HowOften: Large Scale Incidence Rate Calculation of Every Side Effect for Every Drug

Elise Ruan¹, Karthik Natarajan¹, Ruijun Chen², Jungmi Han¹, Mark Velez³, Taha Abdul-Basser⁴, Edwin M. Cruz¹, Hsin-Yi Chen¹, Patrick Ryan¹-³, George Hripcsak¹
¹Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, ²Department of Medicine, Weill Cornell Medical College, New York, NY, ³Janssen Research & Development, Titusville, NJ

Background
Adverse drug reactions (ADRs) are a common cause for emergency department visits, hospitalizations and ambulatory visits and cost healthcare systems billions of dollars, yet there is still little known about the real-world incidence of possible side effects following initiation of a drug.¹ Current strategies to identify ADRs, such as randomized control trials, observational studies and adverse event reporting systems,² identify some side effects for some drugs, but the majority of potential drug-side effect pairs remain untested.

While not causal, incidence rates still have clinical value as seen in prior work with OHDSI, which for example calculated incidence rates of clinical conditions that can occur after COVID-19 vaccination. The results helped inform the European Medicines Agency (EMA) decision to reinstate the AstraZeneca COVID-19 vaccine despite reports of clotting.³,⁴ In this study, we used real world data to calculate the incidence of every clinical condition following initiation of every drug and prepared the results for publication to a website to aid clinicians in decision-making.

Methods
This study is an observational cohort study using eleven databases which had been converted to the OMOP CDM. Analysis was done on electronic health record (EHR) data from Columbia University Irving Medical Center (both inpatient and outpatient) and the Premier Hospitalization database, as well as administrative claims data from IBM MarketScan Commercial Claims and Encounters (CCAE), Medicare Supplemental Beneficiaries (MDCR), Multi-state Medicaid (MDCD), Japan Medical Data Center (JMDC), Optum, and IMS Australia, Germany, and France which has both EHR and claims data. SNOMED CT condition codes were used to define clinical conditions and RxNorm terms for marketed drug ingredients were used to define instances of drug exposure.

Analysis was run in October 2017 and cohorts were based on the first exposure of a drug. Patients required at least one year of data prior to the drug initiation and could not have had prior codes for that drug. The incidence proportion for a clinical condition was defined as the number of patients who have a new occurrence of the condition during the time at risk, which was 1-year post exposure, divided by the number of patients in the cohort. Patients previously diagnosed with the condition prior to drug initiation were excluded. Given patients may not have had data captured during the entire time at risk, we calculated the incidence proportion using only patients who had data entries throughout the entire at-risk period as well as using all patients regardless of how long they were observed post-exposure.

We pre-selected 10 known and studied drug-side effect pairs: 1) lisinopril and angioedema, 2) lisinopril and cough, 3) sertraline and suicidal ideation, 4) sertraline and sexual dysfunction, 5) lamotrigine and stevens-johnson syndrome 6) atorvastatin and muscle pains, 7) levofloxacin and tendon rupture, 8) canagliflozin and urinary tract infection, 9) prednisone and deep venous thrombosis, 10) warfarin and bleeding. We compared the calculated incidences to incidences identified in the literature.
Results

We evaluated 13,005,797 unique drug-outcome pairs from 2,072 drug concepts and 21,433 outcome concepts. Results were displayed on an internal site that allowed users to select a drug and a condition to calculate an incidence rate (Figure 1). If no condition is selected, ADRs extracted from the Food and Drug Administration (FDA) drug label are listed with their corresponding calculated incidences.

Upon searching for a drug-condition pair, a range of incidences is provided with the specific proportions listed for each database that had patients with a new incidence of the condition after drug initiation. Two incidences were calculated for each database; one using only patients who were observed for the entire at-risk period and another using all patients with the drug exposure. Both were included in our target-pair comparisons.

The results for one particular database, Premier Hospitalization Database, which only included data on hospitalized patients, was often significantly higher than the rest of the databases when only using patients who had data for the entire time-at-risk (Figure 2). For our target-pair comparisons, we excluded the Premier results as patients with hospitalizations for the entire 1 year at-risk period likely represent a biased population.
For some adverse outcomes, there were only studies evaluating the entire drug class, such as serotonin reuptake inhibitors (SSRI) instead of sertraline or angiotensin-converting enzyme (ACE) inhibitors for lisinopril. We also did not discriminate based on medication indication, dose, or frequency as our analysis included all patients with the drug exposure.

We found that for our ten target pairs, nine of the calculated incidence ranges overlapped with rates found in the literature (Table 1), which included meta-analyses, systematic reviews, individual randomized control trials and observational studies, and UpToDate. For five pairs, the calculated ranges almost fully subsumed all results found in literature and four pairs had some overlap between the calculated range and range identified through literature. One pair, sertraline and sexual dysfunction did not have any overlap with our calculation significantly underestimating the incidence rate. This may be due to an underrepresentation in billing codes of certain clinical conditions that may be considered sensitive.

### Table 1. Comparison of incidence rates calculated in HowOften project with rates reported in literature and UpToDate.

<table>
<thead>
<tr>
<th>Drug</th>
<th>Condition</th>
<th>Number of Patients in Cohort</th>
<th>HowOften Calculated Incidences (%)</th>
<th>Comparison Sources Incidences (%)</th>
</tr>
</thead>
</table>
| Lisinopril | Angioedema         | 6,693,344                    | 0.12 - 0.77                        | 0.1 - 0.7<sup>5,6</sup>  
0.2<sup>7</sup>  
0.49<sup>8</sup> |
| Lisinopril | Cough              | 5,514,370                    | 3.33 - 13.83                       | 5 - 20<sup>9</sup>  
2.7 - 12.3<sup>10</sup>  
3.9 - 35<sup>11</sup> |
| Sertraline | Suicidal Thoughts | 3,420,513                    | 0.26 - 3.67                        | 0.39 - 1.82<sup>12</sup> |
| Sertraline | Sexual Dysfunction | 3,604,138                    | 0.04 - 1.17                        | 15 - 80<sup>13</sup> |
Conclusion

The calculated incidence proportions for known adverse effects for drugs were largely well-aligned with previous knowledge and suggests that large-scale incidence rate calculations may allow for evaluation of every possible ADR without manual curation. Although not causal, which should be communicated clearly to users, incidence rates, especially low ones, can be clinically relevant in whether the potential effect influences a clinical decision. The clinical condition that was not consistent with literature suggests sensitive conditions may be underrepresented in coding data and additional work should be done to characterize these conditions. There is ongoing work to better define outcome cohorts using both diagnosis codes and phenotyping work by OHDSI workgroups, calculating incidence proportions using different time-at-risks and stratifying by patient characteristics and data source.

References


