Attributes	**GEMINI**[1–3]	**Cosmos**[4]	**MIMIC-IV**[5,6]	**CPRD Gold**[7]
Geographic Coverage	Ontario, Canada	United States	Beth Israel Deaconess Medical Center in Boston, USA	United Kingdom
Cohort Size	~2.5 million	310 million	364,627	>21.3 million
Healthcare Setting	All inpatient medical specialties (e.g., general medicine, respirology, oncology, cardiology, etc.) and ICU	Inpatient and outpatient care	Inpatient, emergency department and ICU	Primary care
Data Sources	Clinical and administrative data extracted from participating hospital information systems and EHRs	Epic EHR systems across participating organizations	EHR at Beth Israel Deaconess Medical Center	EHR from General Practitioners
Available Clinical Data Elements (non-exhaustive)	Diagnoses, procedures, labs, vitals, ICU events, radiology reports, clinical notes, transfers, transfusions, medication orders, demographics, administrative data	Diagnoses, medications, patient-generated health data, birth records, vitals, social drivers, specialty-specific data	Microbiology cultures, labs, medications, administrative data, ICU events, discharge summaries radiology reports, demographics	Demographics, symptoms, conditions, referrals, vaccinations, prescriptions, measurements, and laboratory test results
Date range	2015 onwards	Variable	2008-2019	1987 onwards
Linkage capability	Linkage across participating hospitals via health card numbers; linkage with longitudinal population-based data at ICES, project-specific linkages as needed	Patient matching within Cosmos across participating Epic sites via integrated patient records	Internal linkage within hospital via patient and encounter identifiers	Linkage available for patients from practices in England (e.g., secondary care data, disease/treatment-based registries, death registration data, and socioeconomic datasets)

REFERENCES

1. Verma AA, Pasricha SV, Jung HY, et al. Assessing the quality of clinical and administrative data extracted from hospitals: the General Medicine Inpatient Initiative (GEMINI) experience. *Journal of the American Medical Informatics Association*. 2021;28(3):578-587. doi:10.1093/jamia/ocaa225

2. The GEMINI Database. GEMINI. 2026. https://geminimedicine.ca/the-gemini-database/

3. GEMINI Data Dictionary Version 4.0.0. GEMINI Data Dictionary. November 8, 2024. https://geminimedicine.ca/wp-content/uploads/2023/11/GEMINI_Data_Repository_Data_Dictionary_v4.0.0.html

4. Epic Cosmos. Accessed January 29, 2026. https://cosmos.epic.com/

5. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*. 2023;10(1):1. doi:10.1038/s41597-022-01899-x

6. MIMIC. MIMIC-IV. Accessed July 1, 2026. https://mimic.mit.edu/docs/IV/

7. Sanchez-Santos MT, Axson EL, Dedman D, Delmestri A, Data Resource Profile Update: CPRD GOLD, *International Journal of Epidemiology.* 2025:54(4). https://doi.org/10.1093/ije/dyaf077