MarketScan

Real-world data: The growth guide for biotech and pharma

merative

Navigating the complexities of drug development and market entry is a daunting task. Fortunately, research teams can leverage new innovations within real-world evidence (RWE) to significantly increase success. But harnessing its power is no easy task. To generate high-quality, comprehensive evidence, you first need to build a strong research foundation by carefully selecting the best real-world data (RWD) vendor for your needs. While RWD can be used for a variety of analyses, this guide will focus primarily on the studies requiring the most rigor: Health Economics & Outcomes Research (HEOR), epidemiology, and regulatory. So where do you start and what factors should be considered?

This detailed guide breaks down the top characteristics your RWD needs to conduct research on the U.S. population and helps answer four key questions:

- What makes a database a research-grade database?
- How can I enrich my existing data?
- How can I empower my research teams?
- What should I consider when outsourcing RWE generation?

Study types	Example of applications	Example of stakeholders
Regulatory Obtain regulatory approvals and satisfy requirements including post-authorization studies and new drug applications	– Post-authorization safety studies – Mother-infant linkages – Synthetic control arms – Label expansions	**Regulatory agencies** Inform clinical trials strategy and response to regulatory requirements
HEOR Generate evidence on the value and efficacy of specific treatments and intervention	– Comparative effectiveness – Burden of illness – Treatment patterns & sequences (adherence, persistence, time to treatment)	**Payers** Obtain and optimize formulary status, and maximize reimbursements
Epidemiology Understand real-world disease prevalence and management practices supporting product commercialization	– Natural history of disease & patient journeys – Prevalence and incidence – Time to event analyses – Surveillance (adverse events, infection rates)	**Clinicians** Target patient populations that will benefit most from products

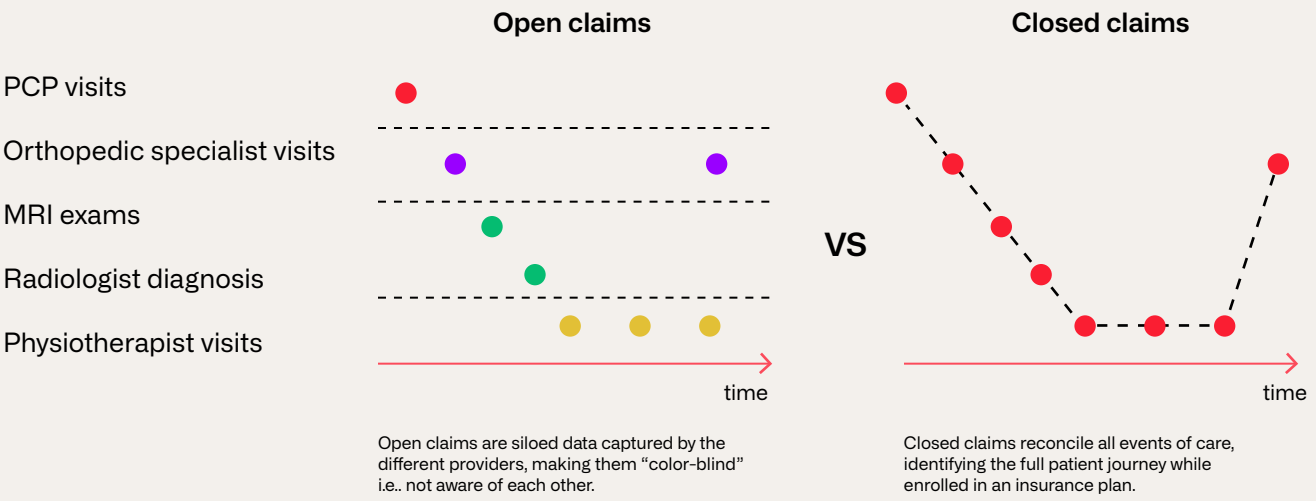
What makes a database a research-grade database?

Generating robust RWE requires rigorous analysis of research-grade RWD. When determining the value of research-grade datasets, we recommend evaluating sources based on these four critical factors: data completeness, actual healthcare cost, data representativeness, and longitudinal patient-view.

Data completeness

When evaluating RWD sources, it's important to select options that provide a complete and holistic view of the patient to help advance studies and better inform decisions. Traditional open claims data are usually generated by health provider systems, and are often fragmented, riddled with duplicates, and missing events. Closed claims data on the other hand, captures all events and settings of care processed by a patient's insurance and filters out duplicates. This ability to precisely track events provides a more complete and accurate account of the patient's journey and is what sets closed claims data apart.

The following diagram represents the patient journey of someone experiencing muscular shoulder pain after a bad fall.



When evaluating closed claims data sources, it's important to understand the concept of adjudication. Adjudication is what underscores the meticulous quality control applied to closed claims data, but not all closed claims are fully-adjudicated. When they are, it denotes that they've followed an exhaustive, multi-month-long process to further ensure the accuracy of the delivered care and costs reported to insurance. This process instills a higher degree of confidence in data quality, further reinforcing the rigor and appeal of closed claims data.

Looking ahead, fully-adjudicated closed claims datasets are seen as the linchpin for successful Regulatory studies and robust HEOR and Epidemiology studies.

It's important to note that some vendors may attempt to replicate closed claims data by combining multiple open claims data sources. This piecemeal approach, while potentially reducing duplicates, can't ensure the complete capture of all claims and may miss vital records, reducing the credibility of studies.

Actual healthcare cost

Actual costs, often referred to as real or true costs, represent the real financial payments for each event of care. In simpler terms, these costs reflect the tangible amount of money that changes hands, including granular details like co-pays and out-of-pocket patient expenses. When conducting HEOR studies, actual costs are indispensable and provide an authentic, tangible measure of healthcare payments.

In an industry dominated by imputed costs, which are merely statistical estimates, true costs lend precision to the data that's impossible to achieve with imputed costs. And, when actual costs are coupled with strong, representative data, they enable more robust and conclusive evidence for HEOR studies.

Lastly, while actual costs don't usually have significant impact on regulatory and epidemiology studies, they still provide valuable insights into specific events of care. For example, in treatment pattern studies, real costs can help with understanding certain details about a particular treatment or procedure based on the payments made.

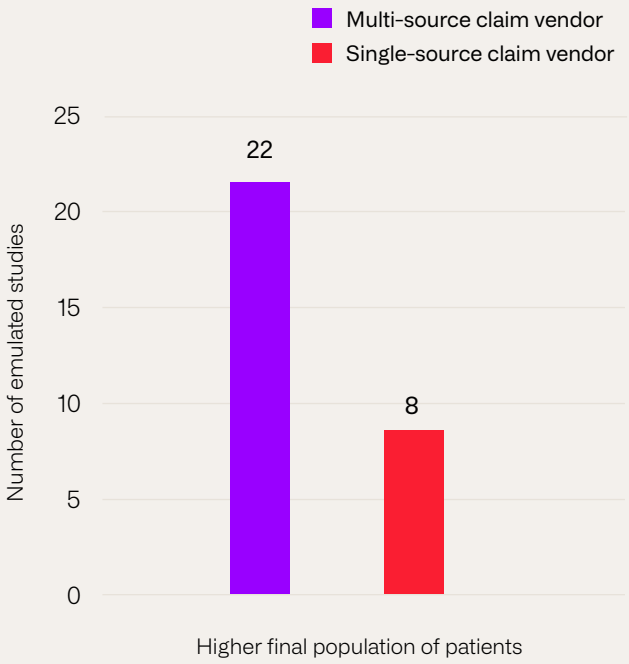
Data representativeness

A representative dataset collects several diverse demographic variables from a wide range of trusted sources – leading to more comprehensive and reliable insights. It sounds simple, but often it goes beyond just capturing a diverse population. For example, in the context of the United States, it can be easy to miss important factors that are required to achieve nationally representative data.

The first and most obvious factor is capturing a broad demographic representation, including members across all 50 states, from multiple age groups, and other socio-economic characteristics. The second and less intuitive factor, is attributed to the complexities of the U.S. healthcare system. Contrary to most developed countries, the U.S. government doesn't provide universal healthcare. Instead, most of the U.S. population receives health insurance through their employers, resulting in hundreds of different health plans and policies impacting care decisions. Therefore, data aiming to be representative of the U.S. must also capture data across a variety of health plans, policies, and insurance firms.

Prioritizing diversity in data collection minimizes the risk of bias and provides researchers with additional variables to better understand your population health trends. Having access to a more holistic picture of healthcare drastically enhances the quality of research outcomes. For example, when conducting HEOR studies, the ability to combine actual cost data with diverse, representative data leads to more robust studies. For epidemiology studies, representativeness is critical in collecting accurate measures of incidence and prevalence rates.

To have representation, you need a large population and a good distribution. Looking at cohort size, and analyzing the studies used for RCT emulation, a multi-source closed claim vendor had a larger final patient cohort in 73% of studies.



Source: Emulation of randomized clinical trials with non-randomized database analyses; Results of 32 clinical trials. The Journal of the American Medical Association - April 2023, DOI: 10.1001/jama.2023.4221

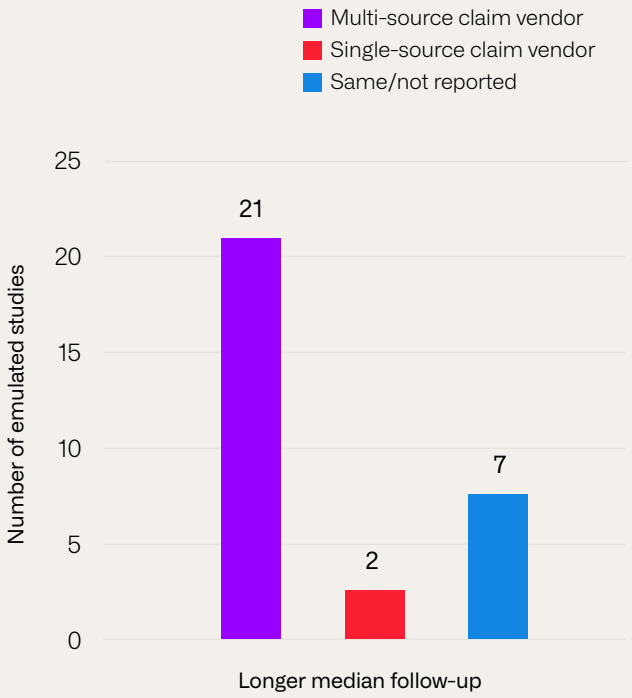
Longitudinal patient-view

For research teams, the sky is the limit with longitudinal data. Longitudinal patient view is the ability to track and examine a specific patient's healthcare journey over a long period of time. It's a critical characteristic that helps researchers better understand patient health status, medical history, treatment episodes, disease progression, and healthcare outcomes.

For HEOR and certain epidemiology studies, researchers often need patients with at least two consecutive years of history. Although in practice, a five-year data window may be requested to allow for the accumulation of study patients and/or allow for longer follow-up periods.

For regulatory studies, teams can use the exhaustive patient records to safely evaluate the impact of exposure to a new drug by making comparisons across large, diverse cohorts. But, what happens when data isn't strong enough to provide a longitudinal patient-view? Cohorts size drops fast! While you may start with a large pool of candidates, as the seemingly qualified patient records are excluded, there is typically a dramatic attrition in the final pool.

Analyzing the studies used for RCT emulation, a multi-source closed claim vendor had the same or longer patient follow-up in 93% of the studies.



Source: Emulation of randomized clinical trials with non-randomized database analyses; Results of 32 clinical trials. The Journal of the American Medical Association - April 2023, DOI:10.1001/jama.2023.4221

"The longitudinal data enabled us to follow patients for 5 years, demonstrating potentially long-term consequences that weren't previously identified."

Division Chief for Biomedical Informatics &
Internal Medicine Department faculty member

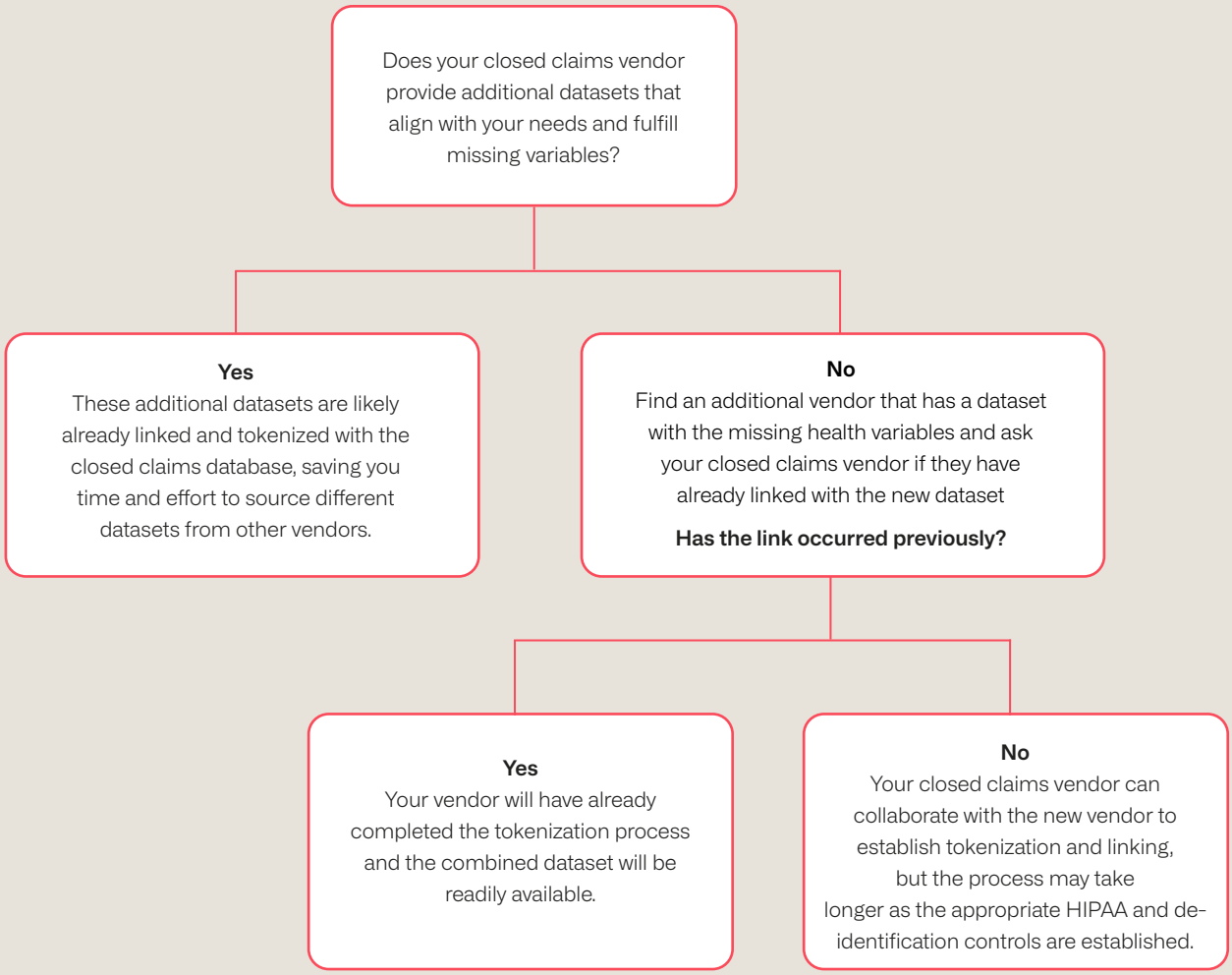
Explore case studies

How can I enrich my existing data?

Once you have your foundational research layer, you can provide your researchers with the means to run studies and compute precise metrics. But the work doesn't always stop there. Your data may still lack specific variables needed to create your relevant cohorts, resulting in a quest for more data. Accomplishing this step requires your data vendor to make connections across the different data sources, a capability known as 'linking.'

However, this aggregation of data may inadvertently reveal some protected health information (PHI), thus risking patient identification. In accordance with Health Insurance Portability and Accountability Act (HIPAA) regulations, data vendors are mandated to sell only de-identified data, even when linking datasets. Consequently, your vendor should provide the capability to tokenize and de-identify data. Tokenization pertains to the creation of an encrypted identifier or 'token', guaranteeing that patient data will remain de-identified.

While linking to your foundational layer, consider the following:



Regardless, whenever you're assessing multiple databases, ensure you determine the quality of the patient pool that you end up with, such as data completeness, meets your research needs. Fortunately, by choosing a robust foundational research layer, you've already taken steps to enhance database overlap.

How can I empower my research teams?

Getting good real-world data is half the battle to getting to real-world evidence. The other half is making sure researchers have the resources and data they need to run comprehensive analysis. However, real-world databases are large and require a lot of storage, making it a challenge for organizations to maintain the IT infrastructure internally. Fortunately, the advent of cloud technologies has helped reduce these constraints, democratizing RWD access to small and large firms, regardless of their internal IT capabilities.

Enhance access and collaboration

Cloud technology helps healthcare organizations access and share information easily. Multiple users can access data simultaneously across different locations and interfaces, helping streamline research projects and optimize clinical decision-making.

Provide data with speed

Data files are large and managing different sources and formats can get complex. Cloud-based RWD helps make data available for research immediately, reducing the time and money spent on manually deploying data and managing multiple versions.

Perform analyses with ease

Some cloud vendors provide easy-to-use analytics, helping researchers of various coding skill levels glean the insights they need to advance studies. For researchers using advanced tools, cloud technology offers seamless integrations to a variety of analytic applications such as Tableau, Power BI, and other SQL compatible tools helping advance initial findings.

Use a scalable, cost-effective model

The elasticity of the cloud allows teams to scale their IT requirements as needed to meet fluctuating demands. This feature, coupled with a pay-as-you-go model, makes cloud services a cost-effective option, eliminating the hefty upfront costs of traditional on-premise infrastructures.

What should I consider when outsourcing the generation of RWE?

When resources are tight or studies become more complex, it's not unusual for teams to partner with consultant services on RWE generation. Consultants can help with a wide variety of projects, but often teams outsource work when they need to focus their energy elsewhere, lack the necessary in-house expertise, or only need the results of an isolated study and don't need the underlying data. Regardless of the reasons, it's important to consider the following factors when searching for external support.

Level of expertise

Conduct an exhaustive review of the researchers' academic credentials and practical qualifications in fields like epidemiology, statistics, health economics, and other relevant disciplines to determine the proficiency of the team. It's important to look for an organization that can offer a diverse, multidisciplinary team, bringing a wider perspective and higher value.

Past experiences

Look for teams that have experience with handling RWD and evidence, especially in the therapeutic areas (TAs) of interest (oncology, cardiovascular disease, etc). TAs have their own unique intricacies and complexities, so it's important to find teams that have prior experience working on those areas. Similarly, consider the firm's ability to handle diverse types of real-world data and transform it into meaningful evidence. Consider requesting some case studies and examples of successful projects that are similar to your objectives to learn more about their past work

Professional handling of healthcare idiosyncrasies

Though it may seem obvious, choosing a consulting firm that comprehends the complexities of the U.S. healthcare system is vital. Given that healthcare procedures, protocols, and policies differ across countries, disease and impact proxies need to be defined based on the practical workings of the healthcare system. Likewise, consultants should possess an understanding of regulatory policies, both domestically and internationally, that govern the use of RWE in drug approval or health assessment.

Willingness to partner

Skills and results are among the biggest deciding factors, but the cultural fit and working style is equally as important. Ensure that they can translate intricate data analyses into clear, actionable insights that align with your business goals. And make sure that they aren't just going to communicate results back to you, but that they also help convey results to your internal and external stakeholders, such as regulatory agencies. Outsourcing RWE generation isn't just about freeing up resources, it's about finding the right partner for the job that will bring the most meaningful insights to life.

Inside knowledge

There can be advantages to partnering with a data vendor that also offers consulting services. Such a firm will have an intimate understanding of their own data, potentially resulting in stronger evidence and faster delivery times. Sometimes, there are certain insights associated with building a real-world database that only the creator possesses and can access.

In conclusion

This guide was designed to help your organization leverage the full potential of real-world evidence effectively. When pursuing rigorous outcomes for projects like HEOR, epidemiology studies, and regulatory approvals, it is essential to build a solid research foundation. To do so, you need to meticulously review real-world data vendors to select one that best fits your needs, understanding the attributes of a research-grade database, and strategies to enrich your data through linking. The work doesn't stop there. It's equally as important to equip your researchers with the right tools, or alternatively, to engage the most suitable consultants, to attain the valuable real-world evidence you seek.

MarketScan is a premium portfolio of data, services, and tools designed to help you strengthen, accelerate, and simplify your real-world evidence studies. With a unique employer-sourced model, the MarketScan research-grade datasets offer an exceptional level of completeness and representativeness, along with actual healthcare cost and an extensive longitudinal-patient view. Cited by more than 3,500 peer-reviewed publications and an enduring reputation of high-precision claims, MarketScan is the foundation layer for rigorous Health Economics and Outcomes Research (HEOR), epidemiology studies, and regulatory submissions.

Explore MarketScan

About MarketScan

MarketScan by Merative provides deidentified, longitudinal, patient-level closed claims and specialty data for 293M+ patients sourced directly from a diverse pool of payers. Industry-leading researchers rely on MarketScan to derive valuable insights pertaining to health economics and outcomes research, treatment patterns, and disease progression across the industry resulting in more than 3,500 peer-reviewed manuscripts.

Learn more at
merative.com/real-world-evidence

About Merative

Merative is a data, analytics and technology partner for the health industry, including providers, health plans, employers, life sciences companies and governments. With trusted technology and human expertise, Merative works with clients to drive real progress. Merative helps clients orient information and insights around the people they serve to improve decision-making and performance. Merative, formerly IBM Watson Health, became a new standalone company as part of Francisco Partners in 2022.

Learn more at www.merative.com

Produced in the United States of America
July 2023

MSN-4876677816 Rev 3.0