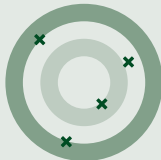


Unlocking precision in drug discovery:

A comparison of real-world data sources

The success of your drug discovery program may depend on the foundational data you're currently using. With different real-world data (RWD) sources available, it can be challenging to determine which source offers the insights, structure, and clinical depth required for critical R&D decisions.

	Lab data	EHR data	Claims data
	Quantitative and qualitative results from diagnostic tests	Digital record of a patient's clinical history, recorded at the point of care	Record of services and medications billed to an insurance company
Lab tests ordered	✓	✓	✓
Lab test results	✓ Structured, raw values	✓ Often stored as attachments or mentioned in physician notes	
Genetic test results	✓ Source of truth for genetic insights	✓ Often stored as attachments or mentioned in physician notes	
Test ordering physician	✓	✓	✓
Diagnosis codes	✓	✓	✓
Demographics	✓	✓	✓
Medications filled			✓
Biometrics		✓	
Physician notes		✓	
Procedures		✓	✓
Clinical granularity	Very high. Delivers precise and objective clinical values	Med/high. Offers detailed clinical notes but often relies on unstructured notes that can be hard to analyze at scale	Low. Contains billing codes for diagnoses and procedures but lacks clinical results
Data availability	Near real-time (typically 24-72 hours)	Real-time to 1 week	30 to 90+ days
Data structure	Highly structured. Organized in a standardized format with distinct fields, making it ideal for large-scale analysis	Hybrid. Includes a mix of structured fields and unstructured text notes, which hold rich detail but may be hard to analyze	Highly structured. Organized in a format based on universal billing codes. Easy to process and query at scale
Interoperability	High. Standardized with universal codes (like LOINC codes)	Low. Often siloed in proprietary systems and contains large amounts of unstructured text	High. Built on universal coding systems (ICD, CPT) for billing
Role in predicting clinical outcomes	High. Objective biomarkers and genetic markers are often the most direct predictors of clinical outcomes	Med/high. Holds potential due to its rich clinical context, but unstructured data can make it challenging for modeling	Low. Lacks the necessary clinical detail
Role in disease diagnosis	 Extremely high clinical accuracy and high precision. Provides the definitive evidence for a disease diagnosis	 High clinical accuracy and moderate/low precision. Provides the comprehensive clinical context	 Low to moderate clinical accuracy and low precision. Provides an administrative signal that a diagnosis was made for billing