

Name:	Karin Johnson
Affiliation:	IMS Health, Epidemiology and Drug Safety
Email:	kjohnson@us.imshealth.com
Presentation type (select one):	Poster

Assessing Data Availability for Pharmacoepidemiology Research in 3 US Healthcare Databases Using the OMOP Common Data Model

**Karin Johnson, PhD¹, Deb Casso, MPH¹, Amanda Anderson, MS¹,
Syd Phillips, MPH¹, Christian Reich, MD, PhD², Urvi Shah², Mui Van Zandt²,
Susan A. Oliveria, ScD, MPH, FISPE¹**

¹IMS Health, Epidemiology and Drug Safety and ²IMS Health Real World Evidence Solutions, Plymouth Meeting, PA, USA

Abstract

IMS Health databases originating from medical claims and electronic health records offer opportunities for conducting pharmacoepidemiology research. The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) provides a framework to assess the availability of variables needed for common research questions. We summarize data availability in US IMS Health databases that have been mapped to the OMOP CDM version 4 (CDM v4). The assessment found that, while single IMS Health databases cover many but not all data elements, they have a large potential cohort size. Depending on the research question, pharmacoepidemiology studies using multiple OMOP domains are feasible by linking databases.

Introduction

Population based pharmacoepidemiology studies rely on information about exposures and outcomes that occur in general populations in clinical practice. Real-world data (e.g., administrative claims and electronic medical records (EMR)) can be used to conduct these studies. However, real-world data sources contain data that were originally collected to support patient care and associated billing. Therefore, it is important to determine the suitability of these data sources for research end use, either individually or linked to other data sources.

IMS Health maintains several real-world databases that have been converted to the OMOP CDM v4, providing an opportunity to evaluate availability of key variables.

Methods

First, a framework was created based on the OMOP domains to define priority data elements for pharmacoepidemiology research studies. 11 of 18 domains of the OMOP CDM were deemed priorities for epidemiology and medical product surveillance.

Second, the US IMS databases that have been converted to the OMOP CDM (Table 1) were systematically reviewed to determine which data sources contained variables of interest in each domain.

Table 1. IMS Health US Databases that have been mapped to OMOP CDM v4

Database	Data Type	Number of Patients
PharMetrics Plus	Medical and Pharmacy Claims	105,009,000
Ambulatory EMR	EMR	37,860,000
Oncology EMR	EMR	2,059,000

Results

PharMetrics Plus contains variables in all of the priority domains for epidemiology and medical product surveillance except for Death (Table 2). However, some individual concepts, for example values for laboratory observations, are not available. PharMetrics Plus also contains variables in the additional domains to support health economics, comparative effectiveness, and quality of care studies. The Ambulatory and Oncology EMRs each contain variables in all priority domains for medical product surveillance except for Procedure Occurrence. Some concepts that are not available in PharMetrics Plus (e.g., lab results) are available in the EMR data sources. Linkage across datasets is possible using a Health Insurance Portability and Accountability Act (HIPAA) compliant deterministic algorithm.

Table 2. OMOP CDM v4 concepts available in IMS Health US Databases

Domain	Database		
	PharMetrics Plus	Ambulatory EMR	Oncology EMR
Priority domains for epidemiology and medical product surveillance			
Cohort	●	●	●
Condition Occurrence	●	●	●
Condition Era	●	●	●
Death	○	●	●
Drug Exposure	●	●	●
Drug Era	●	●	●
Observation	●	●	●
Observation Period	●	●	●
Person	●	●	●
Procedure Occurrence	●	○	○
Visit Occurrence	●	●	●
Additional domains			
Provider	●	●	●
Organization	○	○	●
Care Site	○	●	●
Location	●	●	●
Drug Cost	●	○	○
Procedure Cost	●	○	○
Payer Plan Period	●	○	○

Key:

- All concepts available for this domain
- Some concepts available for this domain
- No concepts available for this domain

Conclusion

Real-world data assets contain data elements necessary for pharmacoepidemiology studies. Individual IMS Health databases cover many but not all data elements and each has a large potential cohort size. Depending on the research question, pharmacoepidemiology studies using all priority OMOP 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 Ambulatory EMR and PharMetrics Plus. Additional variables can be added by linking to other data sources, such as registries that include patient-reported outcomes.