

Background

Population-based pharmacoepidemiology studies rely on information about exposures and outcomes that occur in general populations in clinical practice. Real-world data, such as automated medical claims and electronic medical records, are commonly used when conducting pharmacoepidemiology studies.

Real-world data are collected to support patient care and associated billing. Systematically assessing the availability of variables in these data sources helps determine their suitability for research use. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) provides a structure for organizing and standardizing observational data. Three United States (US)-based IMS Health databases have been converted to the OMOP CDM v4 (Table 1).

The objective of this project is to assess data availability for pharmacoepidemiology research in the converted IMS databases using the OMOP CDM domains.

Table 1. IMS Health US Real-World Databases Converted to OMOP CDM v4

Database	Data Type	# Patients
Adjudicated Claims (PharMetrics Plus)	Prescriptions, hospital, medical claims	105,009,000
Electronic Medical Records (EMR)	Electronic medical records	37,860,000
Oncology EMR	Electronic medical records	2,059,000

Methods

1. A framework was created based on the OMOP CDM v4 specifications to define priority data elements for pharmacoepidemiology research studies. 11 of 18 domains of the OMOP CDM were deemed priorities for pharmacoepidemiology.
2. The Extract, Transform, Load specifications for the converted databases were systematically reviewed to determine the number of OMOP CDM concepts available, by domain.
 - Within a domain, the availability of concepts was classified as complete, partial or missing.
 - The review did not assess the completeness of specific measures (e.g., lab values) for all patients, but instead focused on the data structure available for the majority of patients.

Results

The availability of concepts within pharmacoepidemiology priority domains varies by database (Table 2).

- The EMR database contains information in all pharmacoepidemiology priority domains.
- PharMetrics Plus contains variables in all of the priority domains for pharmacoepidemiology except for Death. PharMetrics Plus also contains variables in the additional domains to support health economics, comparative effectiveness, and quality of care studies.
- The Oncology EMR contains variables in all priority domains for pharmacoepidemiology except for Procedure Occurrence.
- Availability of specific variables within a domain (e.g., drug use details such as stop reason) differs across databases.

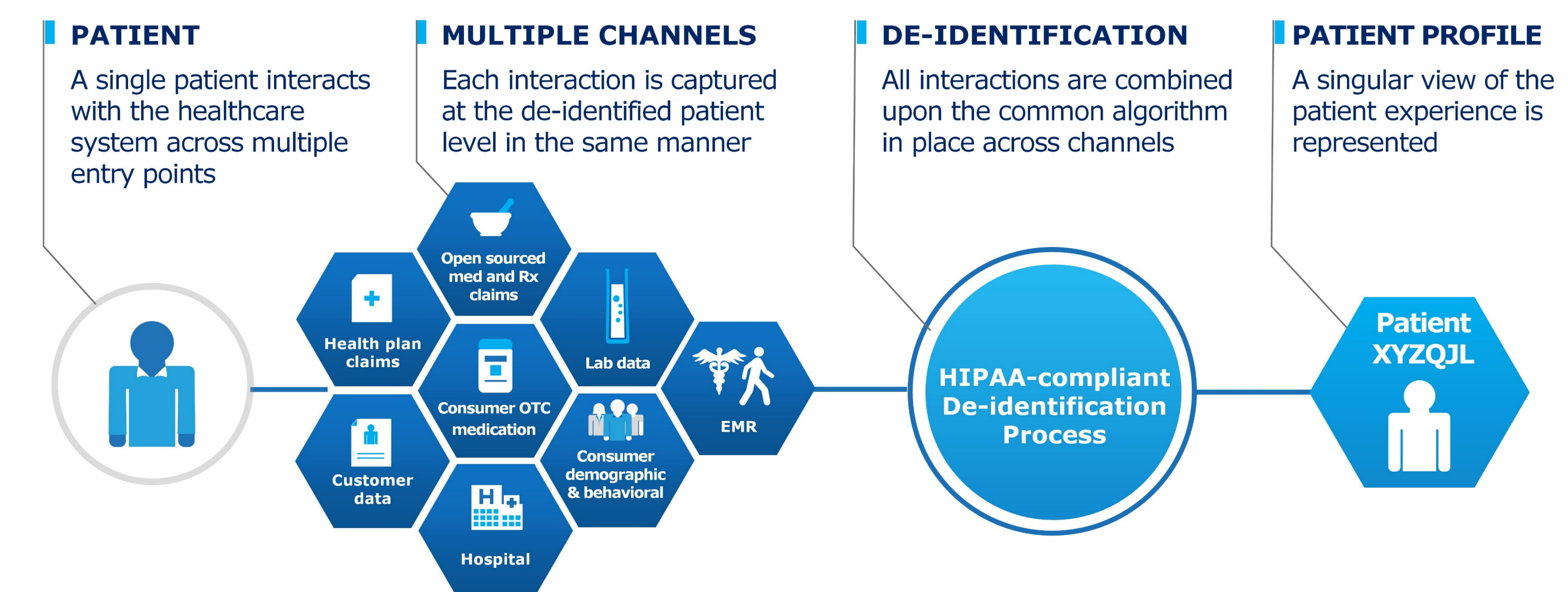
Linkage across datasets is possible using a Health Insurance Portability and Accountability Act (HIPAA) compliant deterministic algorithm (Figure).

Table 2. OMOP CDM v4 Concepts Available in IMS Health US Databases, by Domain

Domain	Description	PharMetrics Plus	EMR	Oncology EMR
Priority domains for pharmacoepidemiology				
Cohort	Person, provider, or visit cohorts	●	●	●
Condition Occurrence	Diagnosis or condition at a certain time	●	●	●
Condition Era	Diagnosis or condition over a period of time	●	●	●
Death	Time and cause of death	○	●	●
Drug Exposure	Person/drug association at a specific time	●	●	●
Drug Era	Person/drug association over a time period	●	●	●
Observation	Clinical facts (e.g., lab tests, signs/symptoms)	●	●	●
Observation Period	Time intervals when observations may be available	●	●	●
Person	Demographic information	●	●	●
Procedure Occurrence	Procedures carried out	●	●	○
Visit Occurrence	Visits for health care services	●	●	●
Additional domains				
Provider	Information about health care providers	●	●	●
Location	Addresses (patients, organizations, care sites)	●	●	●
Organization	Information about health care organizations	○	○	●
Care Site	Information about the site of care	○	●	●
Drug Cost	Drug exposure cost/payment information	●	○	○
Procedure Cost	Procedure cost/payment information	●	○	○
Payer Plan Period	Coverage plan of the person	●	○	○

Key: ● All concepts available for this domain ● Some concepts available for this domain ○ No concepts available for this domain

Figure. IMS Health Deidentification and Linking Process



Conclusions

- IMS Health real-world databases can be used to create large patient-level cohorts to support pharmacoepidemiology research. These databases contain many of the variables in OMOP CDM domains that are priorities for pharmacoepidemiology research.
- Depending on the research question, pharmacoepidemiology studies using all priority OMOP CDM domains are feasible by linking databases.
- Among the IMS Health data sources, a linked database in OMOP CDM v4 format is already available for 5,041,000 patients with records in both the EMR and PharMetrics Plus databases. Additional variables can be added by linking to other data sources, such as registries that include patient-reported outcomes.