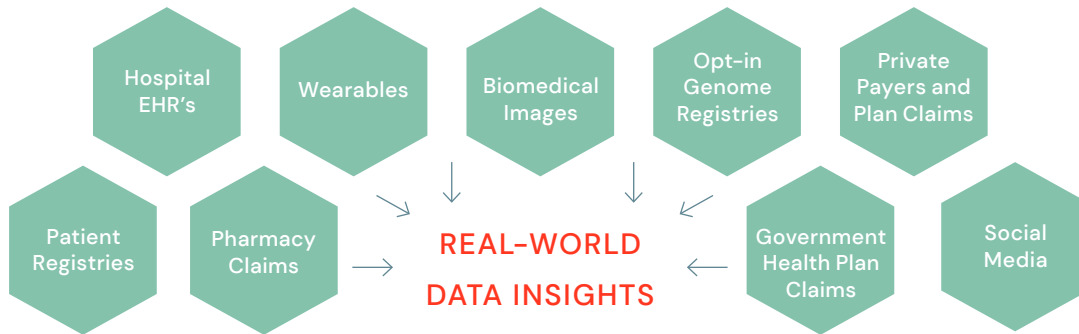# From molecule to market: Delivering on the promise of real-world data

Real-world data (RWD) is revolutionizing how biotech and pharmaceutical organizations develop, deliver and measure the effectiveness of medical treatments in the real-world. Yet, most organizations struggle to transform their large volumes of unstructured RWD into novel insights. Generating real-world evidence (RWE) at scale requires a new approach to data and AI that can only be achieved with a modern data lakehouse.

What is real-world evidence?

Real-world evidence is generated from data collected outside of a traditional clinical trial. This data — called real-world data — is gathered from sources like electronic medical records, claims, and wearable devices. These data sets provide novel insights into the usage, and potential benefits or risks, of a drug.

Hospital EHR's · Wearables · Biomedical Images · Opt-in Genome Registries · Private Payers and Plan Claims · Patient Registries · Pharmacy Claims · **REAL-WORLD DATA INSIGHTS** · Government Health Plan Claims · Social Media

Real-world Evidence Use Cases

Regulatory Approval

Regulators see huge promise in RWE to help accelerate the approval process for new treatments.

Drug Safety

Longitudinal RWE studies can replace or supplement voluntary reporting to provide real-time insight on efficacy or adverse effects.

Trial Design

Drug makers can use RWE to improve the design of clinical trials and fast-track the development of new therapies.

Coverage Decisions

Healthcare companies can use RWE to develop guidelines and support coverage decisions.
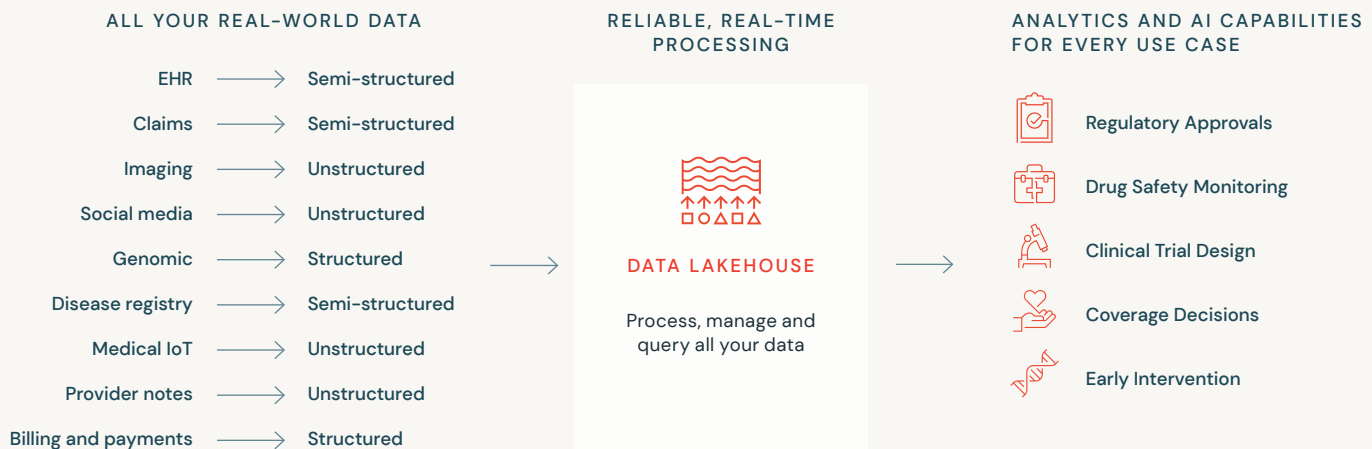
Early Intervention

Integrated risk reports built from diverse sets of RWD help identify patients who are at high risk for chronic conditions.

databricks

Real-world evidence at scale with the Databricks Lakehouse Platform

Bring together all of your real-world data into a single, open and collaborative platform, that supports all of your analytics and AI workloads.

ALL YOUR REAL-WORLD DATA

EHR	→	Semi-structured
Claims	→	Semi-structured
Imaging	→	Unstructured
Social media	→	Unstructured
Genomic	→	Structured
Disease registry	→	Semi-structured
Medical IoT	→	Unstructured
Provider notes	→	Unstructured
Billing and payments	→	Structured

RELIABLE, REAL-TIME PROCESSING

DATA LAKEHOUSE

Process, manage and query all your data

ANALYTICS AND AI CAPABILITIES FOR EVERY USE CASE

- Regulatory Approvals
- Drug Safety Monitoring
- Clinical Trial Design
- Coverage Decisions
- Early Intervention

Challenge	The Databricks Lakehouse Solution
Real-world data is messy A major challenge in generating reliable evidence from real-world data is transforming the data from its raw unstructured form to a proper data model that is optimized for longitudinal studies.	**Ingest and curate all your real-world data** Bring together all your structured and unstructured data in a single platform for data + AI. Databricks' partner solutions make it easy to ingest unstructured real-world data (e.g. NLP for text) and standardize with the OMOP common data model.
Real-world data is massive Legacy on-premises data architectures are complex to manage and costly to scale for today's massive volumes of real-world data, including growth in imaging and genomics.	**Analyze real-world data at scale** Cloud-native and built for speed, the Databricks Lakehouse Platform enables organizations to analyze petabytes of real-world data in near real-time.
Regulatory standards are hard to meet Most organizations have disconnected platforms for data science and data warehouses. This creates challenges when trying to build trust and reproducibility in real-world evidence analytics.	**Build analytics you can trust** Track your data and machine learning models from source through results in a single, secure platform. This enables reproducibility wherever your research teams are located.

Leading life sciences organizations innovate on Databricks

AMGEN

Centralized terabytes of real-world patient data (e.g. EHR, claims, purchased) on Databricks to accelerate clinical trial recruitment with advanced analytics.

SANOFI

Applied machine learning to 17M+ electronic health records to identify new treatment indications for approved therapies while reducing data processing costs by 30%.

Get Started with Databricks' Solution Accelerators for Real-world Evidence

Our Solution Accelerators help your teams get started faster with prepackaged analytics templates that include sample data, prebuilt code, and step-by-step instructions ready to go in a Databricks notebook.

OMOP and Propensity Score Matching

Automate the ingestion of real-world data into your lakehouse and standardize with OMOP for observational analysis at scale.

Generate Oncology RWE with NLP

Extract insights from PDF pathology reports for use cases ranging from measuring disease progression to exploring optimal treatment pathways.

Detect Adverse Drug Events with NLP

Improve drug safety monitoring by extracting adverse drug events from health text data and building visualizations to show the frequency of correlations.

Databricks customers in Healthcare and Life Sciences

AstraZeneca   BAYER   Biogen   CVSHealth   Cerner   gsk GlaxoSmithKline   Providence St.Joseph Health

novo nordisk   SANOFI   Eisai   SANFORD HEALTH   Genome Institute of Singapore GIS   MERCK INVENTING FOR LIFE   REGENERON

databricks

Questions? Learn more at **dbricks.co/lifesciences** and **dbricks.co/healthcare**