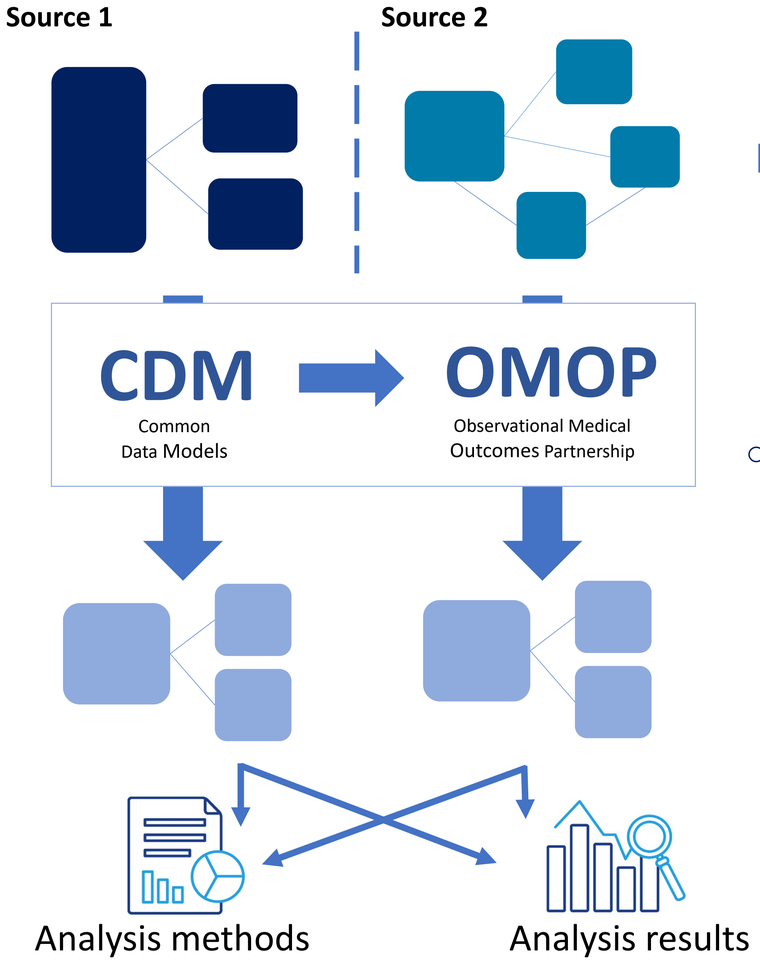
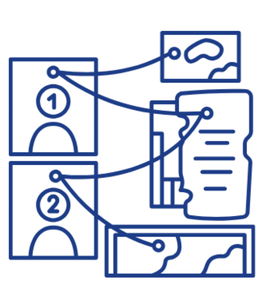
PT7

DATA STANDARDIZATION IN BRAZIL: AN OMOP COMMON DATA MODEL APPROACH IN A DATASUS COHORT

Julio Cesar Barbour Oliveira*[1], Guilherme Silva Julian[2], Jessica Mayumi Maruyama[1],
[1] Precision Data, São Paulo, Brazil, [2] Pfizer, São Paulo, Brazil
* jb@precisiondata.com.br

OBJECTIVES





quick multi-site and country real world evidence generation

- This study aims to describe the methods and quality results of the **creation of an OMOP dataset** with linked **outpatient** and **inpatient** information of Brazilian claims data.

METHODS

Data selected:

Hospital Information System (SIH)

Outpatient procedures, consultations, ICD-10 codes of a primary and secondary diagnosis, medicines, and personal data

Outpatient Information System (SAI)

Inpatient personal information, procedures, treatments and separation (hospital discharges, transfers, and deaths)

Record linkage and OMOP CDM harmonization:

Pre-processing and cleaning stage

Deterministic linkage algorithm

Transformation into OMOP CDM Model

Exclusion: inconsistencies in basic information (date of birth or gender); Patients with different primary keys but with matching basic information

Key information: zip code, date of birth, and gender

Exclusion: Patients from zip codes with more than 2500 distinct individuals linked to them, patients with more than 3 distinct zip codes, and patient's ID before 2012

Data Quality Dashboard: overall pass rate of 98%

RESULTS

5.82 million

patients included in the final dataset.

960 queries testing quality

Table 1. Data Quality Dashboard of OMOP Brazilian Data

	% Pass
Plausibility	92,6
Conformance	92,5
Completeness	100

* indicating a **satisfactory index** of the mapping quality.

Table 2. Total pass per populated OMOP table

OMOP Table	% Pass
CARE_SITE	94,4
CONDITION_ERA	92,1
CONDITION_OCCURRENCE	96,6
DEATH	96,9
DEVICE_EXPOSURE	92,6
DRUG_ERA	95,2
DRUG_EXPOSURE	91,2
LOCATION	100,0
MEASUREMENT	95,9
OBSERVATION	95,9
OBSERVATION_PERIOD	93,9
PERSON	94,7
PROCEDURE_OCCURRENCE	92,3
PROVIDER	100,0
VISIT_OCCURRENCE	89,7

Table 3. Total pass considering entire OMOP base

Total of tests	Total of success	% Pass
960	907	94,50%

*Some tables were not completed for this OMO model due to the lack of certain information in the original database

Table 4. Total of distinct SUS patients after linkage and CDM harmonization per ICD-10 chapter

ICD-10 Chapter	Total Distinct Patients
I. Certain infectious and parasitic diseases	833.125
II. Neoplasms	1.340.140
III. Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	118.241
IV. Endocrine, nutritional and metabolic diseases	608.668
V. Mental and behavioural disorders	1.555.624
VI. Diseases of the nervous system	450.725
VII. Diseases of the eye and adnexa	629.169
VIII. Diseases of the ear and mastoid process	2.514.916
IX. Diseases of the circulatory system	517.390
X. Diseases of the respiratory system	709.830
XI. Diseases of the digestive system	879.286
XII. Diseases of the skin and subcutaneous tissue	350.571
XIII. Diseases of the musculoskeletal system and connective tissue	1.330.198
XIV. Diseases of the genitourinary system	999.132
XV. Pregnancy, childbirth and the puerperium	760.840
XVI. Certain conditions originating in the perinatal period	96.070
XVII. Congenital malformations, deformations and chromosomal abnormalities	23.198
XVIII. Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	100.317
XIX. Injury, poisoning and certain other consequences of external causes	1.061.239
XX. External causes of morbidity and mortality	68.572
XXI. Factors influencing health status and contact with health services	2.018.576
XXII. Codes for special purposes	553

CONCLUSION

- Our study provides the first results of a **high-quality cohort** with outpatient and inpatient information in **Brazil** in OMOP format.
- CMD approaches in Latin America are very scarce and may be a crucial tool to boost **real-world evidence** generation in the region.
- Futhermore, we anticipate conducting and publshing real-world evidence studies utilizing this dataset in the upcoming months.