

**The observational databases and population size used in the OMOP experiment.**

Name	Description	Population Size (M)
GE Healthcare (GE)	Derived from data pooled by providers using GE Centricity Office (an ambulatory electronic health record) into a data warehouse in a HIPAA-compliant manner. The database contains a variety of data including diagnoses, orders, and clinical observations spanning 1998 to 2009.	11.2
MarketScan® Research Databases from Thomson Reuters	<i>MarketScan Lab Database (MSLR)</i> Represents privately-insured population, with administrative claims from inpatient, outpatient, and pharmacy services supplemented by laboratory results. Covers 2003 to 2009.	1.5
	<i>MarketScan Medicaid Multi-State Database (MDCD)</i> Contains administrative claims data for Medicaid enrollees from multiple states. Covers 2003 to 2009.	11.1
	<i>MarketScan Medicare Supplemental and Coordination of Benefits Database (MDCR)</i> Captures administrative claims for retirees with Medicare supplemental insurance paid for by employers, including services provided under the Medicare-covered payment, employer-paid portion, and any out-of-pocket expenses. Covers 2003 to 2009.	4.4
	<i>MarketScan Commercial Claims and Encounters (CCAЕ)</i> Captures administrative claims containing privately insured lives with patient-level de-identified data from inpatient and outpatient visits and pharmacy claims of multiple insurance plans. Covers 2003 to 2009.	58
Humana Inc. (HUM)	Contains medical (inpatient, outpatient, and emergency room), pharmacy, and laboratory data (including test results) available back to the year 2002 from Humana’s administrative claims database of over 11 million medical members.	6.5
i3 Drug Safety, Ingenix Normative Health Information	Contains healthcare claims from across the US. Detailed data captured since May 2000 includes more than 50 million lives. For 2008, approximately 14 million individuals with medical and pharmacy benefit coverage were available. Data is available from 1993 to 2009.	1
Partners HealthCare System (PHCS)	Includes data from the PHCS clinical transaction-based data repository as well as inpatient and outpatient billing feeds being collected in the Research Patient Data Registry, an analytic-structured database. The number of patient- concept pairs or clinical “facts” about the patients exceeds 900 million spanning 21 years.	3

Name	Description	Population Size (M)
Regenstrief Institute (RI)	Healthcare data from 1995 is available from the Indiana Network for Patient Care (INPC), a health information infrastructure that was built upon the foundation of the Regenstrief Medical Record System. Records exist for more than 10 million patients, and contains population-based, longitudinal, and structured coded and text data captured from hospitals, physician practices, public health departments, laboratories, radiology centers, pharmacies, pharmacy benefit managers, and payers.	2
SDI Health (SDI_MID)	Contains data from all 50 US states, HIPAA-compliant, de-identified, encrypted patient-level data from hospitals, clinics and physician offices, and retail and specialty pharmacies into a data warehouse with over 160 million unique patients from 1999 to 2009. The majority of this information is captured via standard electronic transmission of claims and transactions.	90
Department of Veterans Affairs Center for Medication Safety VA MedSAFE (VA)	National Patient Care Database consists of the medical and clinical databases for patients treated in the Veterans Health Administration. Data sources include all services provided at the system's medical centers, ambulatory care and community-based outpatient clinics, extended care facilities and several homecare programs. It includes demographic data, information about hospitalizations and outpatient visits, and drug utilization for prescriptions filled within the VA system.	3.2