Machine-based mapping found only 62% to 95% of infectious-disease measurements

Comparison of machine- and expert-based mapping of lab test data in a nationwide register

Background: Harmonization to OMOP CDM needs to be validated for each specific use case. Intending to cover all laboratory tests in Finland, starting from 2017, Kanta-patient data repository is a very useful data source. However, this repository was primarily developed for manual use and its effective use is difficult. The process of harmonizing its data is an ongoing process. We set out to report to what extent the laboratory test result labels were correctly mapped in a machine-based, preliminary stage of mapping. Our focus is on infection diagnostics.

Result: A first-line method mapped correctly up to 16% of the expert mapped source codes yet majority of the measurements.

Microbe	Regex found codes	Mapped codes	Same concept id	Accuracy*	Mapped measurements	Same concept_id	Accuracy*
Hepatitis	303	144	23	16.0%	332740	315690	94.9%
Influenza	362	254	16	6.3%	341858	272073	79.6%
RSV	777	87	4	4.6%	133323	82595	62.0%

Methods

$$* \ Accuracy = \frac{Same\ concept\_id}{Total\ expert\ mapped}$$

Example mappings: preliminary (blue) and expert mapping (orange)

sourcename	Frequency	conceptid	conceptname	conceptid	conceptname
-infanho[]	97131	3044938	Influenza virus A RNA [Presence] in Specimen by NAA with probe detection	3044938	Influenza virus A RNA [Presence] in Specimen by NAA with probe detection
-infbnho[]	75998	3038288	Influenza virus B RNA [Presence] in Specimen by NAA with probe detection	3038288	Influenza virus B RNA [Presence] in Specimen by NAA with probe detection
u-inf[]	51773	0	Unmapped	0	Unmapped
infanho[]	12270	3044938	Influenza virus A RNA [Presence] in Specimen by NAA with probe detection	3044938	Influenza virus A RNA [Presence] in Specimen by NAA with probe detection
ps-infaag[]	11446	3046524	Influenza virus A Ag [Presence] in Nasopharynx by Immunoassay	3046524	Influenza virus A Ag [Presence] in Nasopharynx by Immunoassay
s-infli[mg/l]	1076	3042299	inFLIXimab [Mass/volume] in Serum or Plasma	3042299	inFLIXimab [Mass/volume] in Serum or Plasma
-influenssaa,b,rs-virusjacovid-19-koronavirustauti,nukleiinihaponosoitus[]	570	0	Unmapped	36031861	Influenza virus A and B and SARS-CoV-2 (COVID-19) and Respiratory syncytial virus RNA panel - Respiratory system specimen by NAA with probe detection
-influenssaa-virus,nukleiinihappo(kval)[]	160	0	Unmapped	3044938	Influenza virus A RNA [Presence] in Specimen by NAA with probe detection

Mapping characteristics

	Preliminary mapping	Expert mapping
Focus	Any lab measurements	Hepatitis, influenza and RSV
Coverage goals	Most frequent	All
Method	Usagi-based process + human checking	Human checking

Limitations:

1. Accuracy is different for each use case and we only show one example.
2. In the future, source codes are subject to changes, even towards a less harmonized direction.

OHDSI

Petteri Hovi, Eero Poukka, Ida-Liisa Kolari, Javier Garcia Tabuenca, FinnGen, Gustav Klingstedt. *Institute for Health and Welfare, Helsinki, Finland*

thl