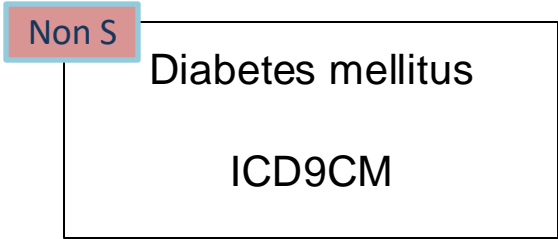
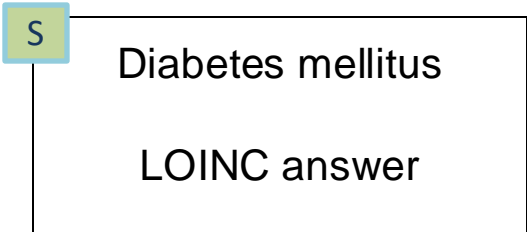
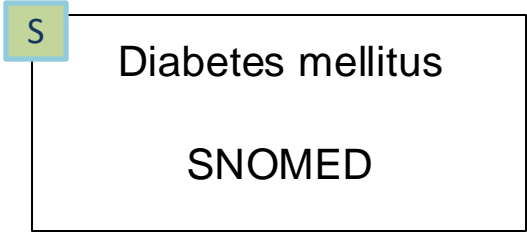
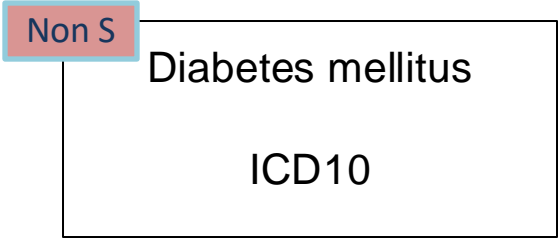


EAV Data handling in OMOP CDM

Observation / Measurement tables

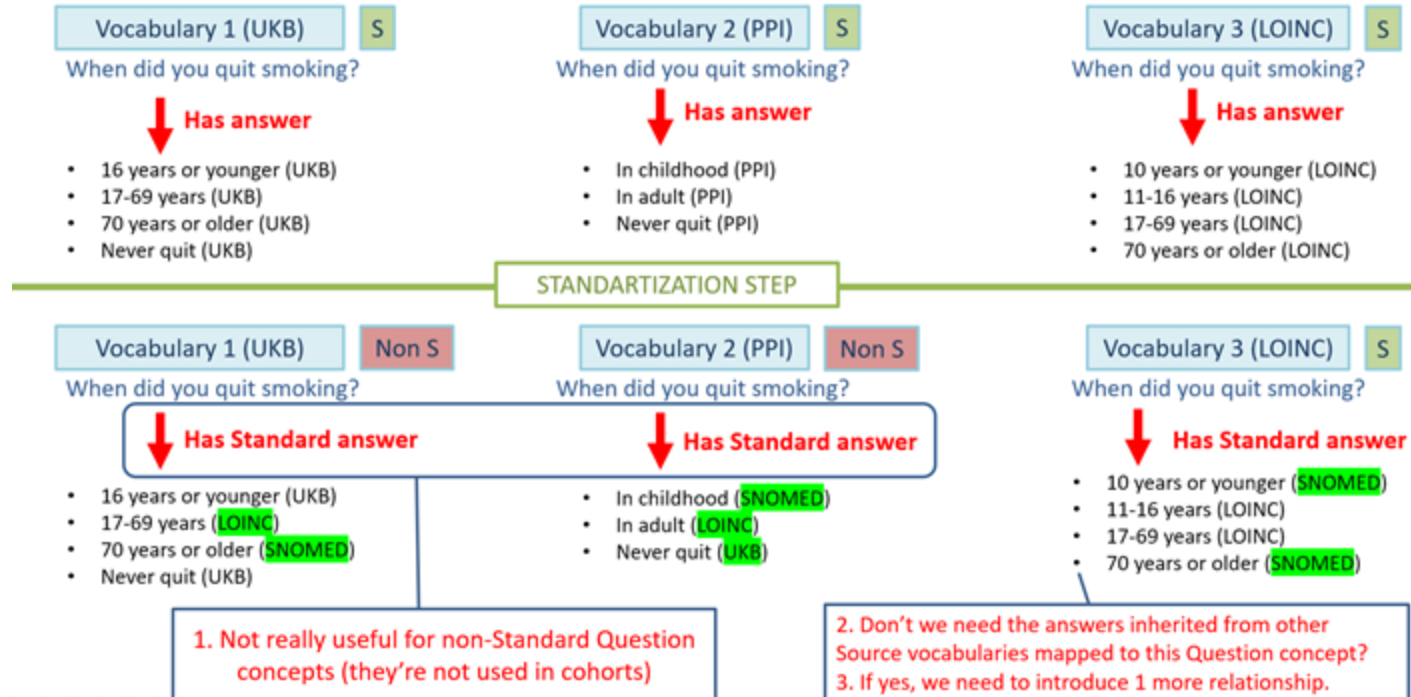
What's the problem?

- Survey vocabularies are very heterogeneous in structure
- They bring a lot of Standard concepts, that are often misused by OMOP users:
 - LOINC ~11,000 concepts
 - PPI more than 2,000 concepts
 - UK Biobank ~4,000 concepts
- They violate vocabulary principles:
 - **Unique Standard Concepts** => For each Clinical Entity there is only one concept representing it.
 - **No Flavors of NULL** => No standard concept should indicate flavors of null (unknown, no reported).
 - **No Negative Information** => No concept should represent absence of evidence.
 - **No Timing** => temporality should be enforced from the data and not rely on the Vocabularies.
(eg. *Have you had AMI during the last 10 years? - Yes*)



Mapping attempt between the Surveys

What's the borderline of the concept semantics?



How to handle EAV variable/value pairs in MEASUREMENT or OBSERVATION - Call for input from the community

Vocabulary Users



Christian_Reich

1 Apr '21

Apr 2021

Friends:

The data in most of the OMOP CDM tables are statements of fact. But we have those EAV tables in which the information is organized in variable/value or question/answer pairs. The variable/question is always a concept, and the value/answer may or may not be a concept. I am talking about the ones where they are, and there are several sources of these:

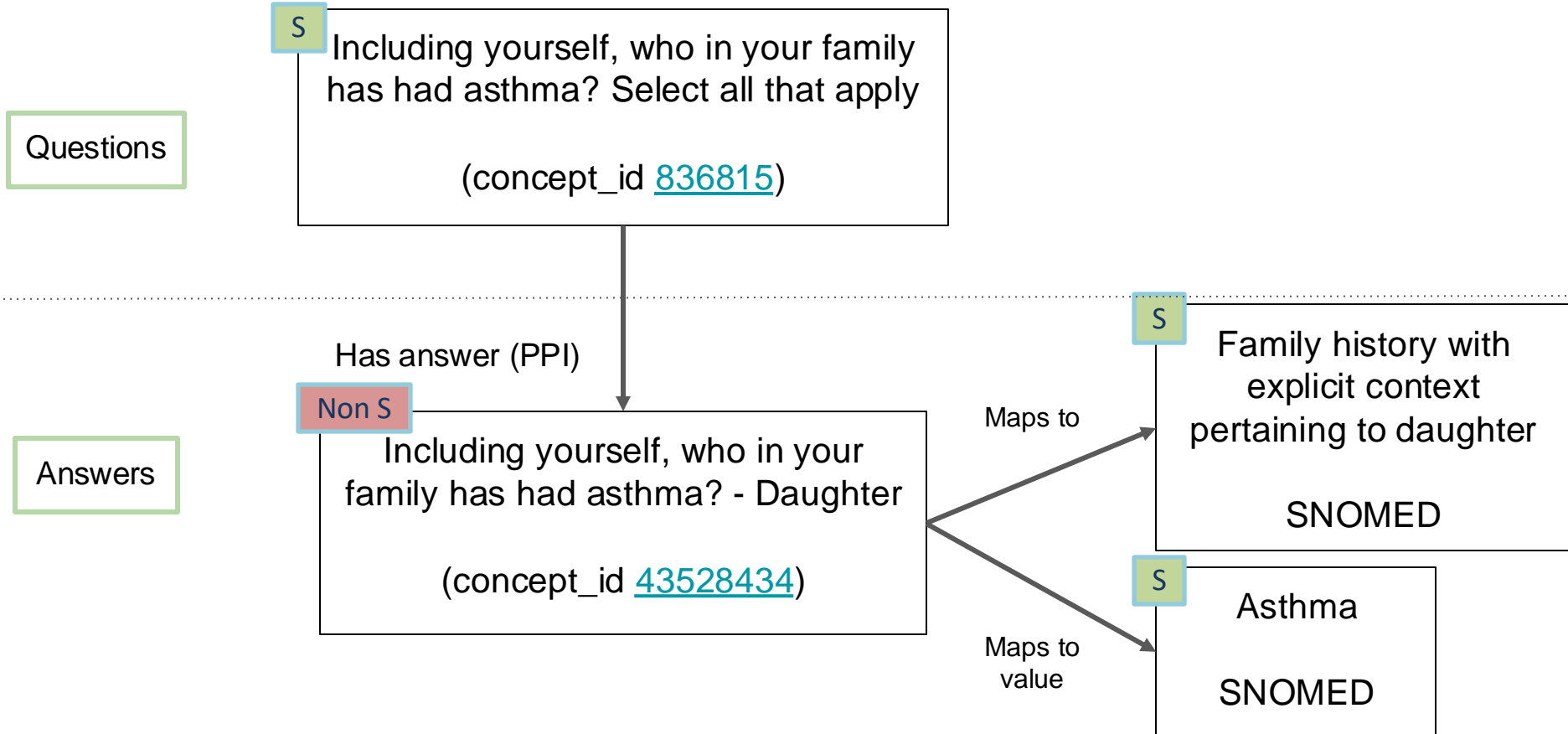
1 / 6

Apr 2021

So, what should we do? How should we handle these?

1. We make all pairs non-standard. Then you can use them if you know they are there, but otherwise we don't claim to be able to effectively utilize them.
2. We make them all standard (Observation or Measurement domain). Then we will get a **lot** of standard concepts of varying benefit, cluttering the tables.
3. We make them non-standard and map them to single standard concepts what they really are.
4. We make them non-standard and create new single pre-coordinated standard concepts combining the information. We are doing this in the Oncology WG.
5. We make the variable or question non-standard, and alter the concept_name of the values/answers in such a way that they contain a concatenated concept_name of variables/questions with the values/answers. The AllOfUs folks did this for their PPI concepts.

Current approaches to content mapping



Current approaches to content mapping

Questions

In your day-to-day life, how often did this happen to you during the past month? People act as if they are afraid of you.

(concept_id [1333314](#))

Answers

S

A few times a month

(concept_id [1332783](#))

Non S

Almost everyday

(concept_id [1332918](#))

Has answer (PPI)

S

Almost every day

LOINC

Maps to

A-to-A or A(V)-to-E mappings. Examples

cope_a_332

1

Maps to

LA6306-0

One

PPI Answer
[Meas Value]

LOINC Answer
[Meas Value]

placenta @2850 @37

(20, 30, or 33) + (10)||Metastasis to regional lymph node(s) plus lung

Maps to

OMOP4999962

Metastasis to lung

Metastasis
[Measurement]

NAACCR Value
[Meas Value]

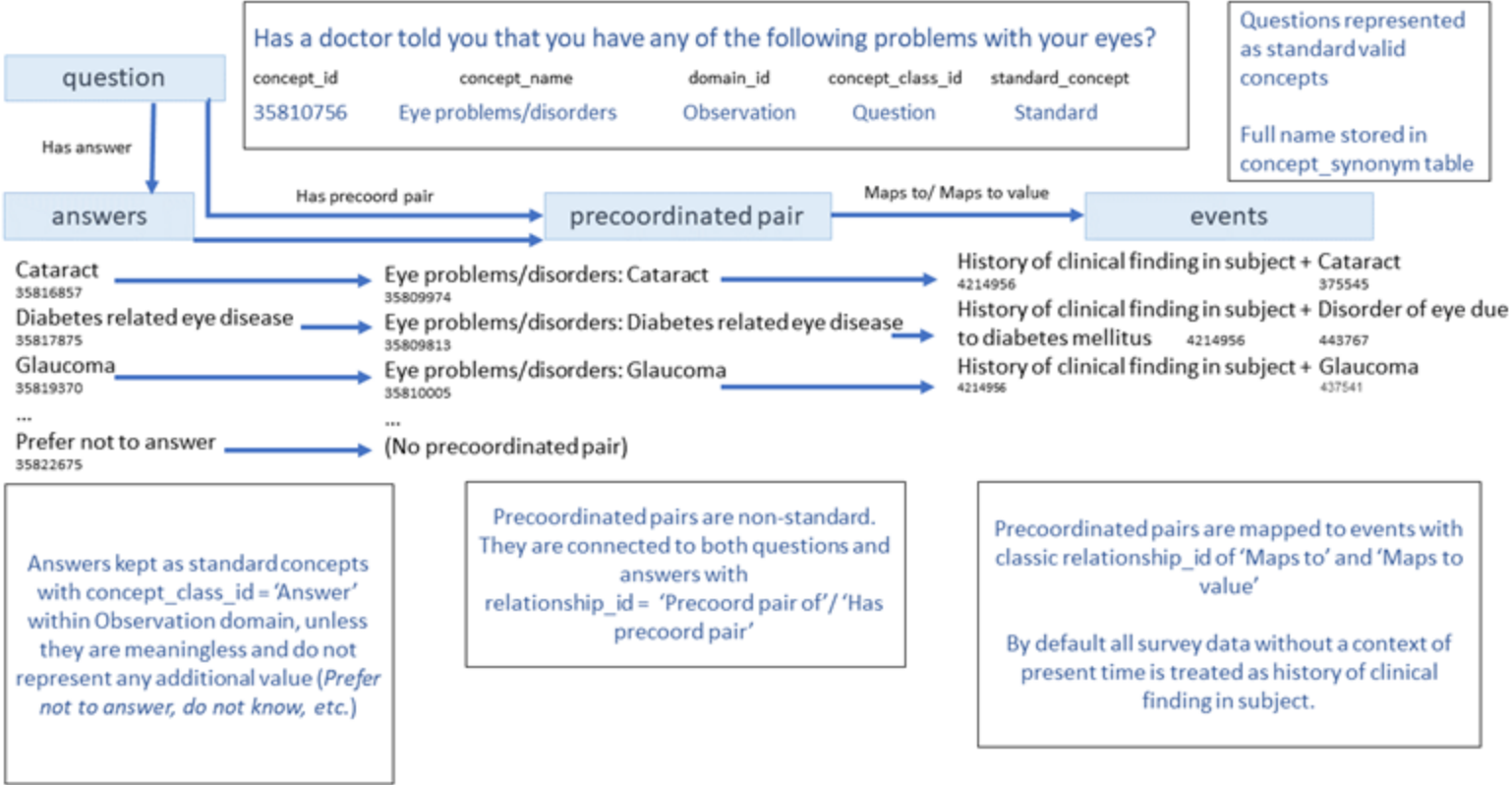
Maps to

OMOP4998946

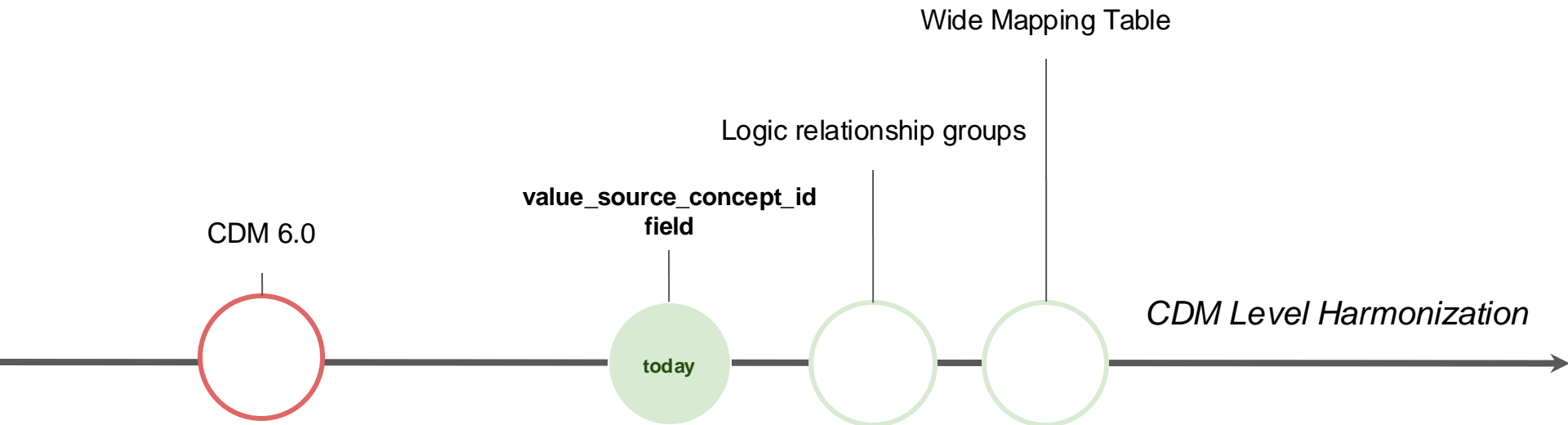
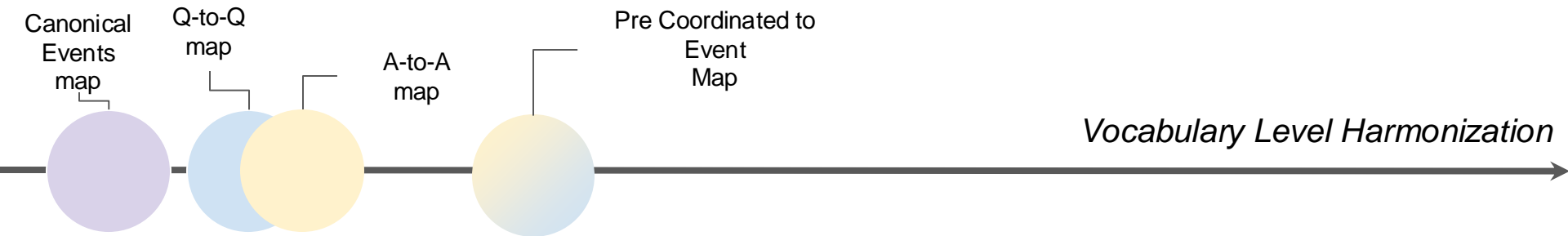
Regional spread to lymph node

Nodes
[Measurement]

Introducing pre-coordinated pairs in UK Biobank



Post-coordination harmonization attempts



Proposal (the same applies to the Measurement table)

| Field in the Observation table | How they should be populated |
|--------------------------------|---|
| observation_source_value | Including yourself, who in your family has had asthma? Select all that apply. |
| value_source_value | Daughter |
| observation_source_concept_id | 836815 Including yourself, who in your family has had asthma? Select all that apply. |
| value_source_concept_id | 43528434 Daughter |
| observation_concept_id | 4054433 Family history with explicit context pertaining to daughter |
| value_as_concept_id | 317009 Asthma |

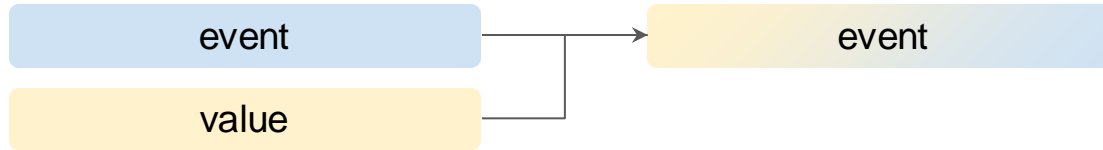
Missing field



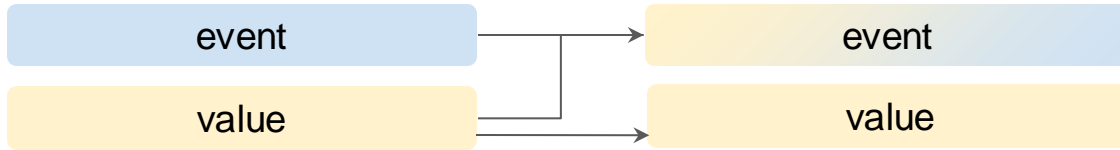
CDM fields definition: violation or not?



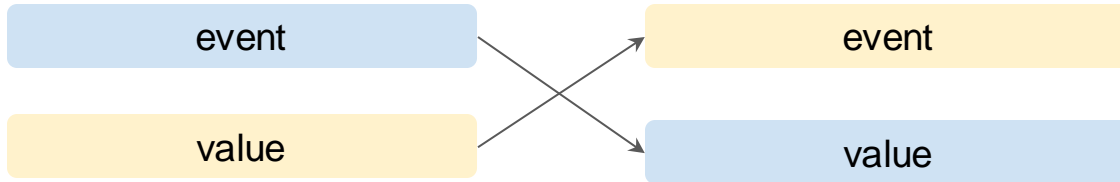
One-to-one maps
STRUCTURE fits SEMANTIC



Many-to-one maps
STRUCTURE ~ SEMANTIC



many-to-many maps
STRUCTURE <> SEMANTIC



swap maps
STRUCTURE <> SEMANTIC

Summary

- We add one field (`value_source_concept_id`) into two tables (Observation, Measurement)
- We align the definition of the CDM fields with the mapping practice and expectations
- In the future, the survey data can be modeled through the non-Standard concepts:
 - vocabulary source structure is preserved as is (Q/A are available)
 - both source questions and answers are accessible in CDM
 - mappings are done through the pre-coordinate pairs where applicable (real clinical facts)
 - analytics is based on either real clinical facts or source non-Standard survey data

References

- [Survey vocabularies in OMOP \(forum\)](#)
- [How to handle EAV variable/value pairs \(forum\)](#)
- [Who is working with UK Biobank? \(forum\)](#)
- [UKB Survey data in the OMOP Vocabulary \(presentation\)](#)
- [UK Biobank vocabulary release \(forum\)](#)
- [UK Biobank vocabulary \(wiki\)](#)
- [Wide MAPPING table \(forum\)](#)
- [Wide mapping table \(poster\)](#)