

# Mapping phenotype data of a biobank to OMOP common data model

Sulev Reisberg  
sulev.reisberg@ut.ee

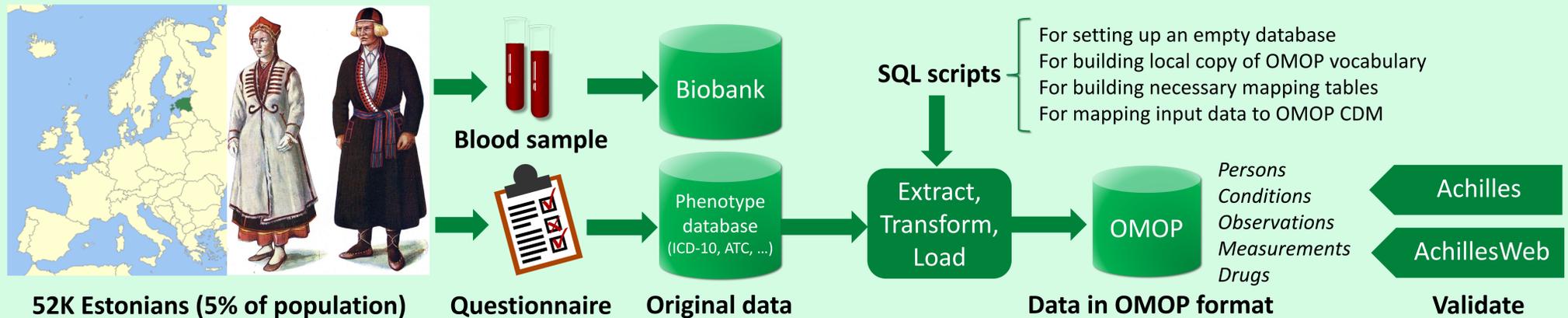
Kristjan Metsalu

Harry-Anton Talvik

Jaak Vilo

## Data (Estonian Genome Center)

## Mapping to OMOP



## Mapping challenges

a) Too many options: *Body weight findings in OMOP vocabulary:*

Which one is the "correct" OMOP concept for measured "body weight"?

Code	Name	Table
3013762	Body weight Measured	Measurement
3013853	Body weight Measured –ante partum	Measurement
3015644	Body weight Measured –postoperative	Measurement
3019336	Body weight Measured –pre dialysis	Measurement
3022281	Body weight Measured –pre pregnancy	Measurement
3023166	Body weight Stated	Measurement
3025315	Body weight	Measurement
3026600	Body weight Estimated	Measurement
3027348	Body weight special circumstances	Measurement
3027348	Body weight percentile range Categoriza...	Measurement
3042378	Body weight Set	Observation
4022831	Body weight AND/OR growth problem	Condition
4099154	Body weight	Observation

b) Missing options:

There is a concept for "red-blond hair", but not for "blond hair"

c) Country-level non-standard "ATC" codes for drugs containing several components

*TOP 5 non-mapped ATC codes*

Code	Meaning
B01AC80_1	Hjertemagnyl (acetylsalicylic acid and magnesium (hydr)oxide)
M01AB81	Diclofenac+pyridoxine+thiamine+cyanocobalamine
C01DA14	Isosorbite mononitrate
C09DA01_7	Lozap H
A12CX80_1	Panangin

d) Product names differ in different countries, no central registry

*Example of product names of Lozap H in 4 countries<sup>1</sup>*

Country	Domestic product name
Estonia	Lozap H
Hungary	TERVALON HCT 50 mg/12,5 mg filtabletta
Latvia	Lozap H 50 mg/12,5 mg apvalkotās tabletes
Lithuania	Lozap H 50/12,5 mg plėvele dengtos tabletės

e) Historical codes/code ranges

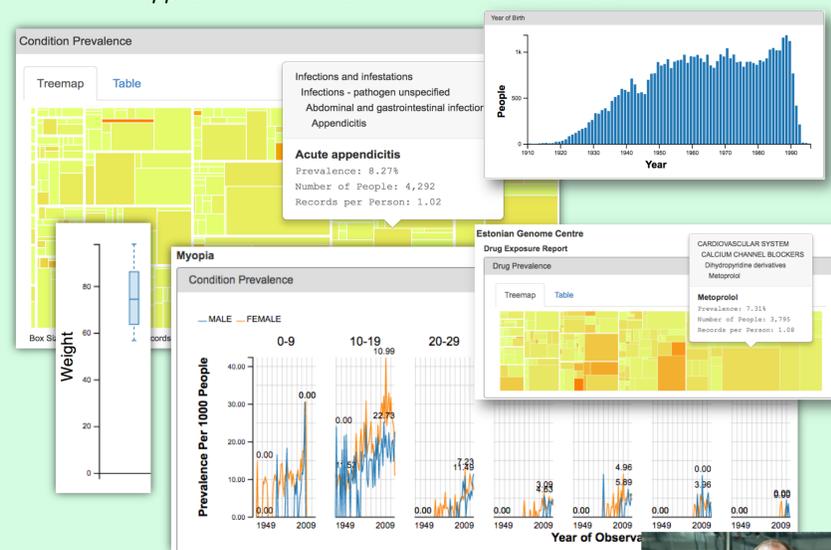
Diagnosis **F45.3** was valid in ICD10 2008 ("still valid in Estonia"), but missing from **current** ICD10 version.

When a **range** is recorded instead of exact diagnosis, e.g **I10-I15** (hypertensive diseases), no standard mapping given in OMOP vocabulary.

## Results

Table	Mapped records		Mapped codes	
	%	n	%	n
Persons	100%	51,890	-	-
Conditions	99%	402,376	93%	7,740
Observations*	100%	2,448	100%	334
Measurements (only weight and height)	100%	107,939	100%	2
Drugs	86%	56,773	63%	1,638

\* All non-mapped observations are shown under Conditions



"The visualizations are exactly what we need"

Tõnu Esko

Estonian Genome Center, University of Tartu  
Vice Director, Senior Research Fellow, PhD

## Lessons learned

- Do not leave unmapped records out from OMOP model, mark them with `CONCEPT_ID=0` instead
- Keep your mapping scripts under version control (git)
- Make the whole mapping process easily repeatable (need to run it several times)
- In a secure system one has to build mapping scripts on test data, which is challenging (it is also challenging to provide representative test data)

## Acknowledgements

Peter R. Rijnbeek, Michel Van Speybroeck, Mairo Puusepp, Tõnu Esko

## References

1. MRI Product Index, The Heads of Medicines Agencies:  
<http://mri.cts-mrp.eu/Human/Product/Details/6979>