

Standardizing Austrians Claims Data using the OMOP Common Data Model: A Feasibility Study

Andrea HABERSON^{a,1}, Christoph RINNER^a and Walter GALL^a
^a*Center for Medical Statistics, Informatics and Intelligent Systems,
Medical University of Vienna*

Abstract. The suitability of the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) for Austrian pseudonymized claims data from social security institutions and information about hospital stays is evaluated. 1,023 (99.7%) of ATC codes and 3,695 (98.6%) of ICD10 codes coincide with the OMOP vocabulary. Mappings for the local vocabularies like the Austrian pharmaceutical registration numbers, the Socio-Economic Index and professional groups, to the OMOP vocabulary do not exist. A standardization with the OMOP CDM is possible, however an initial, not negligible effort is required to adapt and incorporate the vocabulary.

Keywords. standardized health data, secondary use, claims data, drug safety, common data model, OMOP

1. Introduction

Joint analysis of data from various data sources results in a large number of time-consuming and cost-intensive, non-reusable processes due to the different types of data models, coding systems, and terminologies used [1]. The use of claims data for health analysis is an application of secondary use of data [2]. In the case of the Austrian health claims data the same data are repeatedly reprocessed in different studies and various assumptions make it difficult to objectively compare different study results.

This article evaluates whether Austrian health claims data can be standardized using the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) as well as the required effort and the extent of necessary expert knowledge. The project „Adverse drug events in relation to inappropriate medication of geriatric patients with renal insufficiency - a retrospective register-based cohort study” (ADE-PIM) is used as an application case.

2. Methods

The Main Association of Austrian Social Security Institutions (HVB) brings together claims data of all insured inhabitants in the Austrian province of Lower Austria in the

¹ Corresponding Author, Andrea Haberson, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria; E-mail: a.haberson@gmail.com

years 2008 to 2011 from various social security institutions (i.e. MBDS data) in the database GAP-DRG and makes them available for research projects on health care issues. For studies, all local vocabularies must be prepared in such a way that all relevant medical terms are linked to equivalent concepts of the OMOP vocabulary. Thus, the data preparation process begins with the identification of relevant medical terms and their categorizations. The amount of effort needed for this process depends on the particular state of the vocabulary, which has to be incorporated. The CDM is supported and improved by the international, interdisciplinary Observational Health Data Sciences and Informatics (OHDSI) community and currently provides 81 OMOP vocabularies [3].

3. Results

The study cohort consists of more than 11,500 patients with a diagnosis of renal insufficiency in the years 2008 to 2011. The local vocabularies ICD10 and ATC are included in the OMOP vocabulary. From a total of 3,749 different ICD10 diagnoses in the cohort, 3,695 codes (98.6%) match the ICD10 OMOP vocabulary. Out of a total of 1,026 different local ATC codes in the cohort, 1,023 (99.7%) coincide with the ATC OMOP vocabulary. Local vocabularies like the Austrian pharmaceutical registration numbers, the Socio-Economic Index and professional groups do not exist in the OMOP vocabulary.

4. Discussion

No standard concepts for the mapping of the Socio-Economic Index could be found and for a large part of the Austrian pharmaceutical registration numbers no equivalent standard concept exists yet. In general, it is possible to model all local terms in the structures of OMOP, but this requires technical and temporal resources. Experts must be able to assess the equivalence in the fields of diagnosis and medication. Furthermore, in consultation with the OHDSI community, new concepts must be integrated and published in the OMOP CDM. According to this approach, it is possible to standardize the Austrian health claims data in an OMOP compliant manner. Standardizing these data using a common data model is highly recommended for reasons of the comparability of studies, the reusability of methods, and the potential for costs and time savings. In the long run, the benefits of standardization and reproducibility of research should outweigh this initial drawback.

5. References

- [1] O.I. Ogunyemi, D. Meeker, H.E. Kim, N. Ashish, S. Farzaneh, A. Boxwala, Identifying appropriate reference data models for comparative effectiveness research (CER) studies based on data from clinical information systems, *Med Care* 51(8 Suppl 3) (2013), 45-52.
- [2] C. Safran, M. Bloomrosen, W.E. Hammond, S. Labkoff, S. Markel-Fox, P.C. Tang et al., Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper, *J Am Med Inform Assoc.* 14(1) (2007), 1-9.
- [3] Y. Xu, X. Zhou, B.T. Suehs, A.G. Hartzema, M.G. Kahn, Y. Moride et al., A Comparative Assessment of Observational Medical Outcomes Partnership and Mini-Sentinel Common Data Models and Analytics: Implications for Active Drug Safety Surveillance, *Drug Saf.* 38(8) (2015), 749-65.