

# OMOP Myth Busters

## *Dispelling misconceptions about common data model transformations*

### THE PROBLEM

Observational health databases are not created equal. The purpose for collecting the data, the format of the data and terminologies used differ among healthcare settings and data types (e.g. electronic health record, patient registries, administrative claims). Without standardization, conducting a study that uses multiple observational databases has proven to be expensive, time-consuming and difficult to replicate.

### THE SOLUTION

Converting the data in these disparate databases into a common data model (CDM), with common representation, (terminologies, vocabularies, coding schemes) provides researchers with the ability to conduct studies that are cost-effective, faster, and more reliable.

OMOP<sup>1</sup> is a common data model that allows for the systematic analyses of disparate observational databases using standardized analytical methods and tools developed by the OHDSI<sup>2</sup> community. The beneficial combination of a common data model along with common methodologies, provides researchers with the ability to perform health research analytics at scale.

### THE MYTHS AND THE TRUTHS

As interest in using a common data model to analyze multiple heterogeneous databases increases, so do questions about the accuracy and effectiveness of these data conversions. IQVIA is a global leader in OMOP common data model conversions. We currently host more than 12 datasets in the OMOP format and have conducted over 20 conversions.

---

Here are some common misconceptions we have heard over time and the truths behind them:

- 1 **Myth #1**  
"Loss of Data"
- 2 **Myth #2**  
"Loss of Accuracy in Conversion"
- 3 **Myth #3**  
"Loss of Accuracy in Vocabulary Mapping"
- 4 **Myth #4**  
"It Takes Too Much Time"
- 5 **Myth #5**  
"You Don't Have My Use Case in OMOP"
- 6 **Myth #6**  
"I Have to Learn New Medical Terminology"

<sup>1</sup>OHDSI (Observational Health Data Science and Informatics) is a public initiative independent of IQVIA

<sup>2</sup>OMOP (Observational Medical Outcomes Partnership) was a public partnership between FDA and industry, developing the OMOP CDM and Standardized Vocabularies; now maintained by OHDSI

## Myth #1 “Loss of Data”

# Data Quality

### MYTH

“Converting to a CDM will result in “losing” data because it does not map to the standard.”

### TRUTH

After converting data from source to OMOP, IQVIA has a standard quality control service including:

- OHDSI Data Quality Dashboard
- Quality Control checks
- On-premise checks
- Data Profiling checks

## RESOURCES

	Verification			Validation			Total					
	Pass	Fail	Total	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	
Plausibility	1611	228	1839	87%	274	13	287	95%	1885	241	2126	89%
Conformance	590	91	681	87%	97	7	104	93%	687	98	785	88%
Completeness	329	57	386	85%	13	2	15	87%	342	59	401	85%
<b>Total</b>	<b>2530</b>	<b>376</b>	<b>2906</b>	<b>87%</b>	<b>384</b>	<b>22</b>	<b>406</b>	<b>95%</b>	<b>2914</b>	<b>398</b>	<b>3312</b>	<b>88%</b>

General Database Information	Data Collection: Adjudicated claims (accepted and paid by the payer)	Patient Type: Inpatient, outpatient and emergency room
	Data Coverage: Number of Patients = 150 Million	History / Update Frequency: Date start date = 2/2000 Refresh Period = Quarterly
	Geographical distribution: Subsidiary regions of the United States that Progen are commercially present	
Key Patient Information	Patient demographics: Year of birth Gender 3 digit zip code	Conditions: Total number of records = 11.8B Percentage mapped to Standard Concept = 98% Number of relevant Condition Concepts (97th percentile) = 3.61k
	Drugs: Total number of records = 6.32B Percentage mapped to Standard Concept = 99% Number of relevant Drug Concepts (97th percentile) = 2.78k	Procedures: Total number of records = 8.35B Percentage mapped to Standard Concept = 99% Number of relevant Procedure Concepts (97th percentile) = 1.84k
Key Physician & Practice Information	Providers: Number of providers = 52.9M	Care Sites: Number of care sites - Not reported These are not identified but typical organization settings
	Visits: Number of outpatient visits = 4.89B Number of inpatient visits = 360M	

OID	ORIGINTABLE	ORIGINFIELD	TARGETTABLE	TARGETFIELD	VALIDATIONFUNCTION	ORIGINTABLE_VALIDATIONRULE
13001	care_site	care_site_id	visit_occurrence	visit_occurrence_id	A unique identifier for each Care Site	visit_occurrence_id = care_site_id
13002	care_site	care_site_name	visit_occurrence	visit_occurrence_id	From the domain and standard concept = 0. 10 rows to find the best	visit_occurrence_id = care_site_name
13003	condition_occurrence	condition_occurrence_id	person	person_id	This is a required field. A foreign key identifier to the Person in the person table.	person_id = condition_occurrence_id
13004	condition_occurrence	condition_occurrence_start_date	person	person_id	This is a required field. It is populated with a concept_id from condition_occurrence_start_date. Condition and standard concept = 0.	person_id = condition_occurrence_start_date
13005	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a date after 0th person year of birth.	person_id = condition_occurrence_start_date
13006	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a date before 0th person year of birth.	person_id = condition_occurrence_start_date
13007	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a date after 0th person year of birth.	person_id = condition_occurrence_start_date
13008	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a date before 0th person year of birth.	person_id = condition_occurrence_start_date
13009	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a concept_id from domain_id = 7999 Concept, procedure = condition_type, and standard concept = 0. 10 rows to find the best	person_id = condition_occurrence_start_date
13010	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a concept_id from domain_id = 7999 Concept, procedure = condition_type, and standard concept = 0. 10 rows to find the best	person_id = condition_occurrence_start_date
13011	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a concept_id from domain_id = 7999 Concept, procedure = condition_type, and standard concept = 0. 10 rows to find the best	person_id = condition_occurrence_start_date
13012	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a concept_id from domain_id = 7999 Concept, procedure = condition_type, and standard concept = 0. 10 rows to find the best	person_id = condition_occurrence_start_date
13013	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a concept_id from domain_id = 7999 Concept, procedure = condition_type, and standard concept = 0. 10 rows to find the best	person_id = condition_occurrence_start_date
13014	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a concept_id from domain_id = 7999 Concept, procedure = condition_type, and standard concept = 0. 10 rows to find the best	person_id = condition_occurrence_start_date
13015	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a concept_id from domain_id = 7999 Concept, procedure = condition_type, and standard concept = 0. 10 rows to find the best	person_id = condition_occurrence_start_date
13016	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a concept_id from domain_id = 7999 Concept, procedure = condition_type, and standard concept = 0. 10 rows to find the best	person_id = condition_occurrence_start_date
13017	condition_occurrence	condition_occurrence_start_date	person	person_id	Must be populated with a concept_id from domain_id = 7999 Concept, procedure = condition_type, and standard concept = 0. 10 rows to find the best	person_id = condition_occurrence_start_date
13018	death	death_date	person	person_id	This is a required field. A foreign key identifier to the Person in the person table. Each person_id can only have 1 death_date.	person_id = death_date
13019	death	death_date	person	person_id	Must be populated with a date after 0th person year of birth.	person_id = death_date

## Myth #2 “Loss of Accuracy in Conversion”

# Retaining the Accuracy of Source Data

### MYTH

“Using OMOP standards can degrade the accuracy of the data. There could be issues in the conversions ability to accurately reflect a data set.”

### TRUTH

- Validation studies have found minimal differences in the source to OMOP data
- DA France / LPD France validation study found consistency between native and OMOP data sets

Reference: Schwalm M, Raoul T, Chu D, Shah U, Potdar M, Van Zandt M, Coffin G, Jouaville SL. Conversion of a French Electronic Medical Record (Emr) Database into the Observational Medical Outcomes Partnership Common Data Model. Research on Methods – Databases & Management Methods. 2017 Oct 01; vol 20.issue 9, PA741

## RESOURCES

**CONVERSION OF A FRENCH ELECTRONIC MEDICAL RECORD DATABASE INTO THE OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP COMMON DATA MODEL**

Marie-Sophie Schouder<sup>1</sup>, Thomas Raoul<sup>2</sup>, Diane Chou<sup>3</sup>, Jiji Shah<sup>4</sup>, Maghaneh Potdar<sup>5</sup>, Mia Van Zandt<sup>6</sup>, Gabriel Coffin<sup>7</sup>, Sophie L. Jouaville<sup>8</sup>, IQVIA, Biologie Bilancomet, France, <sup>1</sup>IQVIA, New Jersey, USA, <sup>2</sup>QuintilesIMS, Pennsylvania, USA, <sup>3</sup>IQVIA, California, USA

<sup>1</sup> Corresponding Author: sophie.jouaville@iqvia.com

**Table 3 : Patients profile : comparison between the 3 data sources : DA FR Native / DA FR OMOP / LPD Native**

	DA FR NATIVE	DA FR OMOP	LPD NATIVE
	N=12 302*	N=12 382*	N=15 623
Males	7 179 (58.4)	7 231 (58.4)	9 291 (59.5)
Age (in year)	74.6 (± 11.1)	74.4 (± 11.1)	74.6 (± 11.1)
Age ≥75 years	7 055 (57.3)	7 032 (56.8)	8 981 (57.5)
BMI	28.1 (± 5.3)	28.2 (± 5.3)	27.9 (± 5.3)
Diastolic blood pressure	76.2 (± 12)	79.3 (± 20)	76 (± 9)
Systolic blood pressure	131.9 (± 15)	132.7 (± 15)	133 (± 15)
Co-treatments			
NSAIDs	5 320 (43.3)	5 341 (43.1)	7 492 (48.0)
Anti-arrhythmic drug	6 014 (48.9)	6 018 (48.6)	7 425 (47.5)
Injectable anticoagulants	373 (3.0)	377 (3.0)	482 (3.1)
CHA <sub>2</sub> DS <sub>2</sub> Score <sup>(2)</sup>			
0	2 262 (18.4)	2 311 (18.7)	2 638 (16.9)
1	3 997 (32.5)	4 027 (32.5)	5 026 (32.2)
≥2	6 043 (49.1)	6 044 (48.8)	7 959 (50.9)
CHA <sub>2</sub> DS <sub>2</sub> -Vasc Score <sup>(3)</sup>			
0	822 (6.7)	844 (6.8)	998 (6.4)
1	1 559 (12.7)	1 591 (12.8)	1 774 (11.4)
≥2	9 921 (80.6)	9 947 (80.3)	12 851 (82.3)

\*OMOP model assign an occurrence date to all events including clinical measures and lab results. As a result there is a slight difference in visit number between DA FR OMOP and DA FR NATIVE, which explains a slightly elevated number of included patients in DA FR OMOP.

## Myth #3 “Loss of Accuracy in Vocab Mapping”

# Evaluating the Accuracy of Vocabulary Mapping

### MYTH

“OMOP vocabulary mappings are incorrect. There could be issues in the preservation of source information as it is translated to standard concepts.”

### TRUTH

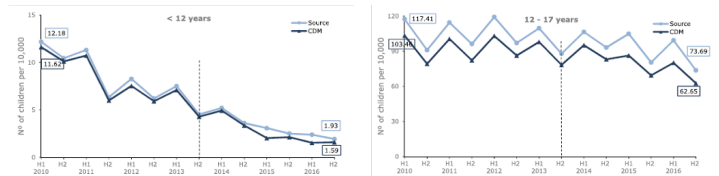
- Validation studies have found minimal differences in the source to OMOP data
- EMA Validation study of IQVIA IMRD UK found consistency between source and OMOP CDM data

## RESOURCES

> Clin Pharmacol Ther. 2020 Apr;107(4):915-925. doi: 10.1002/cpt.1785. Epub 2020 Mar 2.

### Can We Rely on Results From IQVIA Medical Research Data UK Converted to the Observational Medical Outcome Partnership Common Data Model?: A Validation Study Based on Prescribing Codeine in Children

Gianmario Candore <sup>1</sup>, Karin Hedenmalm <sup>1</sup>, Jim Slattery <sup>2</sup>, Alison Cave <sup>2</sup>, Xavier Kurz <sup>2</sup>, Peter Arlett <sup>2</sup>  
Affiliations — collapse



### SIX-MONTHLY PREVALENCE (PER 10,000) OF CODEINE PRESCRIBING FOR PAIN IN 0-17 YEARS

Reference: Candore G, Hedenmalm K, Slattery J, Cave A, Kurz X, Arlett P. Can We Rely on Results From IQVIA Medical Research Data UK Converted to the Observational Medical Outcome Partnership Common Data Model?: A Validation Study Based on Prescribing Codeine in Children. Clin Pharmacol Ther. 2020 Apr;107(4):915-925.

## Myth #4 “It Takes Too Much Time”

# OMOP Conversion Overview

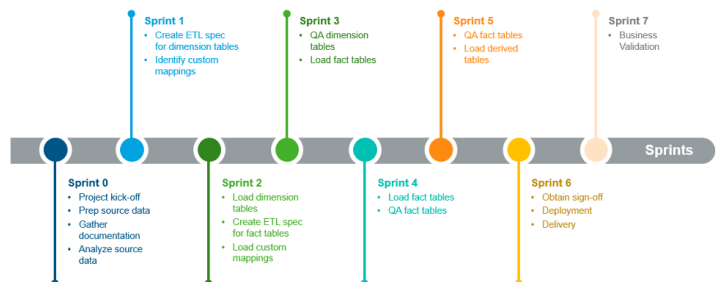
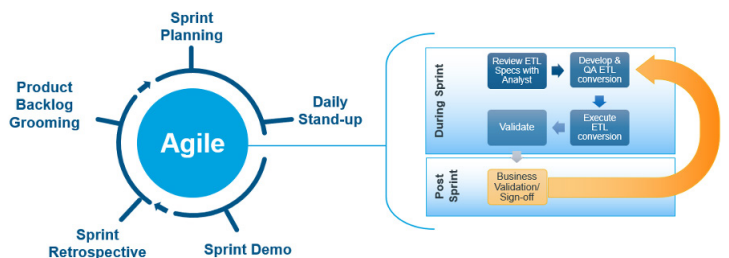
### MYTH

“Taking data from source format to OMOP common data model is tedious and time consuming.”

### TRUTH

- Yes, it takes time to convert data into the OMOP CDM
- We spend time cleaning the data and removing data that cannot contribute to analytical use cases
- We push down common business rules (e.g. patient eligibility criteria, observational period, validity of conditions) into the ETL process. However, this saves significant time during execution of the analytics study packages

## RESOURCES



## Myth #5 “You Don’t Have My Use Case in OMOP”

# OMOP Evolves to Meet Analytical Needs

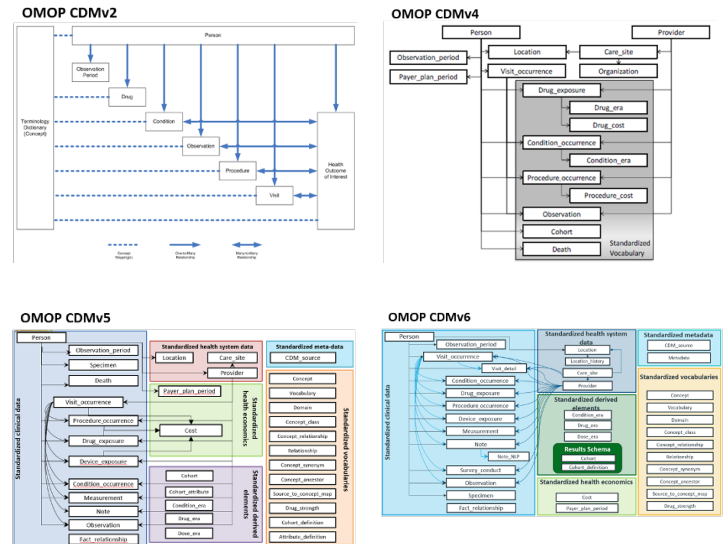
### MYTH

“OMOP is not good enough for my analytical use case or doesn’t cover the therapeutic area that I want to study.”

### TRUTH

- It’s true, OMOP was not built for every use case
- OMOP continues to evolve to support additional use cases when there is enough of a common need. For example, to support oncology data, OHDSI’s Oncology working group designed an Oncology extension to house oncology-specific information in the OMOP CDM

## RESOURCES



## Myth #6 “I Have to Learn New Medical Terms”

# OMOP Vocabulary Hierarchy

### MYTH

“I have to learn new medical terms. OMOP is forcing me to speak in SNOMED, RxNorm and LOINC codes.”

### TRUTH

- OMOP CDM preserves the source codes from the native data and creates a map to a standard concept that is interoperable across all data assets
- You can always start with source codes (e.g., ICD-9/ ICD-10) and use the maps to relationships to find standards
- The hierarchy structure in the standard vocabulary is easily navigated using ancestors and descendants

## RESOURCES

### Ancestry Relationships

