MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230926

A High-Fidelity Combined ATC-Rxnorm Drug Hierarchy for Large-Scale Observational Research

Anna OSTROPOLETS[a,b,1], Polina TALAPOVA[c], Marcel DE WILDE[d], Hamed ABEDTASH[e] , Peter RIJNBEEK[d] and Christian G REICH[d,f]
[a] *Columbia University Irving Medical Center, NY, USA*
[b] *Odysseus Data Services Inc, Cambridge, MA, USA*
[c] *Sci-Force, Kharkiv, Ukraine*
[d] *Erasmus University Medical Center, Rotterdam, The Netherlands*
[e] *Bristol Myers Squibb, Lawrence Township, NJ, USA*
[f] *Northeastern University, Portland, ME, USA*

Abstract. Observational research utilizes patient information from many disparate databases worldwide. To be able to systematically analyze data and compare the results of such research studies, information about exposure to drugs or classes of drugs needs to be harmonized across these data. The NLM's RxNorm drug terminology and WHO's ATC classification serve these needs but are currently not satisfactorily combined into a common system. Creating such system is hampered by a number of challenges, resulting from different approaches to representing attributes of drugs and ontological rules. Here, we present a combined ATC-RxNorm drug hierarchy, allowing to use ATC classes for retrieval of drug information in large scale observational data. We present the heuristic for maintaining this resource and evaluate it in a real world database containing drug and drug classification information.

Keywords. ATC, RxNorm, drug surveillance, drug safety, observational research

1. Introduction

Over the past decade, observational data have been extensively used in clinical research and has shown an ability to impact decision-making both for clinicians and for regulatory bodies [1]. Distributed observational data networks such as Sentinel, OHDSI or EU-ADR can address drug surveillance and effectiveness questions not covered by the randomized clinical trials [2-3]. These data contain drug information commonly encoded as National Drug Codes (NDC), Medi-Span's Generic Product Identifiers (GPI), Cerner's Multum, FDB's Clinical Formulation IDs (GCN_SEQNO), VA's VA_Product identifiers or free text [4]. All these are aggregated into RxNorm, provided by the National Library of Medicine (NLM), and used in OHDSI's OMOP CDM. In Europe, apart from local ontologies, the WHO's Anatomical Therapeutic Chemical (ATC) classification [5,6] is most popular. Ideally, ATC and RxNorm should be combined to create a common reference terminology, allowing full interoperability of data and global research networks.

[1] Corresponding Author: Anna Ostropolets, Columbia University, 622 West 168th Street, PH-20, New York, USA, email: ao2671@cumc.columbia.edu.

There are two possible approaches to achieve this goal. One is provided by the NLM [7], attaching the lowest ATC 5th-level concepts to RxNorm active ingredients. This approach results in massive misclassification of RxNorm drug concepts when ATC's route, dose, indication and drug combination information is lost. E.g., prednisolone exists as an agent in dermatology (ATC, D07AA03) or ophthalmology (S01CB02), system immunosupression (H02AB06), topical vasoprotective (C05AA04) or as a nasal formulation (R01AD02). RxNorm has only one prednisolone (RXCUI 8638). The alternative approach is to utilize the list drug products corresponding to each ATC class, however, the WHO does not make that information available to the public.

In this paper, we propose and evaluate a semi-automated RxNorm-to-ATC mapping process and analyze the challenges associated with aligning drug ontologies for observational research. We performed this task using RxNorm and RxNorm Extension, which we maintain to cover international drugs under the same RxNorm system [8].

2. Methods

The goal of this work was to create a joint ATC-RxNorm hierarchy, where ATC terms serve as parents or ancestors and the fully detailed RxNorm concepts as children or descendants. The process has three steps: (1) completion of ingredient and addition of ATC route of administration [RoA] to RxNorm Dose Form mapping, (2) creation of a heuristic for mapping of ATC to drug products, and (3) filling in the missing parts of the hierarchy between ingredients and drug products.

2.1. Attribute mapping

We first introduced NLM's ATC to RxNorm ingredients mapping and added missing links. For ambiguous ATC concepts such as A12CB03 "zinc protein complex" we introduced crosswalks to all possible RxNorm ingredients and matched them with a precedence score based on clinical plausibility. Multicomponent drugs in ATC were split into the components and processed separately. E.g., G03AB08 "dienogest and estradiol" was broken down and mapped to RxNorm 22968 "dienogest" and 4083 "estradiol".

We also mapped all RoA to corresponding RxNorm Dose Forms. E.g., Nasal RoA in ATC was mapped to RxNorm 316962 “Nasal Solution”, 126542 “Nasal Spray” etc. For ATC 5th concepts not explicitly stating the RoA we inferred this information from their ATC hierarchical ancestors. E.g., for the R01AD02 "prednisolone" we gleaned the Dose Forms from R01A "Decongestants and other nasal preparations for topical use".

2.2. ATC-drug product mapping

We then matched ATC 5th-level codes to RxNorm drug products using the attribute maps from step 1 and the RxNorm hierarchy. Mono-ingredient ATCs were mapped to all descendant mono-ingredient drugs of the corresponding RxNorm ingredient. ATCs with no RoA information were also mapped to descendent of the RxNorm ingredient. If ATC codes had RoA, they were mapped to drugs in the hierarchy below RxNorm Clinical Drug Forms. ATC Combinations were mapped similarly.

This process leads to multiple overlapped and collisions, for which we established a ranking system to prioritize matching based on attribute complexity and the mapping plausibility scores. First, we matched precise combinations where all components were

defined (e.g., N02AJ13 "tramadol and paracetamol"). Then we matched the combinations with broader groups (e.g., N02BE71 "paracetamol, combinations with psycholeptics"). Lastly, we matched mono-ingredient ATC 5th-level concepts (N02BE01 "paracetamol") to the remaining unclaimed combinations containing the ingredient.

2.3. Completion of hierarchy between ATC and RxNorm

A joint ATC-RxNorm hierarchy was constructed following strictly the established mappings between ATC and RxNorm. The result differs from the ATC or RxNorm internal hierarchies, in which all lower-level descendants belong to an ancestor if they are connected through a path of intermediate hierarchical concepts. E.g., the hierarchy of an ATC with a parenteral RoA would pass through the generic RxNorm ingredient, but then continue only to those Drug Forms and products that have an appropriate Dose Form.

2.4. Evaluation

We assessed the performance of our approach on the Integrated Primary Care Information Project (IPCI) database [9]. It contains longitudinal data of Dutch patients visiting general practitioners (GPs). Drug data are coded through G-Standaard Z-index, which happens to provide both ATC and drug products information mapped to RxNorm. This allowed us to analyze the quality of our approach and the impact of potential mismatches on actual patient records.

3. Results

3.1. ATC coverage

Of 5,223 valid ATC 5th-level concepts 4,656 (89.2%) were linked to RxNorm ingredients. Unmapped codes included ingredients or combinations not present in RxNorm or European drug ontologies that had contributed to RxNorm Extension (e.g., A10BD12 "pioglitazone and sitagliptin"), and ingredients outside of the scope of RxNorm.

3.2. RxNorm coverage

Our and NLM approaches showed similar coverage: of 34,691 Clinical Drugs in RxNorm, 27,629 (79.6%) were linked to ATC 5th-level using our approach and 27,077 (78.1%) - using NLM approach. Examples of added pairs included combinations (e.g., R01AD59 "mometasone, combinations" to 417615 "mometasone 0.001 MG/MG / salicylic acid 0.05 MG/MG Topical Ointment") and mono-ingredients (L01XC31 "avelumab" to 187,5548 "avelumab 20 MG/ML Injection"). We linked 50,753 (65.3%) out of 77,704 RxNorm Extension Clinical Drugs to ATC, which provided sufficient coverage for the most important international drugs not in the USA-only RxNorm.

3.3. Crosslinks

ATC to RxNorm drug assignment differed for more than half of the drugs (n=15,425). E.g., the five ATC concepts for prednisolone discussed earlier were linked to 605

Clinical Drugs using the NLM approach, but only 85 links in our approach. C05AA04 "prednisolone" (which is a rectal RoA) was linked to oral, rectal, ophthalmic, and injectable products of prednisolone in NLM and exclusively to rectal formulations in our approach.

3.4. Effect on classifying patient exposure

Our approach provided sufficient coverage of patient data with 97.9% of ingredients and 99.8% of patient records covered. When looking at the ATC assignment by the source and our approach (Figure 1), it aligned for 64.7% of Clinical Drugs, which accounted for 77.5% of the patient records.

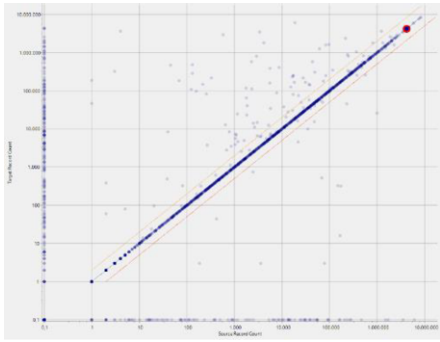

Figure 1. Comparison of ATC code assignment by the data source (x-axis) and the OHDSI Standardized Vocabularies (y-axis), count of records. A dot represents an ATC 5th-level concept.

A substantial proportion of discrepancies accounted for several ATC codes assigned by our approach and one ATC code assigned by the data source. E.g., Atenolol 12.5 MG Oral Tablet was assigned to C07AB03 "atenolol" in the Dutch data source, while we ascribed it to C07AB03 "atenolol" as well as C07AB11 "s-atenolol". In many of these cases, ATC provided little information to make an unambiguous distinction between two ATC concepts on ingredient and RoA information alone. Finally, discrepancies occurred when the data source assigned higher-level ATC codes. E.g., "Mitomycin 0.4 MG/ML Ophthalmic Solution" was assigned L01DC03 "mitomycin" by our approach and S01XA "Other ophthalmologicals" by the data source.

4. Discussion

Here, we presented a semi-automated ATC-RxNorm alignment process, which enabled a more accurate ATC assignment based on ingredient, ingredient combination and RoA information. The joint hierarchy facilitates large-scale phenotyping and covariate construction by enabling researchers to define exposures as drug classes rather than individual drugs.

ATC provides sufficient coverage of the mono-ingredient drugs. But its approach to combination drugs is inconsistent: sometimes it matches single ingredients with drug classes (C09DA10 "fimasartan and diuretics") and other times with other unspecified ingredients (C07FX01 "propranolol and other combinations"). This leads to the possibility of overlap between loosely defined combinations (e.g., a combination of tetracaine, lidocaine, and epinephrine can be classified under N01BA53 "tetracaine, combinations" or N01BB52 "lidocaine, combinations").

It does not help that ATC does not provide a computable form of its assignment rules other than short descriptions in English. RoA is defined at varying hierarchical levels, and inheritance rules to descendants are not explicitly defined. RoA or therapeutic area ("ophthalmologics") is rarely mentioned in the descriptions of the ATC 5th-level concept name, which is why many analysts fail to realize these attributes exist.

An attribute we did not take into consideration at all is the provided recommended daily dose (DDD). Sometimes, it is the only difference between two ATC classes. Drug products with a defined ingredient strength do not provide us with sufficient information about their daily dose since we lack the signature of the prescription (e.g., "twice daily"). Therefore, unfortunately DDD cannot be used for our heuristic approach.

All the above obstructs systematic and consistent classification and can influence patient selection if ATC concepts are used without due diligence. While our solution enables more accurate code assignment, further research is needed to unambiguously assign ATC codes that represent complex drug combination. Data sources can have imperfect assignment, so that it requires both extensive data diagnostics and deep knowledge of ATC to re-assign the codes.

5. Conclusions

We developed a semi-automated ontology alignment process that allows to create a joint ATC-RxNorm hierarchy while preserving dose, route of administration and ingredient alignment. This enables using ATC as a classification system for drug products from the US and international markets for exposure definition and covariate construction.

References

[1] Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. Ann Intern Med. 2009 Aug;151(3):203-5, doi: 10.7326/0003-4819-151-3-200908040-00125.

[2] Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong IC, Rijnbeek PR, van der Lei J, Pratt N, Norén GN, Li YC, Stang PE, Madigan D, Ryan PB. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;216:574-8, doi: 10.3233/978-1-61499-564-7-574.

[3] Ball R, Robb M, Anderson SA, Dal Pan G. The FDA's sentinel initiative--A comprehensive approach to medical product surveillance. Clin Pharmacol Ther. 2016 Mar;99(3):265-8, doi: 10.1002/cpt.320.

[4] Schwartz JJ, Lee E, Butler AP, Facklam DP, Franks B, Spalding JR, Vassilakis ME, Thal GD, Irish WD. The Association of Tacrolimus Formulation Switching with Trough Concentration Variability: A Retrospective Cohort Study of Tacrolimus Use Post-Kidney Transplantation Based on National Drug Code (NDC) Numbers. Adv Ther. 2019 Jun;36(6):1358-1369, doi: 10.1007/s12325-019-00950-5.

[5] Pratt NL, Kerr M, Barratt JD, Kemp-Casey A, Kalisch Ellett LM, Ramsay E, Roughead EE. The validity of the Rx-Risk Comorbidity Index using medicines mapped to the Anatomical Therapeutic Chemical (ATC) Classification System. BMJ Open. 2018 Apr;8(4):e021122, doi: 10.1136/bmjopen-2017-021122.

[6] Zhou JP, Chen L, Guo ZH. iATC-NRAKEL: an efficient multi-label classifier for recognizing anatomical therapeutic chemical classes of drugs. Bioinformatics. 2020 Mar;36(5):1391-6, doi: 10.1093/bioinformatics/btz757.

[7] Bodenreider O, Rodriguez LM. Analyzing U.S. prescription lists with RxNorm and the ATC/DDD Index. AMIA Annu Symp Proc. 2014 Nov;2014:297-306.

[8] A. Ostropolets, “OHDSI RxNorm Extension: 5 major drug markets in RxNorm notation, comparison and lessons learned,” presented at the DailyMed/RxNorm Jamboree Workshop, 2017.

[9] de Ridder MAJ, de Wilde M, de Ben C, Leyba AR, Mosseveld BMT, Verhamme KMC, van der Lei J, Rijnbeek PR. Data Resource Profile: The Integrated Primary Care Information (IPCI) database, The Netherlands. Int J Epidemiol. 2022 Dec;51(6):e314-23, doi: 10.1093/ije/dyac026.