we omitted that outcome for this drug class. These differences can be visualized in Figure 1.

We conducted a variance components analysis in order to quantify the relative heterogeneity that can be attributed to indication, compared to age, biological sex, and database (Table 1). Of the 8 drug classes in our study with more than one indication, trimethoprim had the highest heterogeneity attributed to indication, measured by the median variance component for indication across 73 outcomes (0.49), and GLP-1 antagonists had the lowest heterogeneity attributed to indication (<0.001). Within drug classes, when we compare the variance components for indication, age, biological sex, and database, the variance components for age and database tends to be the largest, followed by indication, and gender almost always had the smallest variance component.

**Conclusion**

Meta-analyzed incidence rates for the drug classes considered in this study were sensitive to stratifications by indications, and among the eight drug classes included in our analysis, trimethoprim was most sensitive to stratification by indication, and GLP-1 least sensitive. As we move on to expanding

this project to explore all exposures and all outcomes of interest, this study has shown, for some health outcomes, it may be important to nest exposures within different indications.

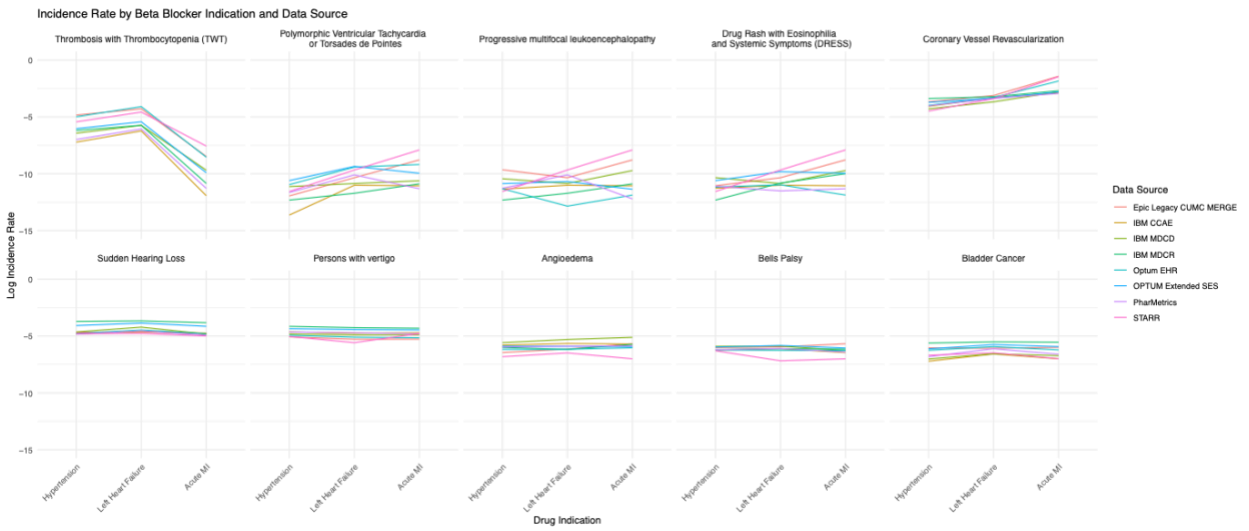


*Figure 1. Incidence rates (log scale) across outcomes with the 5 highest (top row) and 5 lowest (bottom row) indication variance components after adjusting for age, gender, and database. Variance components estimates the contribution of multiple factors (in this case, drug indication, age, gender, and database) to the variability of incidence rate estimates. In other words, the five outcomes with the highest indication variance components are the outcomes where indications matter "most" after adjusting for database, age, and gender. Note: we only plot the databases where there is data available for all indications.*

| ***Drug class*** | ***Indications*** | ***Median VC*** |
|---|---|---|
| Beta Blockers | 1) Essential Hypertension, 2) Left Heart Failure, 3) Acute Myocardial Infarction | 0.1013 |
| SGLT2 Inhibitors | 1) Type 2 Diabetes Mellitus, 2) Left Heart Failure | 0.2642 |
| GLP-1 antagonists | 1) Type 2 Diabetes Mellitus, 2) Obesity | <0.001 |
| Cephalosporins | 1) Urinary Tract Infection, 2) Acute Typical Pneumonia | 0.0397 |
| Fluoroquinolones | 1) Urinary Tract Infection, 2) Acute Typical Pneumonia | 0.0983 |
| Trimethoprim | 1) Urinary tract infection, 2) Pneumonia | 0.4887 |
| JAK inhibitors | 1) Rheumatoid Arthritis, 2) Ulcerative Colitis | 0.0383 * |
| TNF-alpha inhibitors | 1) Plaque Psoriasis, 2) Rheumatoid Arthritis, 3) Ulcerative Colitis, 4) Psoriatic Arthritis, 5) Crohn's Disease | 0.0332 |

*Table 1. 8 drug classes in our study had more than 1 indication. Of these 8 drug classes, trimethoprim had the highest heterogeneity measured by median VC (variance component) between indications among outcomes. Variance components are adjusted for age, gender, and indication*

** For JAK inhibitors, we did not adjust for age and gender, as we had <1000 patients at risk in these strata*

| *Drug class* | *Indications* | *Database* | *Age* | *Biological Sex* |
|---|---|---|---|---|
| Beta Blockers | 0.1013 | 0.1537 | **0.3102** | 0.0204 |
| SGLT2 Inhibitors | 0.2642 | **0.3170** | 0.2779 | 0.0155 |
| GLP-1 antagonists | <0.001 | **0.6117** | 0.3678 | 0.0289 |
| Cephalosporins | 0.0397 | 0.7230 | **1.4631** | 0.0515 |
| Fluoroquinolones | 0.0983 | **0.994** | 0.8573 | 0.0696 |
| Trimethoprim | 0.4887 | 0.2772 | **1.5228** | 0.1219 |
| JAK inhibitors* | 0.0383 | 0.1792 | **0.2055** | 0.0937 |
| TNF-alpha inhibitors | 0.0332 | 0.1675 | **0.1815** | 0.0221 |

*Table 2. Within each class, we broke down the variance components across indications, database, age, and biological sex. Bolded numbers are the largest variance components in each class. Age and database were typically among the factors with the highest variance components, followed by indications, then biological sex.*

** for JAK inhibitors, the 'indications' column was adjusted for database, the 'database' column was adjusted for indications, and the 'age' and 'biological sex' columns were adjusted for indications. This is because as we had <1000 patients at risk in many of the age and gender stratas, leading to few observed incidence rates.*

## References


1. Ruan E, Natarajan K, Chen R, Han J, Velez M, Abdul-Basser T, et al. HowOften: Large Scale Incidence Rate Calculation of Every Side Effect for Every Drug.

2. Li X, Ostropolets A, Makadia R, Shoaibi A, Rao G, Sena AG, et al. Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study. BMJ. 2021 Jun 14;n1435.

3. Voss EA, Shoaibi A, Yin Hui Lai L, Blacketer C, Alshammari T, Makadia R, et al. Contextualising adverse events of special interest to characterise the baseline incidence rates in 24 million patients with COVID-19 across 26 databases: a multinational retrospective cohort study. eClinicalMedicine. 2023 Apr;58:101932.

4. Ostropolets A, Li X, Makadia R, Rao G, Rijnbeek PR, Duarte-Salles T, et al. Factors Influencing Background Incidence Rate Calculation: Systematic Empirical Evaluation Across an International Network of Observational Databases. Front Pharmacol. 2022 Apr 26;13:814198.

5. Rao G (2024). *PhenotypeLibrary: The OHDSI Phenotype Library*. https://ohdsi.github.io/PhenotypeLibrary/, https://github.com/OHDSI/PhenotypeLibrary.