The Book of OHDSI Korea

OHDSI-Korea

2019-06-06

Contents

Chapter 1

Introduction

. , , . , , .

, (Common Data Model, CDM) (Distributed Research Network, DRN) . CDM / DRN , (OMOP-CDM) (OHDSI network) .

OHDSI (open-source research network) , .

, OHDSI .

Chapter 2

The OHDSI Community

Seng Chan You

2.1 OHDSI

OHDSI . OHDSI .

2.2

2.2.1 (Distributed Research Network)

, , , . , , .

. .

, / (, , , ,) . , . .

2.2.2 (Common Data Model)

, . .

. .

.

. ETL(Extract, Transform, Load) .

.

(Observational Health Data and Informatics, OHDSI) FDA (Sentinel CDM), (The National Patient-Centered Clinical Outcomes Research Network, PCORnet) .

OHDSI . OHDSI 2008 Observational Medical Outcomes Partnership(OMOP) . , , , .

OMOP 2013 , OMOP CDM OHDSI . OMOP , OHDSI , , , , .

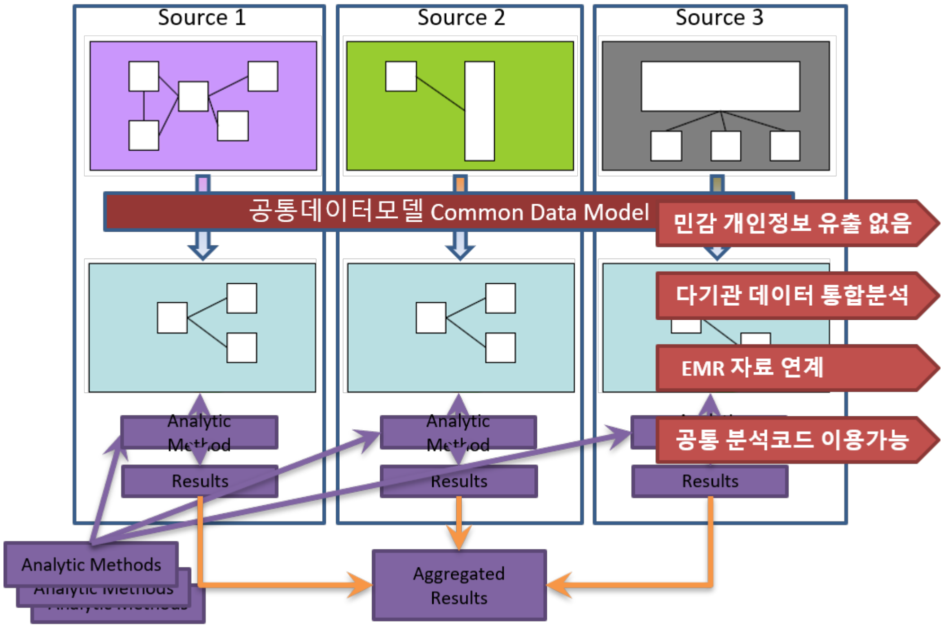


Figure 2.1: Distributed Research Network

CDM Sentinel Initiative (Food and Drug Administration, FDA) . Sentinel . FDA FDA .

Sentinel . Sentinel CDM . Sentinel CDM . . Sentinel Distributed Database(SDD) .

CDM PCORnet The Patient-Centered Outcomes Research Institute(PCORI) 2013 (Electronic health records, EHR) (Comparative effectiveness research, CER) . 50 11 (Clinical data research networks, CDRNs) 18 (Patient-powered research networks, PPRNs) . PCORnet (patient-centered approach) .

2.3 (OHDSI network)

(The Observational Health Data Sciences and Informatics, OHDSI network) OMOP (Observational Medical Outcomes Partnership) . (CDM) (Distributed Research Network) , OHDSI . OHDSI OMOP-CDM , , , .

2.4 OHDSI

2008 (FDA) , , OMOP (Observational Medical Outcomes Partnership) ref. 2009 OMOP-CDM version 1 ref. OMOP common data model (CDM) (claim data) (eletronic heatlh record) , 2013 Reagan-Udall . FDA OMOP , Columbia (coordinating center), George Hripsack (OHDSI) . ODYSSEY , OHDSI . 2014 Columbia Face-to-Face (F2F meeting) 2015 (Washington DC) . (Bethesda) . OHDSI .

Figure 2.2: OHDSI International Symposium 2017 in Korea

Figure 2.3: OHDSI International Symposium 2017 in Korea

2.4.1

2014 OMOP-CDM , 2015 OHDSI . OMOP-CDM, OHDSI , 2016 OHDSI committee Korean chapter , OHDSI . 2017 3 , 3 .

OHDSI 2017 3 7 2 / OHDSI OHDSI .

2.5 , ,

OHDSI mission, vision, value page .

2.5.1 OHDSI

.

> To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care.

2.5.2 OHDSI

.

Figure 2.4: Tutorial in the OHDSI International Symposium 2017

A world in which observational research produces a comprehensive understanding of health and disease.

2.5.3 OHDSI

- **Innovation**: .

  Observational research is a field which will benefit greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.

- **Reproducibility**: , , .

  Accurate, reproducible, and well-calibrated evidence is necessary for health improvement.

- **Community**: .

  Everyone is welcome to actively participate in OHDSI, whether you are a patient, a health professional, a researcher, or someone who simply believes in our cause.

- **Openness**: , , , .

  We strive to make all our community's proceeds open and publicly accessible, including the methods, tools and the evidence that we generate.

- **Collaboration**: , .

  We work collectively to prioritize and address the real world needs of our community's participants.

- **Beneficence**: .

  We seek to protect the rights of individuals and organizations within our community at all times.

OHDSI , ‘ , ’ .

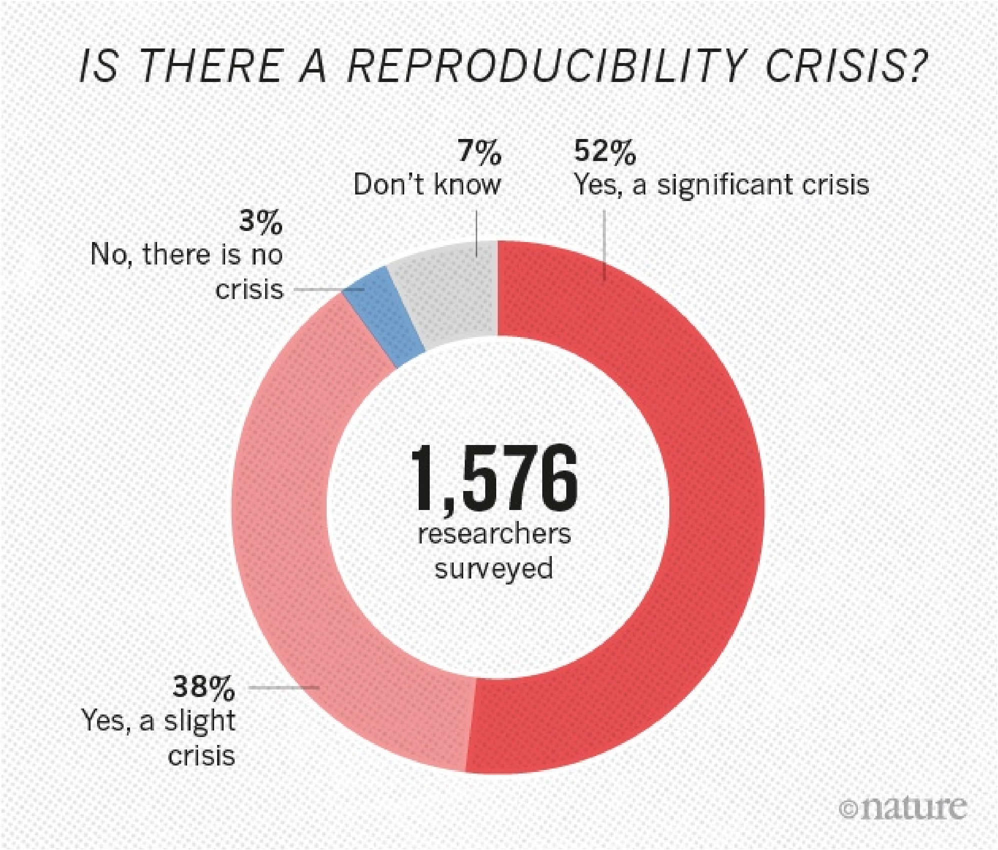


Figure 2.5: ?

2.6 (Open Science and Reproducible Research)

OHDSI CDM DRN , .

2.6.1 (Reproducible Crisis)

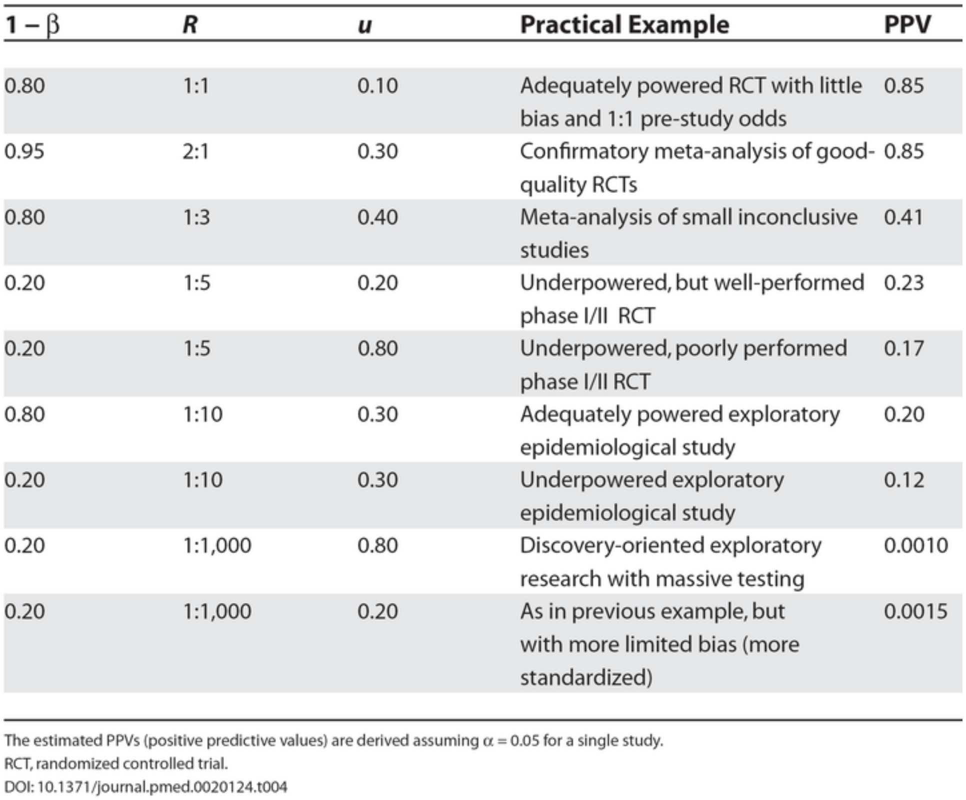
2016 Nature 1576 , 70% , 50% . 52% . ref

(Reproducible research) ’ (raw data) ’ .

, . PLOS Medicine 1/10 , (Discovery-oriented exploratory research with massive testing) 1000 1 [ref, *Ioannidis, 2005 PLOS Medicine*].

? 4 [ref, *Bishop, 2019, Nature*].

- Publication Bias
- Low Statistical Power
- *P* *P*-value Hacking
- HARKing (Hypothesizing After Results are Known)



1 – β	*R*	*u*	Practical Example	PPV
0.80	1:1	0.10	Adequately powered RCT with little bias and 1:1 pre-study odds	0.85
0.95	2:1	0.30	Confirmatory meta-analysis of good-quality RCTs	0.85
0.80	1:3	0.40	Meta-analysis of small inconclusive studies	0.41
0.20	1:5	0.20	Underpowered, but well-performed phase I/II RCT	0.23
0.20	1:5	0.80	Underpowered, poorly performed phase I/II RCT	0.17
0.80	1:10	0.30	Adequately powered exploratory epidemiological study	0.20
0.20	1:10	0.30	Underpowered exploratory epidemiological study	0.12
0.20	1:1,000	0.80	Discovery-oriented exploratory research with massive testing	0.0010
0.20	1:1,000	0.20	As in previous example, but with more limited bias (more standardized)	0.0015

The estimated PPVs (positive predictive values) are derived assuming $\alpha = 0.05$ for a single study.
RCT, randomized controlled trial.
DOI: 10.1371/journal.pmed.0020124.t004

Figure 2.6: PPV of Research Findings for various combinations of power, ratio of Tru to Not-True Relationship, and Bias

2.6.1.1 Publication Bias

. . ’ *(file drawer effect)* . , . [ref, , . 165]

p-value hacking HARKing . impact factor (IF) , IF .

2.6.1.2 Low Statistical Power

(small sample size) (small effect), . (underpowered) , [ref, R. G. Newcombe Br. Med. J. (Clin. Res. Ed.) 295, 656–659; 1987] . Low Statistical Power , *p*-value hacking HARKing . IF .

2.6.1.3 *P* *P*-value Hacking and HARKing (Hypothesizing After Results are Known)

P-value hacking HARKing , ‘Discovery-oriented exploratory research with massive testing’ . IF , *P*-value ($P < 0.05$) , . ’False-Positive Psychology’ Simmons 4 (, , ,) ‘ ’ [ref, Simmons et al., 2011 Psychological Science]. Hack Your Ways To Scientific Glory *P*-hacking , testing 0.05 *P* .

2.6.2

Simmons , 6 (disclosure based solution) . , , , .

```
1. Authors must decide the rules for the terminating data collection before data collectino begins and r
2. Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data coll
3. Authors must list all variables collected in a study
4. Authors must report all experimental conditions, including failed manipulations
5. If observation are elimiated, authors must also report what the statistical results are if those obse
6. If an analysis includes a covariate, authors must report the statistical results of the analysis with
```

OHDSI ‘ ’ . . OHDSI . OHDSI Studies GitHub OHDSI Study Protocol Github .

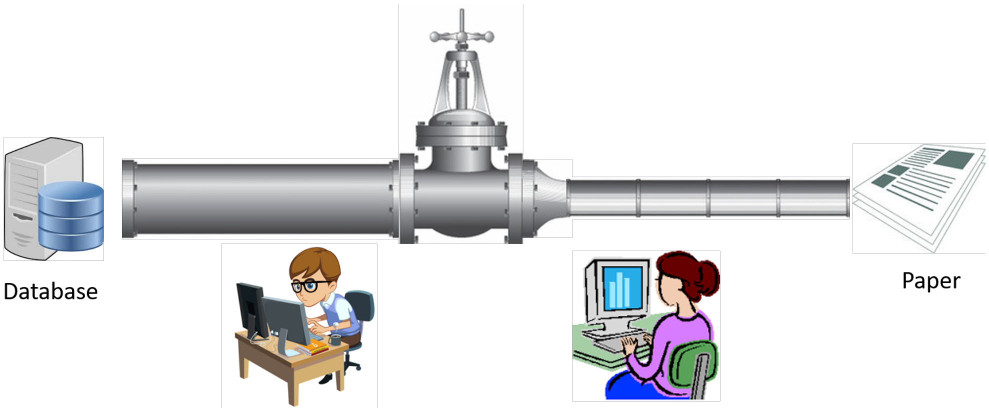


Figure 2.7: An OHDSI study shoul be look like a pipeline

Chapter 3

The OMOP-CDM

- OMOP-CDM : CDM wiki page OMOP-CDM .
- OMOP-CDM Github : OMOP-CDM github
- OMOP-CDM : OMOP-CDM OMOP-CDM .
- OHDSI tutorial : OHDSI past event OMOP-CDM tutorial .

3.1

OMOP-CDM . OMOP-CDM .

- **Suitability for purpose:** The CDM aims to provide data organized in a way optimal for analysis, rather than for the purpose of addressing the operational needs of health care providers or payers.
- **Data protection:** All data that might jeopardize the identity and protection of patients, such as names, precise birthdays etc. are limited. Exceptions are possible where the research expressly requires more detailed information, such as precise birth dates for the study of infants.
- **Design of domains:** The domains are modeled in a person-centric relational data model, where for each record the identity of the person and a date is captured as a minimum.
- **Rationale for domains:** Domains are identified and separately defined in an entity-relationship model if they have an analysis use case and the domain has specific attributes that are not otherwise applicable. All other data can be preserved as an observation in an entity-attribute-value structure.
- **Standardized Vocabularies:** To standardize the content of those records, the CDM relies on the Standardized Vocabularies containing all necessary and appropriate corresponding standard healthcare concepts.
- **Reuse of existing vocabularies:** If possible, these concepts are leveraged from national or industry standardization or vocabulary definition organizations or initiatives, such as the National Library of Medicine, the Department of Veterans' Affairs, the Center of Disease Control and Prevention, etc.
- **Maintaining source codes:** Even though all codes are mapped to the Standardized Vocabularies, the model also stores the original source code to ensure no information is lost.
- **Technology neutrality:** The CDM does not require a specific technology. It can be realized in any relational database, such as Oracle, SQL Server etc., or as SAS analytical datasets.
- **Scalability:** The CDM is optimized for data processing and computational analysis to accommodate data sources that vary in size, including databases with up to hundreds of millions of persons and billions of clinical observations.
- **Backwards compatibility:** All changes from previous CDMs are clearly delineated in the github repository (https://github.com/OHDSI/CommonDataModel). Older versions of the CDM can be easily created from the CDMv5, and no information is lost that was present previously.

3.2 Data Model Conventions

There are a number of implicit and explicit conventions that have been adopted in the CDM. Developers of methods that run against the CDM need to understand these conventions.

3.2.1 General conventions of the model

The OMOP CDM is considered a "person-centric" model, meaning that the people (or patients) drive the event and observation tables. At a minimum, the tables have a foreign key into the PERSON table and a date. This allows for a longitudinal view on all healthcare-relevant events by person. The exceptions from this rule are the standardized health system data tables, which are linked directly to events of the various domains.

3.2.2 General conventions of schemas

New to CDM v6.0 is the concept of schemas. This allows for more separation between read-only and writeable tables. The clinical data, event, and vocabulary tables are in the 'CDM' schema and are considered read-only to the end user. This means that the tables can be queried but no information can be accidentally removed or written over except by the database administrator. Tables that need to be manipulated by web-based tools or end users have moved to the 'Results' schema. Currently the only two tables in the 'Results' schema are COHORT and COHORT_DEFINITON, **add a sentence explaining that these tables describe groups of interest that the user might define, put in links to the later sections** though likely more will be added over the course of v6.0 point releases. These tables can be written to, meaning that a cohort created in ATLAS or by a user can be stored in the COHORT table and accessed at a later date. This does mean that cohorts in the COHORT table can be manipulated by anyone so it is always recommended that the SQL code used to create the cohort be saved along with the project or analysis in the event it needs to be regenerated.

3.2.3 General conventions of data tables

The CDM is platform-independent. Data types are defined generically using ANSI SQL data types (VARCHAR, INTEGER, FLOAT, DATE, DATETIME, CLOB). Precision is provided only for VARCHAR. It reflects the minimal required string length and can be expanded within a CDM instantiation. The CDM does not prescribe the date and datetime format. Standard queries against CDM may vary for local instantiations and date/datetime configurations.

In most cases, the first field in each table ends in '_ID', containing a record identifier that can be used as a foreign key in another table. For example, the CONDITION_OCCURRENCE table contains the field VISIT_OCCURRENCE_ID which is a foreign key to the VISIT_OCCURRENCE table where VISIT_OCCURRENCE_ID is the primary key.

3.2.4 General conventions of fields

Variable names across all tables follow one convention:

Notation	Description
_SOURCE_VALUE	Verbatim information from the source data, typically used in ETL to map to CONCEPT_ID, and not to be used by any standard analytics. For example, CONDITION_SOURCE_VALUE = '787.02' was the ICD-9 code captured as a diagnosis from the administrative claim.

Notation	Description
_ID	Unique identifiers for key entities, which can serve as foreign keys to establish relationships across entities. For example, PERSON_ID uniquely identifies each individual. VISIT_OCCURRENCE_ID uniquely identifies a PERSON encounter at a point of care.
_CONCEPT_ID	Foreign key into the Standardized Vocabularies (i.e. the standard_concept attribute for the corresponding term is true), which serves as the primary basis for all standardized analytics. For example, CONDITION_CONCEPT_ID = 31967 (http://athena.ohdsi.org/search-terms/terms/31967) contains the reference value for the SNOMED concept of 'Nausea'
_SOURCE_CONCEPT_ID	Foreign key into the Standardized Vocabularies representing the concept and terminology used in the source data, when applicable. For example, CONDITION_SOURCE_CONCEPT_ID = 45431665 (http://athena.ohdsi.org/search-terms/terms/45431665) denotes the concept of 'Nausea' in the Read terminology; the analogous CONDITION_CONCEPT_ID might be 31967, since SNOMED-CT is the Standardized Vocabulary for most clinical diagnoses and findings.
_TYPE_CONCEPT_ID	Delineates the origin of the source information, standardized within the Standardized Vocabularies. For example, DRUG_TYPE_CONCEPT_ID can allow analysts to discriminate between 'Pharmacy dispensing' and 'Prescription written'

3.2.5 Representation of content through Concepts

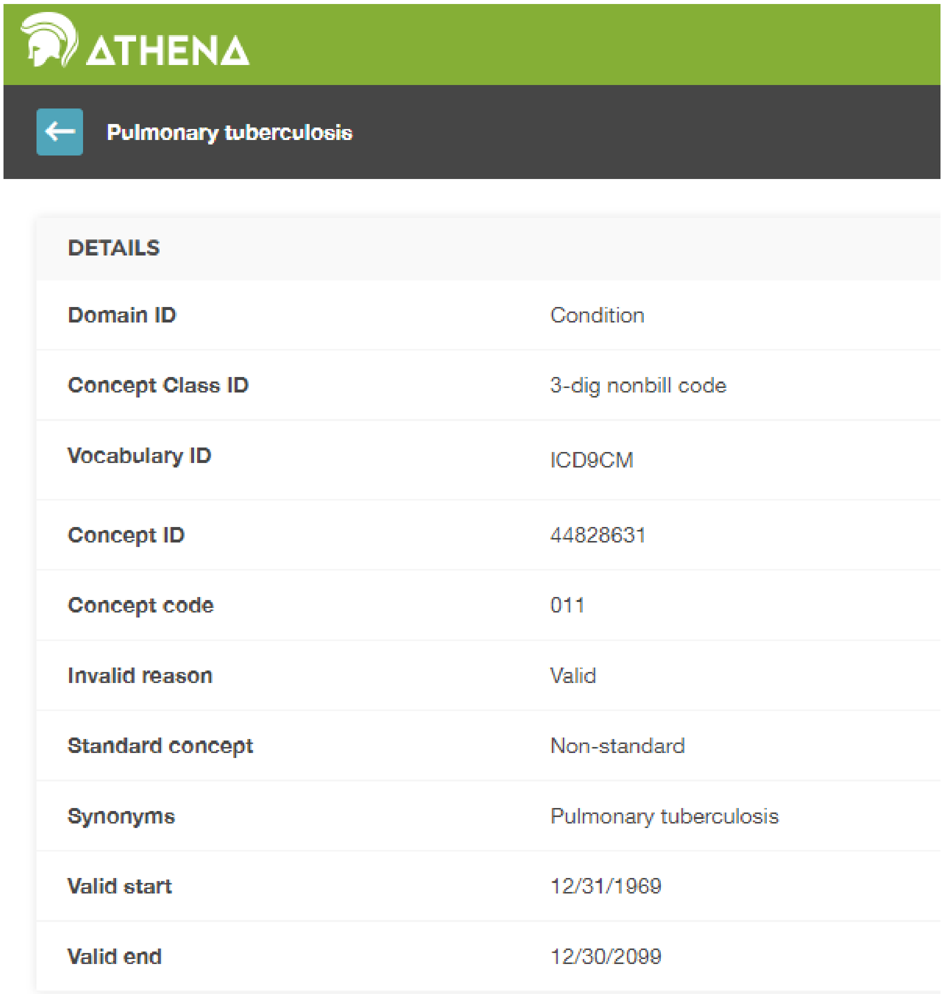
In CDM data tables the content of each record is represented using Concepts. Concepts are stored in event tables with their CONCEPT_IDs as foreign keys to the CONCEPT table, which contains Concepts necessary to describe the healthcare experience of a patient. If a Standard Concept does not exist or cannot be identified, the the CONCEPT_ID 0 is used, representing a non-existing concept or un-mappable source value.

Records in the CONCEPT table contain detailed information about each concept (name, domain, class etc.). Concepts, Concept Relationships, Concept Ancestors and other information relating to Concepts is contained in the tables of the Standardized Vocabularies.

3.2.6 Difference between Concept IDs and Source Values

Many tables contain equivalent information in multiple places: As a Source Value, a Source Concept and as a Standard Concept.

- Source Values contain the codes from public code systems such as ICD-9-CM, NDC, CPT-4, READ etc. or locally controlled vocabularies (such as F for female and M for male) copied from the source data. Source Values are stored in the _SOURCE_VALUE fields in the data tables.
- Concepts are CDM-specific entities that represent the meaning of a clinical fact. Most concepts are based on code systems used in healthcare (called Source Concepts), while others were created de-novo (CONCEPT_CODE = 'OMOP generated'). Concepts have unique IDs across all domains.
- Source Concepts are the concepts that represent the code used in the source. Source Concepts are only used for common healthcare code systems, not for OMOP-generated Concepts. Source Concepts are stored in the _SOURCE_CONCEPT_ID field in the data tables.
- Standard Concepts are those concepts that are used to define the unique meaning of a clinical entity. For each entity there is one Standard Concept. Standard Concepts are typically drawn from existing public vocabulary sources. Concepts that have the equivalent meaning to a Standard Concept are



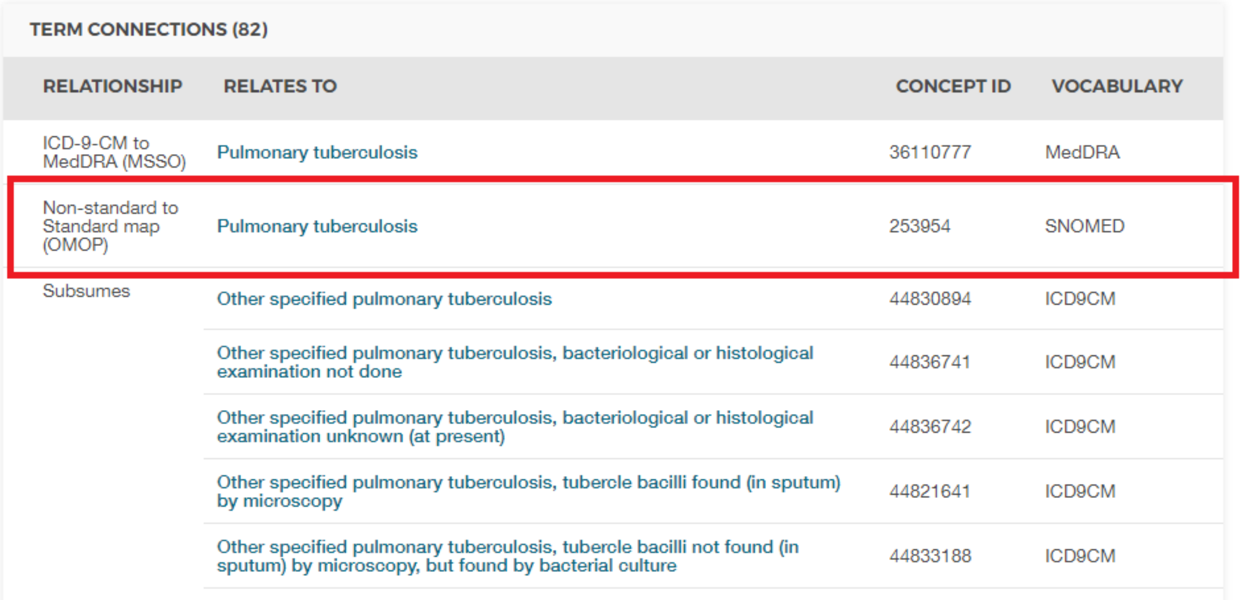
DETAILS	
Domain ID	Condition
Concept Class ID	3-dig nonbill code
Vocabulary ID	ICD9CM
Concept ID	44828631
Concept code	011
Invalid reason	Valid
Standard concept	Non-standard
Synonyms	Pulmonary tuberculosis
Valid start	12/31/1969
Valid end	12/30/2099

Figure 3.1: ICD9CM code for Pulmonary Tuberculosis

mapped to the Standard Concept. Standard Concepts are referred to in the _CONCEPT_ID field of the data tables.

Source Values are only provided for convenience and quality assurance (QA) purposes. Source Values and Source Concepts are optional, while **Standard Concepts are mandatory**. Source Values may contain information that is only meaningful in the context of a specific data source. This mandatory use of Standard Concepts is what allows all OHDSI collaborators to speak the same language. For example, let's look at the condition 'Pulmonary Tuberculosis' (TB). Figure 3.1 shows that the ICD9CM code for TB is 011.

Without the use of a standard way to represent TB the code 011 could be interpreted as 'Hospital Inpatient (Including Medicare Part A)' in the UB04 vocabulary, or as 'Nervous System Neoplasms without Complications, Comorbidities' in the DRG vocabulary. This is where Concept IDs, both Source and Standard, are valuable. The Concept ID that represents the 011 ICD9CM code is 44828631 (http://athena.ohdsi.org/search-terms/terms/44828631). This differentiates the ICD9CM from the UBO4 and from the DRG. The Standard Concept that ICD9CM code maps to is 253954 (http://athena.ohdsi.org/search-terms/terms/253954) as shown in figure 3.2 by the relationship 'Non-standard to Standard map (OMOP)'. This same mapping relationship exists between Read, ICD10, CIEL, and MeSH codes, among others, so that any research that references the standard SNOMED concept is sure to include all supported source codes.



TERM CONNECTIONS (82)

RELATIONSHIP	RELATES TO	CONCEPT ID	VOCABULARY
ICD-9-CM to MedDRA (MSSO)	Pulmonary tuberculosis	36110777	MedDRA
Non-standard to Standard map (OMOP)	Pulmonary tuberculosis	253954	SNOMED
Subsumes	Other specified pulmonary tuberculosis	44830894	ICD9CM
	Other specified pulmonary tuberculosis, bacteriological or histological examination not done	44836741	ICD9CM
	Other specified pulmonary tuberculosis, bacteriological or histological examination unknown (at present)	44836742	ICD9CM
	Other specified pulmonary tuberculosis, tubercle bacilli found (in sputum) by microscopy	44821641	ICD9CM
	Other specified pulmonary tuberculosis, tubercle bacilli not found (in sputum) by microscopy, but found by bacterial culture	44833188	ICD9CM

Figure 3.2: SNOMED code for Pulmonary Tuberculosis

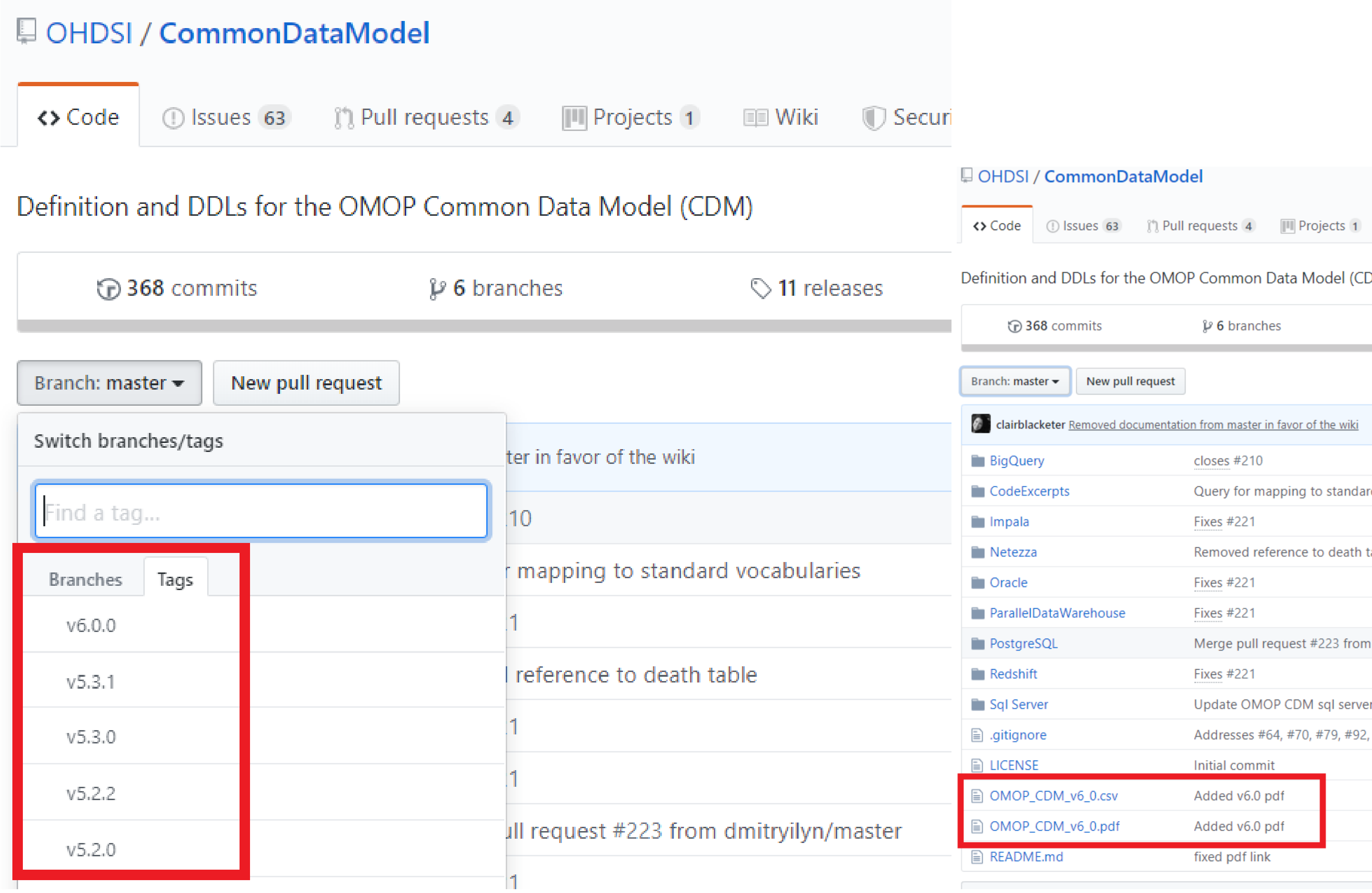
An example of how this relationship is depicted in the tables is shown in figure (**link to figure in CONDITION_OCCURRENCE**)

3.3 OMOP CDM Standardized Tables

The OMOP CDM contains 16 Clinical data tables, 10 Vocabulary tables, 2 Metadata tables, 4 Health System data tables, 2 Health Economics data tables, 3 standardized derived elements, and 2 results schema tables. To illustrate how these tables are utilized in practice the data of one person will be used as a common thread throughout the rest of the chapter. While part of the CDM the Vocabulary tables are not covered here, rather, they are detailed in depth in Chapter **??**.

3.3.1 OMOP-CDM (table specification)

OMOP-CDM OHDSI Common Data Model . 2019 5 25 6.0 . wiki page . OMOP-CDM specification Tags pdf .



BigQuery	closes #210
CodeExcerpts	Query for mapping to standar
Impala	Fixes #221
Netezza	Removed reference to death t
Oracle	Fixes #221
ParallelDataWarehouse	Fixes #221
PostgreSQL	Merge pull request #223 from
Redshift	Fixes #221
Sql Server	Update OMOP CDM sql server
.gitignore	Addresses #64, #70, #79, #92,
LICENSE	Initial commit
OMOP_CDM_v6_0.csv	Added v6.0 pdf
OMOP_CDM_v6_0.pdf	Added v6.0 pdf
README.md	fixed pdf link

3.3.2 OMOP-CDM

Open community OHDSI OMOP-CDM . , OMOP-CDM CDM . OMOP-CDM , .

OMOP-CDM , 2019 5 30 v6.0 .

3.3.3 Running Example: Endometriosis

Endometriosis is a painful condition whereby cells normally found in the lining of a woman's uterus occur elsewhere in the body. Severe cases can lead to infertility, bowel, and bladder problems. The following sections will detail one patient's experience with this disease and how her clinical experience might be represented in the Common Data Model.

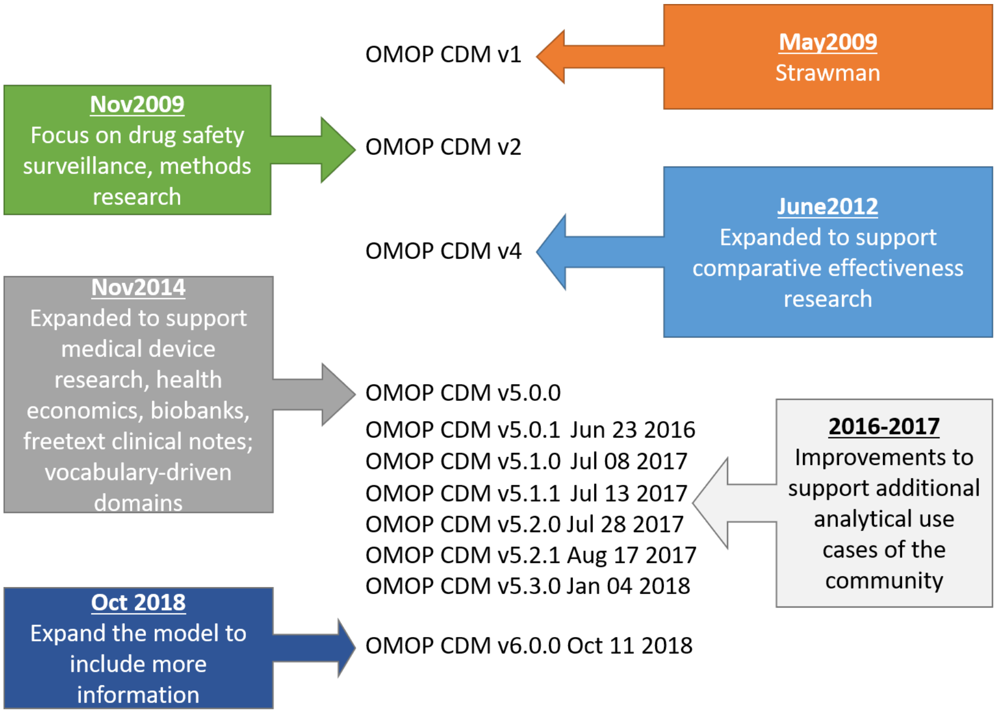

OMOP CDM v1
May2009
Strawman
Nov2009
Focus on drug safety surveillance, methods research
OMOP CDM v2
OMOP CDM v4
June2012
Expanded to support comparative effectiveness research
Nov2014
Expanded to support medical device research, health economics, biobanks, freetext clinical notes; vocabulary-driven domains
OMOP CDM v5.0.0
OMOP CDM v5.0.1 Jun 23 2016
OMOP CDM v5.1.0 Jul 08 2017
OMOP CDM v5.1.1 Jul 13 2017
OMOP CDM v5.2.0 Jul 28 2017
OMOP CDM v5.2.1 Aug 17 2017
OMOP CDM v5.3.0 Jan 04 2018
2016-2017
Improvements to support additional analytical use cases of the community
Oct 2018
Expand the model to include more information
OMOP CDM v6.0.0 Oct 11 2018

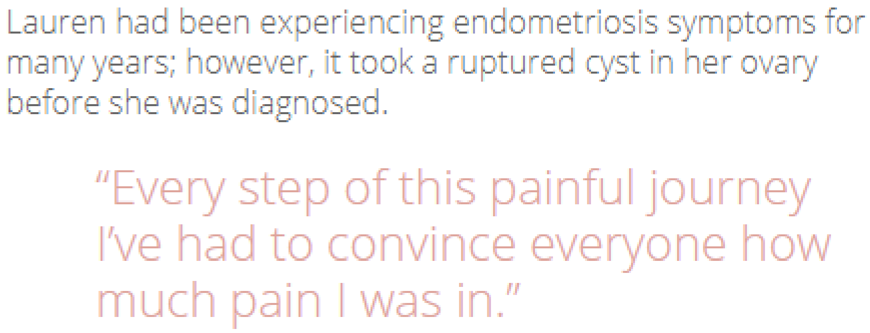
Figure 3.3: Evolution of OMOP-CDM

3.3.4 PERSON table

As the Common Data Model is a person-centric model (see section 3.2.1) let's start with how she would be represented in the PERSON table. For the full PERSON table specification please see the CDM wiki https://github.com/OHDSI/CommonDataModel/wiki/PERSON.

What do we know about Lauren?

- She is a 36-year-old woman
- Her birthday is 12-March-1982
- She is white
- She is english



Lauren had been experiencing endometriosis symptoms for many years; however, it took a ruptured cyst in her ovary before she was diagnosed.

"Every step of this painful journey I've had to convince everyone how much pain I was in."

Figure 3.4: Read more about Lauren and endometriosis at https://www.endometriosis-uk.org/laurens-story

With that in mind, her PERSON table might look something like this:

Column Name	Value	Explanation
person_id	1	Person_id should be an integer, either directly from the source or generated as part of the build process.
gender_concept_id	8532	The concept_id referring to female gender is 8532 (http://athena.ohdsi.org/search-terms/terms/8532).
year_of_birth	1982	
month_of_birth	3	
day_of_birth	12	
birth_datetime	1982-03-12 00:00:00	When the time is not known midnight is used.
death_datetime		
race_concept_id	8527	The concept_id referring to white race is 8527 (http://athena.ohdsi.org/search-terms/terms/8527).
ethnicity_concept_id	38003564	Typically hispanic status is stored for ethnicity. The concept_id 38003564 (http://athena.ohdsi.org/search-terms/terms/38003564) refers to 'Not hispanic'.
location_id		Her address is not known.
provider_id		Her primary care provider is not known.
care_site_id		Her primary care site is not known.
person_source_value	1	Typically this would be her identifier in the source data, though often is it the same as the person_id.
gender_source_value	F	The gender value as it appears in the source is stored here.
gender_source_concept_id	0	If the gender value in the source was coded using a vocabulary recognized by OHDSI, that concept_id would go here. For example, if her gender was 'Sex-F' in the source and it was stated to be in the PCORNet vocabulary concept_id 44814665 (http://athena.ohdsi.org/search-terms/terms/44814665) would go in this field.
race_source_value	white	The race value as it appears in the source is stored here.
race_source_concept_id	0	Same principle as gender_source_concept_id.
ethnicity_source_value	english	The ethnicity value as it appears in the source is stored here.
ethnicity_source_concept_id	0	Same principle as gender_source_concept_id.

3.3.5 OBSERVATION_PERIOD table

The OBSERVATION_PERIOD table is designed to define the amount of time for which a patient's clinical events are recorded in the source system. For US healthcare insurance claims this is typically the enrollment period of the patient. When working with data from electronic health records (EHR) often the first record in the system is considered the observation_period_start_date and the latest record is considered the observation_period_end_date with the understanding that only the clinical events that happened within that particular system were recorded. For the full OBSERVATION_PERIOD table specification please see the CDM wiki (https://github.com/OHDSI/CommonDataModel/wiki/OBSERVATION_PERIOD).

How can we determine Lauren's observation period?

Lauren's information is most similar to EHR data in that we only have records of her encounters from which to determine her observation period.

Encounter_ID	Start_Date	Stop_Date	EncounterClass
70	2010-01-06	2010-01-06	outpatient
80	2011-01-06	2011-01-06	outpatient
90	2012-01-06	2012-01-06	outpatient
100	2013-01-07	2013-01-07	outpatient
101	2013-01-14	2013-01-14	ambulatory
102	2013-01-17	2013-01-24	inpatient

Based on the encounter records her OBSERVATION_PERIOD table might look something like this:

Column Name	Value	Explanation
observation_period_id	1	This is typically an autogenerated field that creates a unique id number for each record in the table.
person_id	1	This comes from the PERSON table and links PERSON and OBSERVATION_PERIOD.
observation_period_start_date	2010-01-06	This is the start date of her earliest encounter on record.
observation_period_end_date	2013-01-24	This is the end date of her latest encounter on record.
period_type_concept_id	44814725	The best option in the Vocabulary with the concept class 'Obs Period Type' is 44814724 (http://athena.ohdsi.org/search-terms/terms/44814724), which stands for 'Period covering healthcare encounters'.

3.3.6 VISIT_OCCURRENCE

The VISIT_OCCURRENCE table houses information about a patient's encounters with the health care system. Within the OHDSI vernacular these are referred to as visits and are considered to be discreet events. There are 12 categories of visits though the most common are inpatient, outpatient, emergency and long term care. For the full VISIT_OCCURRENCE table specification please see the CDM wiki (https://github.com/OHDSI/CommonDataModel/wiki/VISIT_OCCURRENCE).

How do we represent Lauren's encounters as visits?

Revisting the encounters we used to determine her observation period:

Encounter_ID	Start_Date	Stop_Date	EncounterClass
70	2010-01-06	2010-01-06	outpatient
80	2011-01-06	2011-01-06	outpatient
90	2012-01-06	2012-01-06	outpatient

Encounter_ID	Start_Date	Stop_Date	EncounterClass
100	2013-01-07	2013-01-07	outpatient
101	2013-01-14	2013-01-14	ambulatory
102	**2013-01-17**	**2013-01-24**	**inpatient**

As an example let's represent the inpatient encounter as a record in the VISIT_OCCURRENCE table.

Column Name	Value	Explanation
visit_occurrence_id	514	This is typically an autogenerated field that creates a unique id number for each visit on the person's record in the converted CDM database.
person_id	1	This comes from the PERSON table and links PERSON and VISIT_OCCURRENCE.
visit_concept_id	9201	The concept_id referring to an inpatient visit is 9201 (http://athena.ohdsi.org/search-terms/terms/9201).
visit_start_date	2013-01-17	The start date of the visit.
visit_start_datetime	2013-01-17 00:00:00	The date and time of the visit started. When time is unknown midnight is used.
visit_end_date	2013-01-24	The end date of the visit. If this is a one-day visit the end date should match the start date.
visit_end_datetime	2013-01-24 00:00:00	The date and time of the visit end. If time is unknown midnight is used.
visit_type_concept_id	32034	This column is intended to provide information about the provenance of the visit record, i.e. does it come from an insurance claim, hospital billing record, EHR record, etc. For this example the concept_id 32035 (http://athena.ohdsi.org/search-terms/terms/32035) is used as the encounters are similar to electronic health records
provider_id*	NULL	If the encounter record has a provider associated, the id for that provider goes in this field. This should be the provider_id from the PROVIDER table that represents the provider on the encounter.
care_site_id	NULL	If the encounter record has a care site associated, the id for that care site goes in this field. This should be the care_site_id from the CARE_SITE table that codes for the care site on the encounter.
visit_source_value	inpatient	The visit value as it appears in the source goes here. In this context 'visit' means outpatient, inpatient, emergency, etc.
visit_source_concept_id	0	If the visit value from the source is coded using a vocabulary that is recognized by OHDSI, the concept_id that represents the visit source value would go here.
admitted_from_concept_id	0	If known, this is the concept_id that represents where the patient was admitted from. This concept should have the concept class 'Place of Service' and the domain 'Visit'. For example, if a patient was admitted to the hospital from home, the concept_id would be 8536 (http://athena.ohdsi.org/search-terms/terms/8536).

Column Name	Value	Explanation
admitted_from_source_value	NULL	This is the value from the source that represents where the patient was admitted from. Using the above example, this would be 'home'.
discharge_to_concept_id	0	If known, this is the concept_id that represents where the patient was discharged to. This concept should have the concept class 'Place of Service' and the domain 'Visit'. For example, if a patient was released to an assisted living facility, the concept_id would be 8615 (http://athena.ohdsi.org/search-terms/terms/8615).
discharge_to_source_value	0	This is the value from the source that represents where the patient was discharged to. Using the above example, this would be 'assisted living facility'.
preceding_visit_occurrence_id	NULL	The visit_occurrence_id for the visit immediately preceding the current one in time for the patient.

*A patient may interact with multiple health care providers during one visit, as is often the case with inpatient stays. These interactions can be recorded in the VISIT_DETAIL table. While not covered in depth in this chapter, you can read more about the VISIT_DETAIL table on the CDM wiki (https://github.com/OHDSI/CommonDataModel/wiki/VISIT_DETAIL)

3.3.7 CONDITION_OCCURRENCE

Records in the CONDITION_OCCURRENCE table are diagnoses, signs, or symptoms of a condition either observed by a Provider or reported by the patient.

What are Lauren's conditions?

Revisiting her account she says "About 3 years ago I noticed my periods, which had also been painful, were getting increasingly more painful. I started becoming aware of a sharp jabbing pain right by my colon and feeling tender and bloated around my tailbone and lower pelvis area. My periods had become so painful that I was missing 1-2 days of work a month. Painkillers sometimes dulled the pain, but usually they didn't do much."

The SNOMED code for painful menstruation cramps, otherwise known as dysmenorrhea, is 266599000. Let's see how that would be represented in the CONDITION_OCCURRENCE table:

Column	Value	Explanation
condition_occurrence_id	964	This is typically an autogenerated field that creates a unique id number for each condition on the person's record in the converted CDM database.
person_id	1	This comes from the PERSON table and links PERSON and CONDITION_OCCURRENCE.
condition_concept_id	194696	The concept_id that represents the SNOMED code 266599000 is 194696 (http://athena.ohdsi.org/search-terms/terms/194696)
condition_start_date	2010-01-06	The date when the instance of the Condition is recorded.
condition_start_datetime	2010-01-06 00:00:00	The date and time when the instance of the Condition is recorded. Midnight is used when the time is unknown

Column	Value	Explanation
condition_end_date	NULL	If known, this is the date when the instance of the Condition is considered to have ended.
condition_end_datetime	NULL	If known, this is the date and time when the instance of the Condition is considered to have ended.
condition_type_concept_id	32020	This column is intended to provide information about the provenance of the condition, i.e. does it come from an insurance claim, hospital billing record, EHR record, etc. For this example the concept_id 32020 (http://athena.ohdsi.org/search-terms/terms/32020) is used as the encounters are similar to electronic health records. Concept_ids in this field should be in the 'Condition Type' vocabulary.
condition_status_concept_id	0	If known, the condition_status_concept_id represents when and/or how the condition was diagnosed. For example, a condition could be an admitting diagnosis, in which case the concept_id 4203942 (http://athena.ohdsi.org/search-terms/terms/4203942) would be used.
stop_reason	NULL	If known, the reason that the Condition was no longer present, as indicated in the source data.
provider_id	NULL	If the condition record has a diagnosing provider listed, the id for that provider goes in this field. This should be the provider_id from the PROVIDER table that represents the provider on the encounter.
visit_occurrence_id	509	If known, this is the visit (represented as visit_occurrence_id taken from the VISIT_OCCURRENCE table) during which the condition was diagnosed.
visit_detail_id	NULL	If known, this is the visit detail encounter (represented as visit_detail_id from the VISIT_DETAIL table) during which the condition was diagnosed.
condition_source_value	266599000	This is the value from the source that represents the condition. In Lauren's case of dysmenorrhea the SNOMED code for that condition is stored here and the standard concept_id mapped from that code is stored in CONDITION_CONCEPT_ID.
condition_source_concept_id	194696	If the condition value from the source is coded using a vocabulary that is recognized by OHDSI, the concept_id that represents that value would go here. In the example of dysmennorhea the source value is a SNOMED code so the concept_id that represents that code is 194696. In this case it is the same as the condition_concept_id since the SNOMED vocabulary is the standard condition vocabulary
condition_status_source_value	0	If the condition status value from the source is coded using a vocabulary that is recognized by OHDSI, the concept_id that represents that source value would go here.

3.3.8 DRUG_EXPOSURE

The DRUG_EXPOSURE captures records about the utilization of a Drug when ingested or otherwise introduced into the body. Drugs include prescription and over-the-counter medicines, vaccines, and large-molecule biologic therapies. Radiological devices ingested or applied locally do not count as Drugs.

Drug Exposure is inferred from clinical events associated with orders, prescriptions written, pharmacy dispensings, procedural administrations, and other patient-reported information.

What are Lauren's drug exposures?

We know that Lauren was given 60 acetaminophen 325mg oral tablets for 30 days (NDC code 69842087651) at her visit on 2010-01-06 to help with her dysmenorrhea pain. Here's how that might look in the DRUG_EXPOSURE table:

Column	Value	Explanation
drug_exposure_id	1001	This is typically an autogenerated field that creates a unique id number for each drug_exposure on the person's record in the converted CDM database.
person_id	1	This comes from the PERSON table and links PERSON and DRUG_EXPOSURE.
drug_concept_id	1127433	The NDC code for acetaminophen maps to the RxNorm code 313782 which is represented by the concept_id 1127433 (http://athena.ohdsi.org/search-terms/terms/1127433).
drug_exposure_start_date	2010-01-06	The start date of the drug exposure
drug_exposure_start_datetime	2010-01-06 00:00:00	The start date and time of the drug exposure. Midnight is used when the time is not known.
drug_exposure_end_date	2010-02-05	The end date of the drug exposure. Depending on different sources, it could be a known or an inferred date and denotes the last day at which the patient was still exposed to the drug. In this case the end is inferred since we know Lauren had a 30 days supply.
drug_exposure_end_datetime	2010-02-05 00:00:00	The end date and time of the drug exposure. Similar rules apply as to drug_exposure_end_date. Midnight is used when time is unknown
verbatim_end_date	NULL	If the source provides an end date rather than just days supply that date goes here.
drug_type_concept_id	38000177	This column is intended to provide information about the provenance of the drug, i.e. does it come from an insurance claim, prescription record, etc. For this example the concept_id 38000177 (http://athena.ohdsi.org/search-terms/terms/38000177) is used as the drug record is from a written prescription. Concept_ids in this field should be in the 'Drug Type' vocabulary.
stop_reason	NULL	The reason the Drug was stopped. Reasons include regimen completed, changed, removed, etc.
refills	NULL	The number of refills after the initial prescription. The initial prescription is not counted, values start with null. In the case of Lauren's acetaminophen she did not have any refills so the value is NULL.

Column	Value	Explanation
quantity	60	The quantity of drug as recorded in the original prescription or dispensing record.
days_supply	30	The number of days of supply of the medication as prescribed.
sig	NULL	The directions ('signetur') on the Drug prescription as recorded in the original prescription (and printed on the container) or dispensing record.
route_concept_id	4132161	This concept is meant to represent the route of the drug the patient was was exposed to. Lauren took her acetaminophen orally so the concept_id 4132161 (http://athena.ohdsi.org/search-terms/terms/4132161) is used.
lot_number	NULL	An identifier assigned to a particular quantity or lot of Drug product from the manufacturer.
provider_id	NULL	If the drug record has a prescribing provider listed, the id for that provider goes in this field. This should be the provider_id from the PROVIDER table that represents the provider on the encounter.
visit_occurrence_id	509	If known, this is the visit (represented as visit_occurrence_id taken from the VISIT_OCCURRENCE table) during which the drug was prescribed.
visit_detail_id	NULL	If known, this is the visit detail (represented as visit_detail_id taken from the VISIT_DETAIL table) during which the drug was prescribed.
drug_source_value	69842087651	This is the source code for the Drug as it appears in the source data. In Lauren's case she was prescribed acetaminophen and the NDC code is stored here.
drug_source_concept_id	750264	This is the concept_id that represents the drug source value. In this example the concept_id is 750264 (http://athena.ohdsi.org/search-terms/terms/750264).
route_source_value	NULL	The information about the route of administration as detailed in the source.
dose_unit_source_value	NULL	The information about the dose unit as detailed in the source.

3.3.9 PROCEDURE_OCCURRENCE

The PROCEDURE_OCCURRENCE table contains records of activities or processes ordered by, or carried out by, a healthcare provider on the patient to have a diagnostic or therapeutic purpose. Procedures are present in various data sources in different forms with varying levels of standardization. For example:

- Medical Claims include procedure codes that are submitted as part of a claim for health services rendered, including procedures performed.
- Electronic Health Records that capture procedures as orders.

What procedures did Lauren have? From her description we know she had a ultrasound of her left ovary on 2013-01-14 that showed a 4x5cm cyst. Here's how that would look in the PROCEDURE_OCCURRENCE table:

Column	Value	Explanation
procedure_occurrence_id	1277	This is typically an autogenerated field that creates a unique id number for each procedure_occurrence on the person's record in the converted CDM database.
person_id	1	This comes from the PERSON table and links PERSON and PROCEDURE_OCCURRENCE
procedure_concept_id	4127451	The SNOMED procedure code for a pelvic ultrasound is 304435002 which is represented by the concept_id 4127451 (http://athena.ohdsi.org/search-terms/terms/4127451).
procedure_date	2013-01-14	The date on which the procedure was performed.
procedure_datetime	2013-01-14 00:00:00	The date and time on which the procedure was performed. Midnight is used when time is unknown.
procedure_type_concept_id	38000275	This column is intended to provide information about the provenance of the procedure, i.e. does it come from an insurance claim, EHR order, etc. For this example the concept_id 38000275 (http://athena.ohdsi.org/search-terms/terms/38000275) is used as the procedure record is from an EHR record. Concept_ids in this field should be in the 'Procedure Type' vocabulary.
modifier_concept_id	0	This is meant for a concept_id representing the modifier on the procedure. For example, if the record indicated that a CPT4 procedure was performed bilaterally then the concept_id 42739579 (http://athena.ohdsi.org/search-terms/terms/42739579) would be used.
quantity	0	The quantity of procedures ordered or administered.
provider_id	NULL	If the procedure record has a provider listed, the id for that provider goes in this field. This should be the provider_id from the PROVIDER table that represents the provider on the encounter.
visit_occurrence_id	740	If known, this is the visit (represented as visit_occurrence_id taken from the VISIT_OCCURRENCE table) during which the procedure was performed.
visit_detail_id	NULL	If known, this is the visit detail (represented as visit_detail_id taken from the VISIT_DETAIL table) during which the procedure was performed.
procedure_source_value	304435002	The source code for the Procedure as it appears in the source data. This code is mapped to a standard procedure Concept in the Standardized Vocabularies and the original code is, stored here for reference.
procedure_source_concept_id	4127451	This is the concept_id that represents the procedure source value.

Column	Value	Explanation
modifier_source_value	NULL	The source code for the modifier as it appears in the source data.

Chapter 4

The OMOP Vocabulary

- ATHENA: ATHENA OMOP vocabulary .
- : OHDSI wiki page .
- OMOP vocabulary github : OMOP vocabulary github
- OHDSI Github : OHDSI vocabulary github
- OHDSI tutorial : OHDSI past event OMOP vocabulary tutorial .

4.1 Design Principles

Chapter 5

Extract Transform Load

ETL .

- OHDIS ETL best practice: OHDIS ETL best practice
- ETL : ETL github
- ETL sample : ETL sample page
- ETL tools : ETL tool page
- THEMIS WG: Korean THEMIS WG .

5.1 Pre-processing

5.1.1 WhiteRabbit and Rabbit-in-a-Hat

Chapter 6

SQL and R

6.1 Database Connector

- Database Connnector github

6.2 SQL Render

- SQL Render github
- SQL Render
- OHDSI SQL Devloper

Chapter 7

Cohort

7.1 Using SQL

7.2 ATLAS

7.3 Phenotype Library

Chapter 8

Characterization

- Cohort definition tutorial: OHDSI past event Cohort definition .

8.1 FeatureExtraction

8.2 ATLAS

8.2.1 Baseline characteristics

8.2.2 Incidence rate calculation

Chapter 9

Population-Level Estimation

- Population-Level Estimation : OHDSI past event Population-level estimation .

Chapter 10

Patient-Level Prediction

- Patient-Level Prediction : OHDSI past event Patient-Level Prediction .

Chapter 11

Extension of CDM

11.1 Genomic CDM

11.2 Radiology CDM

11.3 AEGIS