Analysis of drug use by dose form in large healthcare databases: Data granularity issues and CDM considerations

Vojtech Huser, MD PhD[1]

[1] Lister Hill National Center for Biomedical Communication, National Library of Medicine, National Institutes of Health

Introduction

Analysis of drug use internationally can be challenging due to differences in approved drug ingredients, different dose forms or different branding. In order to allow multi-site international analyses, medication events need to be represented in a common data model and use the same terminology and granularity level (e.g., ingredient vs. clinical drug). RxNorm terminology provides and maintains drug concepts and relationships for drugs approved in the US and via inclusion of other terminologies, such as ATC in RxNorm, it provides some (but very limited) support for drugs not approved in the US. The OMOP (Observational Medical Outcomes Partnership) Common Data Model version 5 (CDM v5) captures drug events on two levels: (1) drug exposure clinical/branded drug granularity level (and contains constructs for drug delivery route); whereas the derived (2) drug era table targets drug ingredient granularity level (and lacks a drug route construct). For simple drug analyses, the drug era table (organized by ingredients) transcends conveniently many international data differences. We explore an interim drug data representation layer that uses drug ingredients extended with drug route information derived from dose forms (e.g., oral tablet, vaginal gel). The route is inferred from the drug exposure data rows with the help of the OMOP terminology (RxNorm relationships).

Methods

In our pilot study, we consider a case of a healthcare consumer who is interested in understanding the use of a given drug. The OMOP Vocabulary and search capabilities within the OHDSI Atlas tool, allows a consumer to distinguish and link branded drugs to ingredient(s). However, we would like healthcare consumers to also benefit from aggregated observational data (such as pre-computed views from the Observational Health Data Sciences and Informatics1 (OHDSI) Achilles tool for data characterization) that could theoretically allow consumers to also understand the real word use of a drug or drug ingredient (in addition to mere terminology-based insights). In such a scenario, the consumer may be interested in the lowest possible complexity and may prefer to always target the drug ingredient view (drug era table). In our preliminary analysis, we used Truven MarketScan Commercial Claims and Encounters (CCAE) claim-based dataset. We used the research lab within the Innovation in Medical Evidence Development and Surveillance (IMEDS) program of the Reagan-Udall Foundation (RUF) for the Food and Drug Administration (FDA). In CCAE, drug_era table contains 7 times fewer drug concepts (1,854 distinct drug terms) as opposed to drug_exposure table (with 12,989 distinct drug terms). Although the OMOP CDM includes the drug_route_concept_id column in drug_exposure table, our CDM v5-shaped claims-based CCAE dataset did not populate this column (which is why we saw a need to infer it). We used SQL based data transformation and R language to implement the healthcare consumer data view described above.

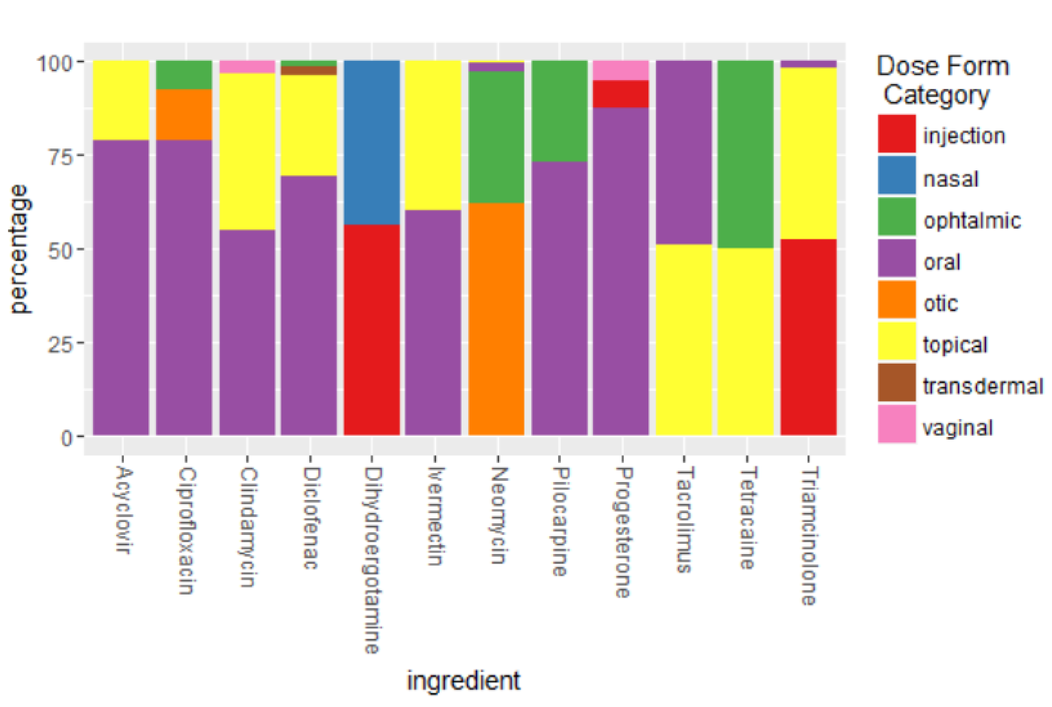

Figure 1: Use of selected drug ingredients by dose form category.

OMOP vocabulary relationships (originating mostly from RxNorm), such as 'has dose form' and 'has ingredient', were used to convert clinical/branded drugs to individual ingredients and pair those with a dose form.

Results

CCAE version 5 dataset within IMEDS cloud lab contains data for over 165 million patients for the period of January 2003 till March 2015. The OMOP terminology used by our project was 'v5.0 3-Apr-2015'. To simplify the implementation (at our site and also possibly for other sites repeating our analysis), we structured the analysis to utilize Achilles pre-computed views for drug data (analysis '704: Number of persons with at least one drug exposure, by drug_concept_id by calendar year by gender by age decile' that we further aggregated to only year level). To account for changing number of patients captured by the CCAE database, we divided raw drug prescription patient counts by the database population for a given year (utilizing again Achilles analysis '109: Number of persons with continuous observation in each year'). We stratified the data views supporting the consumer drug searches by year. The data presented below are for year 2014.

The drug_exposure data extended with inferred dose form contained a total of 74 valid dose forms (out of 108 possible dose forms defined in the 2016-05 release of RxNorm). To improve final data display, we further grouped the number of dose form into categories. For example, 'otic' category grouped dose forms of 'Otic Solution' and 'Otic Suspension'. Although RxNorm provides its own groupings of dose forms, we found it not suitable for our purpose and have used our own manually defined categorization. Because this was only a demonstration pilot, our custom categorization was not comprehensive and did not try to cover all possible RxNorm dose forms but rather tried to merely reduce the number of dose forms presented to the consumer. Figure 1 shows a selected subset of ingredients with y axis representing the percentage of ingredient by dose form category. For example, a consumer may arrive at the acyclovir ingredient and see that 79% of its use is oral and 21% is topical. The source code for our analysis (in SQL RedShift dialect and R) and selected outputs are available at https://github.com/vojtechhuser/OHDSI-drug-route.

Discussion

Our pilot experiment demonstrates the value of route information (or inferred dose form) at the ingredient level. We acknowledge that the drug view we designed may only be beneficial to consumers since clinicians and healthcare researchers are typically intimately familiar with drug route options for a particular ingredient of their interest. An additional motivation for our research, beyond the consumer focus, is the ability to exclude drug events involving non-systemic drugs. The preliminary analysis of the extended ingredient view shows that majority of non-systemic use can be identified from inferred dose form data with the exception of some ingredients where delineation of systemic/non-systemic effect would have to done at clinical drug level (e.g., certain topical or rectal ingredients/dose forms).

Acknowledgement: This research was supported the Reagan-Udall Foundation for the FDA project IMEDS-SA-0011 and the Intramural Research Program of the National Institutes of Health (NIH)/National Library of Medicine (NLM)/Lister Hill National Center for Biomedical Communications (LHNCBC)

Contact: vojtech.huser@nih.gov