Mapping to standardised vocabularies: a process for drug codes in Australia

Ty Stanford, PhD[1], Jodie Hillen, PhD[1], Nicole Pratt, PhD[1]
[1]Quality Use of Medicines and Pharmacy Research Centre, University of South Australia, Adelaide, South Australia, Australia

Abstract

A crucial step in the transformation of observational data into the OMOP common data model (CDM) is the translation of native vocabularies to their respective OHDSI 'standardized vocabularies'. We undertook the translation of an Australian drug prescription dataset, a 10% sample of the national Pharmaceutical Benefits Scheme data, to the OMOP CDM and report her on the process undertaken and the results. A large coverage of patient records was achieved using the OHDSI provided mappings without manual intervention (94.4%). Manual mapping of the remaining records using the OHDSI tool Usagi increased the patient record coverage to 99.5%. We share our process and solutions to unexpected issues in the mapping process to contribute to OHDSI data standards and facilitate the translation of Australian datasets to the CDM, currently underrepresented in the OHDSI network.

Research Category (please highlight or circle which category best describes your research)

Observational data management, clinical characterization, population-level estimation, patient-level prediction, other (if other, please indicate)

Introduction: Mapping of clinical data to standardized vocabularies is an important step in creating an OMOP CDM and the data quality of downstream analyses is reliant on correct classification of medicine exposure and disease outcomes. In Australia, the Pharmaceutical Benefits Scheme (PBS) records all subsidized medicines dispensed to patients. A 10% sample of the nationwide PBS prescription data (PBS10) are available to researchers and this dataset has been used widely for drug utilization studies[1] and policy impact analysis.[2] The PBS10 data have previously been mapped to OMOP CDM to investigate the use of medicines in children.[3] Here we outline our process of updating the mapping the PBS data to OMOP CDM with the aim to (a) provide further documentation of the PBS vocabulary mapping for Australian data, (b) describe the coverage mapping of PBS medicine codes to the OHDSI standard vocabulary, and (c) encourage feedback on our mapping process to establish mapping conventions and further contribute to the OHDSI community.

Objective: As PBS medicine codes are very granular and each medicine code uniquely describes the product, manner of administration, form and strength we aimed to map the native PBS medicine codes to the corresponding OHDSI standardized vocabulary for drugs in the OMOP CDM without loss of the drug information granularity that the PBS source codes represent.

Methods: Mapping of the PBS non-standard vocabulary to an OHDSI standard vocabulary (referred herein as *mapping process*) required several steps. An overview of steps (a) and (b) are depicted in Figure 1.

a) Mapping PBS item codes to the AMT (Australian Medical Terminology) vocabulary.[4]
b) Mapping of the AMT vocabulary to the RxNorm Extension vocabulary using the OHDSI '`concept_relationship`' table.
c) Where the above steps did not find a suitable match - manual mapping of PICs to RxNorm Extension '`concept_id`'s were made using the OHDSI tool Usagi.[5]

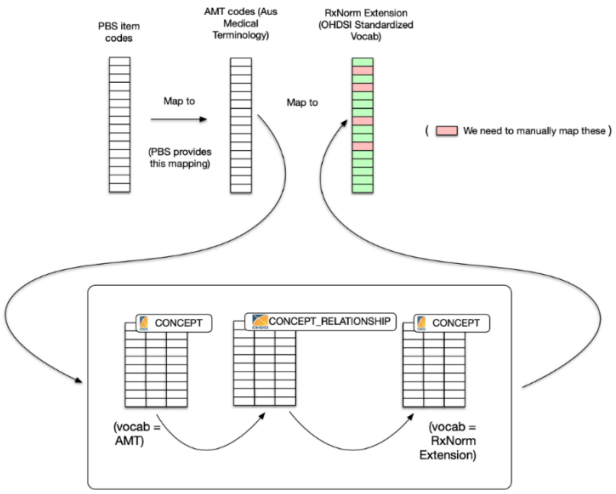

Figure 1. The mapping of PBS (Pharmaceutical Benefits Scheme) data to the OHDSI standardized vocabulary of RxNorm Extension.

Results: Of the 3692 unique PBS medicine codes available in Australia, 74% were directly mapped to the RxNorm Extension standard concepts which accounted for 94% of all medicines dispensed (Table 1). Of the remaining 969 PBS codes, 430 (11.6%) were mapped manually using Usagi, accounting for 5.1% of medicines dispensed. The remaining codes accounted for less than 0.5% of medicines dispensed.

Table 1. Patient record and PBS item code capture through the mapping process.

Step	PBS item codes, *n* (%)	Patient records, *n (%)*
OHDSI assisted automatic mapping (Figure 1)	2 723 (73.8%)	39 192 527 (94.4%)
Manual mapping via Usagi	430 (11.6%)	2 098 247 (5.1%)
Total records mapped to standard concept	**3 153 (85.4%)**	**41 290 774 (99.5%)**

To validate our manual mappings made in Usagi, we used three raters (one pharmacist and two non-pharmacists) to independently map common PBS Medicine codes. Of the 38 randomly selected codes, there was majority agreement in 92.1% (35/38) of cases. Disagreement in mapped '`concept_id`'s between the raters was overwhelmingly a result of concepts with slightly differing descriptions in their form such as "Ophthalmic Suspension" as opposed to "Ophthalmic Solution".

Challenges encountered in the mapping process included: missing information in official PBS description files requiring web scraping of additional information, multiple source codes in the AMT vocabulary for PBS Medicine codes and historical PBS medicine codes in the PBS10 data.

Conclusion: The process of mapping Australian specific PBS medicine codes to the RxNorm vocabulary, facilitated through the Australian Medicines Terminology codes, resulted in 85% of codes with an associated standard concept code. The remaining medicines were rarely dispensed in practice, the most common of these was products with A12 (mineral supplements) Anatomical Therapeutic Classification codes. We aim to make this mapping process available to researchers in Australia through the Transformational Data Collaboration to assist in the uplift of datasets in Australia.

References

1. Mellish L, Karanges EA, Litchfield MJ, Schaffer AL, Blanch B, Daniels BJ, Segrave A, Pearson SA. The Australian Pharmaceutical Benefits Scheme data collection: a practical guide for researchers. BMC research notes. 2015 Dec 1;8(1):634.
2. Kemp-Casey A, Pratt N, Ramsay E, Roughead EE. Using post-market utilisation analysis to support medicines pricing policy: an Australian case study of aflibercept and ranibizumab use. Applied health economics and health policy. 2019 Jun 1;17(3):411-7
3. Brauer R, Wong IC, Man KK, Pratt NL, Park RW, Cho SY, Li YC, Iqbal U, Nguyen PA, Schuemie M. Application of a Common Data Model (CDM) to rank the paediatric user and prescription prevalence of 15 different drug classes in South Korea, Hong Kong, Taiwan, Japan and Australia: an observational, descriptive study. BMJ open. 2020 Jan 1;10(1).
4. PBS Downloads [Internet]. Canberra (ACT): Pharmaceutical Benefits Scheme; 2020 July 1 [cited 2020 July 16]. Available from: https://www.pbs.gov.au/browse/downloads.
5. Observational Health Data Sciences and Informatics. The Book of OHDSI [Internet]. 2019. Chapter 6, Extract Transform Load. [cited 2020 Jul 16]. Available from: https://ohdsi.github.io/TheBookOfOhdsi/ExtractTransformLoad.html.