

# Evaluation of vaccine concept mappings in the OMOP vocabulary: a real-world database study

PRESENTER: **Adam Black**  
**Denys Kaduk**

## INTRO

The OHDSI community and OMOP common data model support robust observational studies across multiple datasets and institutions. However, the quality issues of vaccine related concepts in the OMOP vocabulary pose a significant barrier to efficient and high-quality studies. Following our previous quality assessment of vaccine vocabularies and concepts related to influenza, pneumococcal disease, and shingles<sup>1</sup>, we expanded the evaluation to all vaccine types, but with a focus on the “Maps to” relationship between source concept and standard vaccine concepts.

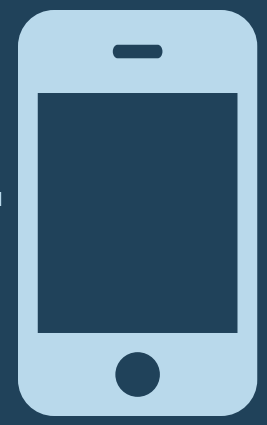
## METHODS

- Iterative regular expressions-based pattern matching and manual review were used to identify a starting set of vaccine concepts.
- The starting set of vaccines concepts was then expanded using concept relationships and hierarchies in the OMOP vocabulary. The expanded list of vaccine concepts was manually reviewed for accuracy.
- Once the comprehensive list of vaccine concepts was complete, all "Maps to" relationships involving vaccines were extracted from the *concept\_relationship* table.
- Occurrences of each source\_concept\_id - standard\_concept\_id pair were counted in the *drug\_exposure* and *procedure\_occurrence* tables in five OMOP CDM databases
- "Maps to" relationships that occurred in at least one dataset were manually reviewed for accuracy by a clinical expert and mapping errors were grouped into four categories

## RESULTS

We found 15,932 vaccine-related concepts in 32 vocabularies (Table 1). From these concepts we extracted 15,220 “Maps to” relationships and reviewed 1,170 source\_concept\_id – standard\_concept\_id pairs with >=1 occurrence in any of the 5 CDM datasets we have access to. The clinical expert on our team identified potential problems with 104 mappings (8.89% of the mappings reviewed), as summarized in Table 2.

# A manual review of vaccine mappings in the OMOP CDM identified several errors which can lead to incorrect results in research.



Take a picture to download the [full paper](#)

**Table 1. Vaccine concept in usage in 5 OMOP datasets**

Vocabulary	Number of vaccine concepts	Number and percent of vaccine concepts used in real world datasets
NDC	1427	1000 (70.1%)
RxNorm	2920	420 (14.4%)
Read	257	257 (100%)
SNOMED	474	213 (44.9%)
Gemscript	289	148 (51.2%)
RxNorm Extension	8229	130 (1.6%)
CVX	161	120 (74.5%)
CPT4	136	118 (86.8%)
HCPCS	22	22 (100%)
ICD9Proc	6	6 (100%)
ICD10PCS	2	2 (100%)
AMT	118	0 (0%)
ATC	66	0 (0%)
BDDPM	128	0 (0%)
CIEL	61	0 (0%)
CTD	10	0 (0%)
dm+d	356	0 (0%)
DPD	130	0 (0%)
GCN_SEQNO	220	0 (0%)
GGR	171	0 (0%)
HemOnc	3	0 (0%)
ICD9ProcCN	4	0 (0%)
JMDC	26	0 (0%)
KDC	3	0 (0%)
MeSH	32	0 (0%)
Multum	53	0 (0%)
NCCD	8	0 (0%)
NDFRT	30	0 (0%)
Nebraska Lexicon	49	0 (0%)
OPCS4	1	0 (0%)
SPL	261	0 (0%)
VA Product	176	0 (0%)

**Table 2. Mapping Issues by category**

Mapping issue category	Definition and Example	# and % with mapping issues
<b>Lack of complete mapping</b>	A vaccine mapping that did not capture all components or ingredients of the vaccine. <i>[45488921] "Third low dose diphtheria, tetanus and inactivated polio vaccination" maps to [529411] tetanus and [529303] diphtheria but not polio.</i>	68 (65.4%)
<b>Incorrect mapping</b>	A vaccine mapping where the standard concept is not synonymous with the source concept. <i>[21601291] "hemophilus influenzae B, purified antigen conjugated; systemic" maps to [515671] "Neisseria meningitidis"</i>	21 (20.1%)
<b>Imprecise mapping</b>	A vaccine mapping where important information is either removed or added. <i>[2213439] "Influenza virus vaccine, trivalent (IIV3), split virus, 0.25 mL dosage, for intramuscular use" maps to [40213153] "Influenza, seasonal, injectable" which drops information about dosage, route, and valence.</i>	6 (5.8%)
<b>Questionable mapping</b>	A vaccine mapping that is not necessarily incorrect but should be reviewed by the vocabulary team <i>[2213449] "Rabies vaccine, for intramuscular use" maps to CVX concept [40213209] "rabies vaccine, unspecified formulation". CVX concept [40213208] "rabies, intramuscular injection" is a standard concept that would be a better fit but has been retired by CDC.</i>	9 (8.7%)

Denys Kaduk, Adam Black, Yupeng Li, Lixia Yao

