

Mapping U.S. FDA National Drug Codes to Anatomical-Therapeutic-Chemical Classes using RxNorm

Fabrizio S. P. Kury, MD¹, Olivier Bodenreider, MD, PhD¹

¹U.S. National Library of Medicine, U.S. National Institutes of Health, Bethesda, MD, USA

Background and Methods

U.S. Food and Drug Administration National Drug Codes (NDCs) are the official identification of drug products in the U.S. market and therefore commonly appear in large drug prescription datasets from this country. For analyzing those datasets, however, it is usually the case that one wishes to identify drugs by classes, rather by NDCs, even if only because the latter is different according to drug manufacturer, dosage forms and packaging. The Anatomical-Therapeutic-Chemical¹ drug classification system is particularly convenient for analysis of large datasets because it necessarily provides a 4 levels of aggregation for each drug – ATC-5 is the drug itself, while ATC-4, 3, 2, and 1 are classes. However, neither ATC nor the FDA provides a map from NDCs to ATC classes. Here we demonstrate how to use the U.S. National Library of Medicine RxNorm² application programming interface to create such a map, and present the statistics of mapping of 71,309 NDCs found in Medicare Part D claims from 2006 to 2013 and 134,580 NDCs found in a commercial all-payer claims dataset from Partners Healthcare from 2011-2012.

Methods and Results

We produced an R script and made it freely available under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license in the GitHub repository at https://github.com/fabkury/ndc_map. The script ingests a table of NDCs, queries the RxNorm API to obtain their RxNorm Concept Unique Identifiers (RxCUIs), then queries the API again to obtain the ATC-4 class or classes of the RxCUIs. Therefore, two mapping issues can occur: NDCs not recognized by RxNorm, and NDCs with RxCUIs without any assigned ATC-4 class. Tables 1 and 2 provides statistics on the counts of NDCs from the mapping of the two datasets.

Table 1. Results of the mapping process.

	Medicare	All-payer
Total unique NDCs	71,309	134,580
NDCs with no RxCUI	10,413 (14.6%)	29,160 (21.7%)
RxCUI but no ATC-4	5,331 (7.5%)	7,459 (5.5%)
Unique NDC—ATC-4	114,069	187,426
Total ATC-4 classes	517 (58.62%)	540 (61.2%)

Table 2. Tallies of ATC-4 classes per NDC.

# ATC-4	Medicare	All-payer
0	15,744	36619
1 to 2	45266	78414
3 to 8	8887	17667
9 to 21	1412	1880
>= 22	0	0

Discussion

Medicare claims contained a relatively limited set of NDCs which were proportionally better covered by RxNorm. One individual NDC could be associated with up to 21 ATC-4 classes in both datasets. This ambiguity happened for two reasons. First, some drugs have more than one active ingredient, and each ingredient has its own ATC class or classes. Second, ATC can assign multiple classes to one same ingredient depending on its usage. For example, *miconazole* has 6 classes in ATC-4 depending on whether the drug is for oral use, gynecological use, ontological use, and so on, but the mapping process does not take that into consideration. Another example, *bacitracin* has 3 ATC-4 classes and polymyxin B has 5 classes, so the drug bacitracin/polymyxin B has 8 classes. If not addressed, this ambiguity can readily duplicate data when the map is applied to common database operations. In the Medicare dataset, only 77.9% of the NDCs were successfully mapped to at least one ATC class, but those NDCs corresponded to 97.6% of all claims in the dataset, denoting that the more popular a given NDC is, the more likely it is to be recognized by RxNorm and have an assigned ATC class.

References

1. World Health Organization (WHO) Collaborating Centre for Drug Statistics Methodology. ATC/DDD Index 2016. 2016. Accessed from http://www.whocc.no/atc_ddd_index/ on March 9, 2017.
2. U.S. National Library of Medicine. RxNorm APIs. Accessed from <https://rxnav.nlm.nih.gov/APIsOverview.html> on March 9, 2017.