UNIVERSITY OF TARTU

1632

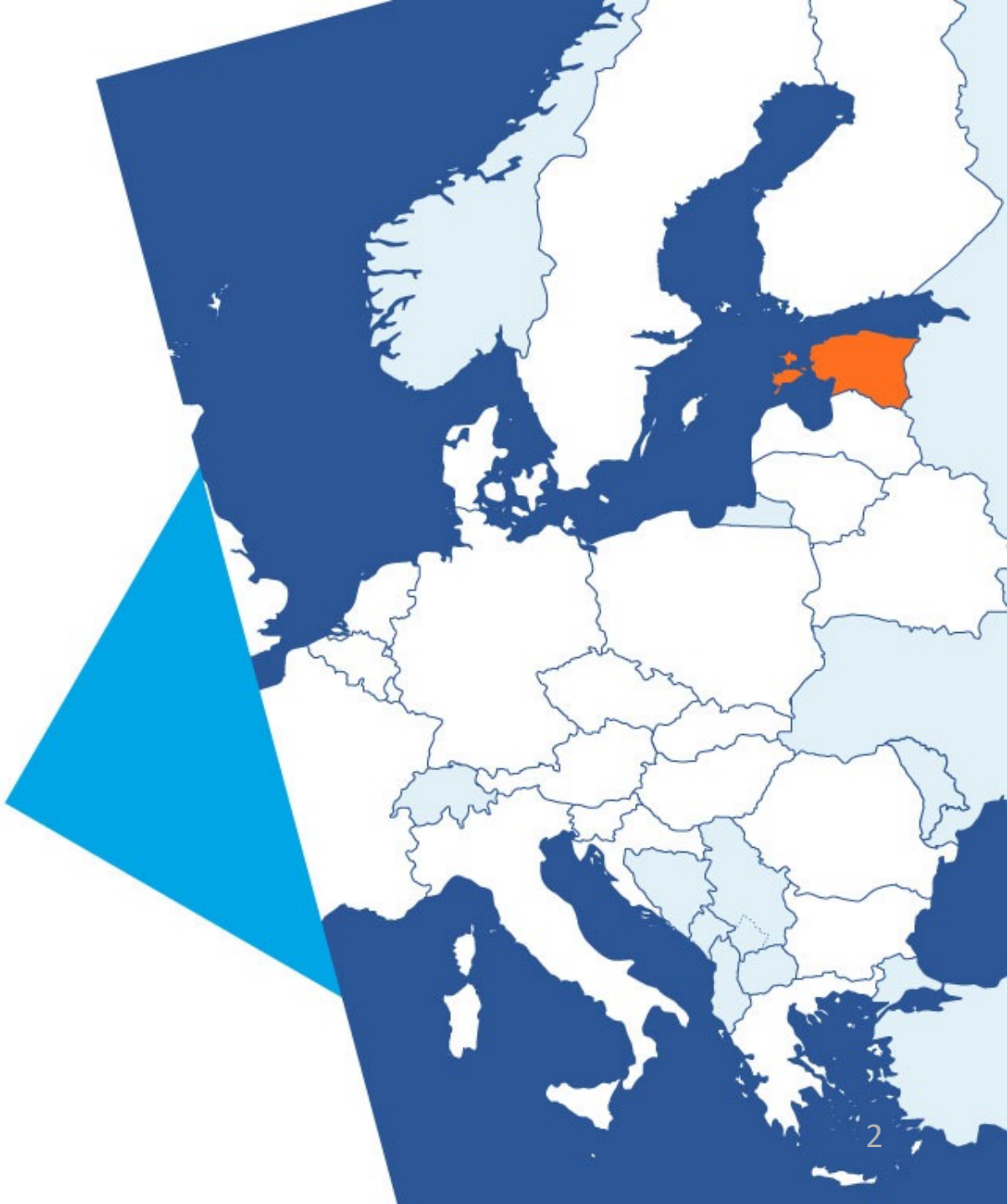
Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned

Sulev Reisberg, Ph.D.

Estonia & University of Tartu

- Small Eastern European country
- Member of European Union and NATO
- Population 1.4 M
- “Estonia is the world's most digitally advanced society” (WIRED)

- University of Tartu - the oldest and largest university in Estonia



Research group of health informatics

Our recent paper

http://doi.org/10.1093/jamiaopen/ooad100

JAMIA Open, 2023, **6(4)**, ooad100
https://doi.org/10.1093/jamiaopen/ooad100
Research and Applications

Research and Applications

Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned

Marek Oja, PhD[1,*], Sirli Tamm, MSc[1], Kerli Mooses, PhD[1], Maarja Pajusalu, MSc[1], Harry-Anton Talvik, MSc[1,2], Anne Ott, MSc[1], Marianna Laht, MD[1], Maria Malk, MSc[1], Marcus Lõo, MSc[1], Johannes Holm, MSc[1], Markus Haug, MSc[1], Hendrik Šuvalov, MSc[1], Dage Särg, MSc[1,2], Jaak Vilo, PhD[1,2], Sven Laur, PhD[1], Raivo Kolde, PhD[1], Sulev Reisberg, PhD[1,2]

[1]Institute of Computer Science, University of Tartu, 51009 Tartu, Estonia, [2]STACC, 51009 Tartu, Estonia

*Corresponding author: Marek Oja, PhD, Institute of Computer Science, University of Tartu, Narva mnt 18, 51009 Tartu, Estonia (marek.oja@ut.ee)

Author Contributions: Dr M. Oja and S. Tamm are considered co-first authors and Dr R Kolde and Dr S Reisberg are considered co-last authors of this work. In addition, they had full access to all the data in the study and take responsibility for the integrity of data and accuracy of the data analysis.

Abstract

Objective: To describe the reusable transformation process of electronic health records (EHR), claims, and prescriptions data into Observational Medical Outcome Partnership (OMOP) Common Data Model (CDM), together with challenges faced and solutions implemented.

Materials and Methods: We used Estonian national health databases that store almost all residents' claims, prescriptions, and EHR records. To develop and demonstrate the transformation process of Estonian health data to OMOP CDM, we used a 10% random sample of the Estonian population ($n = 150\,824$ patients) from 2012 to 2019 (MAITT dataset). For the sample, complete information from all 3 databases was converted to OMOP CDM version 5.3. The validation was performed using open-source tools.

Results: In total, we transformed over 100 million entries to standard concepts using standard OMOP vocabularies with the average mapping rate 95%. For conditions, observations, drugs, and measurements, the mapping rate was over 90%. In most cases, SNOMED Clinical Terms were used as the target vocabulary.

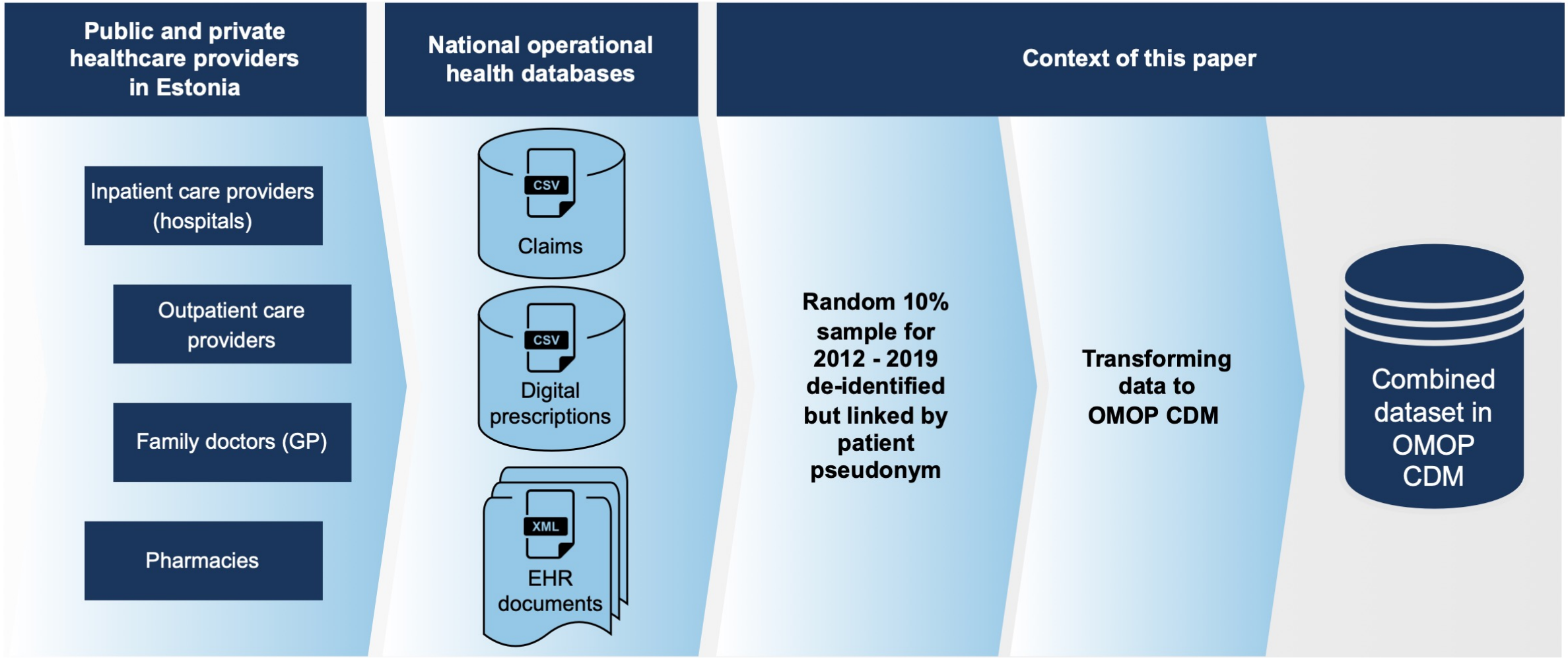
Discussion: During the transformation process, we encountered several challenges, which are described in detail with concrete examples and solutions.

Conclusion: For a representative 10% random sample, we successfully transferred complete records from 3 national health databases to OMOP CDM and created a reusable transformation process. Our work helps future researchers to transform linked databases into OMOP CDM more efficiently, ultimately leading to better real-world evidence.

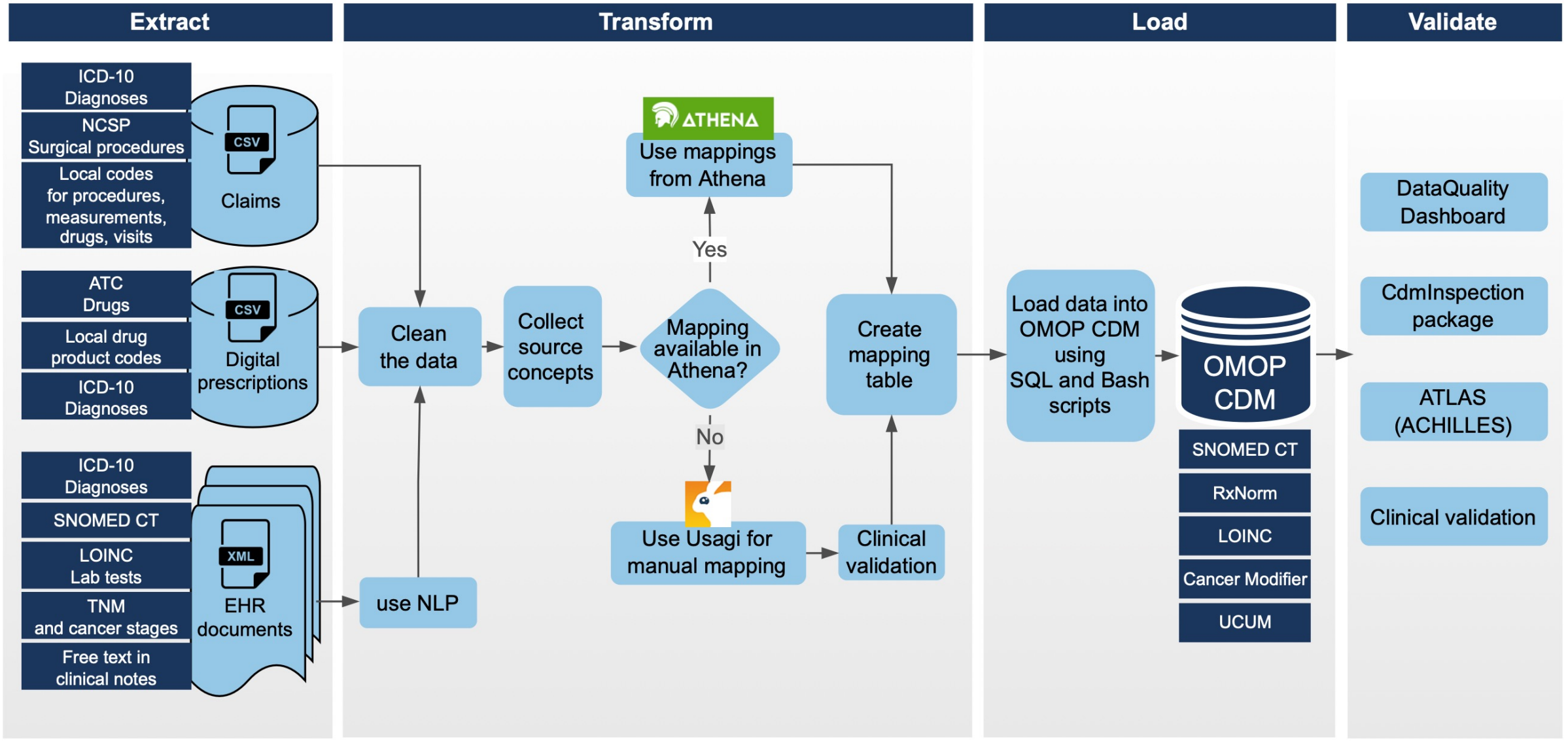
Lay Summary

Health data can be found in various sources and formats, making it challenging for researchers. To address this issue, one possible approach is

Data sources & context of the paper



Overview of ETL



Reviewer 2

> Starting with
> just 10% makes sense,
> but it'd be great
> to have stats
> for the other 90% too.

Highlights

- Extensive table of challenges + examples + solutions
- Dataset used in a number of health studies
- A separate paper will be submitted that focuses on data extraction from source documents

Table 4. Main challenges and solutions of the current work.

Challenge	Example	Solution
The same health event is represented in several source datasets without a clear link between them, potentially leading to duplicates.	The same diagnosis code for a patient may be recorded in an EHR, claim, and prescription files. However, it may be difficult to link these documents to a single event due to the absence of a unique identifier for the case.	Transform each record as they are (even if duplicates) but add the provenance information to the record so one can use it when making cohorts.
No clear guidelines for choosing target vocabulary when multiple standard OMOP vocabularies are available. Additionally, there is no roadmap indicating which standard vocabularies may no longer be considered standard for OMOP CDM in the near future.	Physician Current Procedural Terminology Fourth Edition (CPT4) and SNOMED CT are both standard OMOP vocabularies for procedures; similarly, LOINC and SNOMED CT are for lab tests. The National Cancer Institute Thesaurus (NCIt) was a standard OMOP vocabulary at the beginning of our study, but not standard anymore.	Use the target vocabulary you are more familiar with. Keep in mind that what constitutes a standard OMOP vocabulary may change over time.
Hard to keep manual mapping files up to date as the standard target concepts change over time when updating the vocabularies.	Local code "9124," which is used for vaccination against diphtheria and tetanus, was mapped to SNOMED CT code "73152006" (administration of diphtheria and tetanus vaccine). That target concept changed from standard to nonstandard at some point in time. Thus, we had to remap it to the concept code "1657590" from RxNorm vocabulary (diphtheria toxoid vaccine, inactivated/tetanus toxoid vaccine, inactivated injection).	Whenever updating the vocabulary, recheck the mappings in Usagi before running the transformation. Usagi automatically creates the list of nonstandard mappings so one can fix them before the actual data transformatio
	atypical squamous cells of undetermined (ASC-US) result of the Papanico- corded in our datasets by 006," SNOMED codes	When working with historical codes ar always check the most recent target co this event to reuse the same code.

THANK YOU!

https://health-informatics.cs.ut.ee

ut.ee

info@ut.ee

tartuylikool
tartuuniversity

unitartu
unitartuscience
unitartutiksu