



Data Scientist

AI Data Science Machine Learning

# Analyzing Real-World Evidence at Scale

Learn how to accelerate drug development and improve healthcare delivery with a unified approach to data, analytics and AI



Rockwall  
Analytics

DATA & ANALYTICS ADVISORY

# Background

Big data has proven as revolutionary to medicine as scientific breakthroughs like the microscope and x-ray. By analyzing large, population-scale datasets, researchers and clinicians can now spot trends that were previously not visible in smaller studies. As such, many healthcare organizations are aggregating data across ever-increasing patient populations and data sources including data collected outside of clinical trials. This introduction of real-world data (RWD) to medical research promises a new era of deeper insights, more powerful treatments, and better quality of care.

RWD ranges from electronic health records (EHR), telemetry data from wearables, health insurance claims, biomedical images, disease registries, social media, and even online forums where patients share experiences. Pharmaceutical companies, health insurers and regulators are investing heavily in methods of analyzing RWD in order to derive real-world evidence (RWE) and accelerate scientific research.

**By applying advanced analytics and machine learning to RWD, healthcare and life sciences organizations can deliver on a number of innovative use cases, including:**



## VALUE-BASED CARE

Blending claims with EHR data provides insights into the value and effectiveness of providers or treatment protocols.



## EARLY INTERVENTION

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



## REGULATORY APPROVAL

Regulators see huge promise in RWD to help accelerate the approval process for new and safer drugs.



## PHARMA INNOVATION

Drug makers can use RWE to improve the design of clinical studies and fast track the development of new therapeutics.



## PHARMACOVIGILANCE

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

# Challenges Analyzing RWE at Scale

Drug discovery can take years, and cost billions of dollars. But when researchers tap into the wealth of real-world data contained in medical insurance claims and electronic health records, new worlds of possibility open up for pharmaceutical innovation.

The efficacy of clinical trials can be resolved by analyzing EHR data to optimize enrollment criteria and trial sites. New patterns, previously not visible in smaller sample sizes, can be unearthed to support potential new indications for medicines. From development to approval and beyond, real-world data is revolutionizing the industry.

Despite the promise of real-world evidence and the massive investments being made to build and purchase data assets, big data hurdles are blocking the path forward. In fact, fewer than half of healthcare and life sciences companies report having the capabilities necessary to fully leverage their real-world evidence.

**There are three primary obstacles limiting the ability of healthcare and life sciences organizations to adopt and scale RWE.**

## Inability to Scale to Support Massive Data Sets

While the opportunity to drive innovation in healthcare is massive, obstacles exist. Data scientists, engineers, and bioinformatics teams often struggle to integrate, analyze, and model diverse, often siloed sets of structured and unstructured data. Teams find themselves handicapped by legacy on-premise infrastructure that is difficult to manage, unequipped for interactive exploration of terabytes-to-petabytes of real-world data, and lacking the advanced analytics tools needed to run predictive, machine-learning models at scale.



### STATE OF RWE IN HEALTHCARE AND LIFE SCIENCES



**6.4** terabytes of clinical, imaging, and genomics data produced by an individual over a lifetime\*



**90%** of companies are building capabilities to conduct RWE studies\*



**60%** of companies are using machine learning and 95% plan to use it in the future\*



**Less than Half** of companies have mature enough capabilities to fully leverage RWE \*

\* Source: <https://www2.deloitte.com/us/en/insights/industry/life-sciences/2018-real-world-evidence-benchmarking.html?id=us:2pm:3ad:rwe2018:eng:chs:103018:stat2>

# Challenges Analyzing RWE at Scale

Of course, the introduction of new data sources into an already data-heavy industry such as pharma brings new challenges. Data from outside clinical trials are usually siloed, heterogenous, and unstructured. Deriving clinically valid insights from this type of data requires infrastructure that allows researchers to process and analyze diverse sets of data at scale, while maintaining data integrity to support regulatory and scientific validity.



# Challenges Analyzing RWE at Scale

## Infrastructure Complexity Slows Agility and Innovation

Many organizations use legacy analytics platforms to analyze real-world evidence data. However, modern datasets are becoming too large for these platforms to process efficiently, stymieing biomedical innovation. Shared enterprise data platforms also result in bottlenecks, as teams jockey for compute time and storage upgrades.

This makes developing and deploying pipelines labor-intensive and time-consuming, and makes scaling analytics to current data volumes infeasible. Loading data may fail, resulting in weeks lost to data setup. A basic set of queries can take days or weeks to execute, and advanced analytics, such as cross-cohort analyses and machine learning, are simply not possible.

## Poor Team Collaboration Hurts Data Science Productivity

Finally, like their counterparts in other industries, disparate teams of researchers and analysts that come together to glean real-world evidence from real-world data often find themselves working in silos, which stifles progress. Analysts and data scientists explore limited data sets and conduct experiments on laptops, making it hard to share analyses, combine data sources, and audit completed analyses. The inability to reuse code and share insights can put a damper on operational efficiency, morale, and, ultimately, innovation.

### Is your data ready for RWE at Scale?

This checklist from scientists at Flatiron Health will help you make sure your data sets are regulatory-grade and capable of yielding scientific evidence.

#### HIGH QUALITY

Know where your data comes from and where it's been.

#### COMPLETE

Both structured and unstructured data need predefined rules for abstraction and benchmarked against trusted standards.

#### TRANSPARENT

Clear definitions for study goals and cohort selection are essential for gathering scientifically valid RWE.

#### GENERALIZABLE

Always identify biases, like geographic representation, in data sets to allow for statistical adjustments and clinical interpretations.

#### TIMELY

Because RWE tracks decisions made in real-time, real-world data must be recent and timestamped.

#### SCALABLE

With an ever-growing number of patients and variables, RWE calls for infrastructure that can manage and analyze massive amounts of data.

