

# Evaluation and Challenges of Medical Procedure Data Harmonization to SNOMED-CT for Observational Research

Ines REINECKE<sup>a,1</sup>, Michael KALLFELZ<sup>b</sup>, Martin SEDLMAYR<sup>a</sup>, Joscha SIEBEL<sup>a</sup> and Franziska BATHELT<sup>a</sup>

<sup>a</sup> Carl Gustav Carus Faculty of Medicine, Center for Medical Informatics, Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany

<sup>b</sup> Odysseus Data Services GmbH, Berlin, Germany

**Abstract.** The relevance of health data research on real world data (RWD) is increasing. To prepare national RWD for international research, harmonization with standard terminologies is required. In this paper, we evaluate to what extent the German OPS vocabulary in OHDSI covers codes present in RWD and mappings to SNOMED-CT. The evaluation identified a mapping gap of 21.1% in the RWD set.

**Keywords.** OHDSI, OMOP, SNOMED-CT, OPS, interoperability

## 1. Introduction

Real world data (RWD) empowers researchers to identify patterns in large healthcare data sets with the purpose to improve diagnoses and treatments. Hence, the relevance of observational research networks is increasing in Europe and around the globe [1-3]. The Observational Medical Outcomes Partnership (OMOP) CDM has become widespread in the recent years and can ensure the transferability and the comparability of results [4]. Mapping of national terminologies to international standards is crucial for participation in research networks on OMOP. This paper aims to evaluate the coverage of German OPS codes present in a RWD set to standardized vocabularies provided by OHDSI.

## 2. Methods

The procedure data was provided by the University hospital Dresden (UKD). The data set contains around 1.8 million procedure codes. Data analysis consisted of (a) calculating the mapping relationship between OPS procedures and SNOMED-CT concepts to identify mapping relationships and mappings other than mapping exactly one OPS to one SNOMED-CT code and (b) quantitative analysis to determine the frequency of procedure occurrence and the details on the RWD vocabulary coverage.

---

<sup>1</sup> Corresponding Author, Ines Reinecke, Carl Gustav Carus Faculty of Medicine, Center for Medical Informatics, Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany; E-mail: ines.reinecke@tu-dresden.de

3. Results

The OPS vocabulary contains mappings to SNOMED-CT for 8,149 OPS codes (Table 1). The RWD under review contains 14,041 distinct OPS codes. The vocabulary does not provide a mapping for 10,044 OPS codes that have been present in the data set. Taking the frequency of the recorded procedures in the data set into account, the “no map” group represents 21.1% of the complete data set. 60.3% of the procedure data in the RWD set is covered by the OPS vocabulary with a 1-1 mapping, with a small number of concepts covering a large percentage of the data.

**Table 1.** Descriptive statistics of procedure usage and vocabulary details by mapping

mapping	OPS codes (vocabulary)	OPS codes (RWD )	procedure occurrences		descriptive statistics of procedure code frequency			
			total	percentage	mean	min	max	std
1-1	7493	3,413	1,113,835	60.3%	326	1	104,277	2,452
1-2	552	497	258,737	14.0%	521	1	26,082	1,740
1-3	92	77	83,207	4.5%	1081	1	24,780	3,403
1-4	12	10	1,820	0.1%	182	4	1,047	308
no map	30,306	10,044	390,496	21.1%	39	1	26,620	366
total	38,455	14,041	1,848,095	100%	132	1	10,4277	1,324

4. Discussion

We identified a substantial gap in the coverage of OPS code mapping to standard OHDSI concepts when compared to RWD from the UKD. Mapping data based on the currently available OPS vocabulary expose the risk of information loss. For the purpose of data harmonization, it is crucial for all countries to identify, name, understand and assess those risks when using data translated from national to international terminologies. Our next steps will include implementation of additional mappings based on the most frequent OPS procedure codes present in RWD to minimize the known gaps.

5. Conclusion

In this paper we have shown to what extent the RWD set is covered by the existing OPS vocabulary and its mappings to SNOMED-CT. Next steps have been proposed to proceed with closing identified gaps. This work is funded by the German Ministry of Education and Research (FKZ 01ZZ1801A/L).

References

[1] EMA DARWIN [Internet]. [cited 2022 Jan 17]. Available from: <https://bit.ly/EMA-DARWIN>.  
[2] European Health Data Evidence Network – ehden.eu [Internet]. [cited 2022 Jan 17]. Available from: <https://www.ehden.eu/>.  
[3] Schuemie MJ, et al. Principles of Large-scale Evidence Generation and Evaluation across a Network of Databases. *J Am Med Inform Assoc.* 2020;27:1331–1337.  
[4] Reinecke I, Zoch M, Reich C, et al. The Usage of OHDSI OMOP - A Scoping Review. *Stud Health Technol Inform.* 2021;283:95–103.