

5 Key Indicators of Real-World Data Quality

How to Evaluate the Quality of Real-World Data

The availability and use of real-world data (RWD) are growing rapidly across healthcare with more vendors, additional data types, and an overall higher volume of data. The proliferation of RWD offerings means more options to underpin your use cases. But it also brings some concerns around the reliability and consistency of broad market data sets.

One of those complications stems from the 21st Century Cures Act, signed in 2016, which aims to speed up the development of medical products and make innovative treatments more accessible to patients. One key provision of the act calls for the FDA to provide guidance on using real-world evidence (RWE) in regulatory decisions to ensure data quality and traceability. Traceability indicates that the results can be traced back to the original source data.

Historically when people searched for RWD, they either needed to choose between highly curated data from small samples (e.g., academic medical centers), or broad data sets that covered lots of patients, but the consistency and origin of the data were sometimes questionable or uncertain. The increased need for robust RWE has shifted customers to focus on both the quality and consistency of data.

In this blog, we'll discuss the importance of thinking about the depth of data capture and highlight what to look for the next time you are evaluating a RWD asset to power your RWE generation.

How to Assess Data Quality

Are you achieving the results you planned with your RWE strategy? If you're like most company leaders, when you think about ways to improve your RWD or RWE strategy, you may focus on how to do more with your data, such as invest in more platforms, partner with top analytics companies, or hire great data scientists. While these are important considerations, have you given enough thought to the quality of your foundational data sets?

Your team makes critical decisions based on your RWD and generates analytics and RWE to submit for regulatory approvals, gain market access, determine appropriate pricing, and measure utilization. Critical decisions and outputs require the best quality data and not all data sets are created equal.

Here is how you can assess the quality of the RWD you can license:

Is the data traceable?

Many RWD data sets are assembled from a variety of heterogeneous sources, and many vendors are not able to track the original source of the data, nor go back to data contributors to address questions about the data's accuracy or completeness. As RWD gets both more focus and scrutiny, the ability to trace the lineage of data back to the original source is becoming a market standard.

Blue Health Intelligence® (BHI®) maintains relationships with all of our data contributors and can work with our customers to ensure they fully understand the content of the data and advise on the best ways to use the data to support RWE objectives.

What is the degree of missingness in fields that matter to you?

How many fields or concepts are included for each patient? Do you receive information about patients' insurance coverage, financial responsibility, or participation in value-based care programs?

BHI takes pride in the exceptional completeness of its data, with fill rates surpassing 99% for crucial fields, including allowed amounts, National Provider Identifiers (NPIs), and multiple diagnoses. By maintaining near-perfect fill rates, BHI ensures that no critical details are missing, enabling comprehensive insights and informed decision-making.

How consistent and harmonized is the data?

Even if the data is complete, is it straight from the source or is it derived? If your vendor combined multiple data sets itself, does it use the same terminologies and logic? Can you trust that the data is completely accurate?

We take a standardized approach to data management by implementing a unified data model that is consistently applied across all contributors. This ensures that the data collected from various sources is structured and organized in a uniform manner.

Furthermore, we maintain a rigorous quality assurance process to validate the accuracy and reliability of the data. This process has been certified by Milliman®, using a comprehensive four-stage certification methodology that tracks field accuracy, and data completeness.

Does the data reflect a long enough time period to show the changes you need to see?

According to a study published in the Journal of the American Medical Association, 20% of commercially insured people change health plans each year. For many analyses, you need multiple years of data to understand things like long-term costs and utilization. How you create a cohort for analysis often depends on understanding their history. Maybe you want to focus on members who had at least three years between the initial diagnosis and their

first treatment. You can't find the right patients if you don't have enough historical data for them.

BHI's database encompasses a significant duration of continuous enrollment, specifically exceeding three years for more than 50% of members and surpassing five years for over 40% of members. We also maintain a master member index (MMI) across all of our data contributors to be able to track members as they move across health plans. This extended period of continuous enrollment ensures a substantial and reliable data set, allowing for a comprehensive view of patient information and trends over an extended timeframe.

Does the data reflect the population as a whole and the groups you want to see?

Ensuring the representativeness of the study population is vital when presenting RWD-based analyses to both internal stakeholders and external parties.

Do you have the ability to segment by plan design or formulary? Do you know when a benefit type is carved out of the benefit design?

With BHI you can rely on a robust data set sourced from 29 unique organizations, encompassing over 300 distinct plan designs that cover 100% of ZIP codes in the nation. This extensive coverage allows for the segmentation of data based on various plan designs and formularies, providing a more comprehensive and nuanced understanding of the study objectives, and enhancing the validity and applicability of the analyses conducted.

Shortcomings in these five areas are why many data vendors focus on the number of patient records you can get for the money. They don't want you to peek under the hood for the quality of what you're getting for those patients, how accurate the data is, and how likely that the insights will meet regulatory guidelines. Quality data not only helps meet your RWD needs but saves time and money by ensuring the success of your project.

What Does Good Data Look Like?

Here are five key indicators of quality in healthcare data sets you should look for:

- 1. True Adjudicated Financial Values, Not Proxies**

Will you get the specific data elements you want for each patient rather than having to use a proxy?

- 2. High Fill Rates**

This means you can make use of every data point you pay for.

- 3. Standardization**

The data is comprised using one standardized format. This means that the fields will be consistently defined.

4. **Longitudinal**

Data that reflects several years of a member's medical history enables you to not only analyze care patterns over long periods of time but also identify comparable cohorts for benchmarking, rather than simply finding members with similar point-in-time diagnoses.

5. **Trustworthiness**

For the market to rely on the evidence created from your RWD it needs to have confidence in the underlying data – including its validity, appropriateness, and traceability.

Build Success for the Future

As you think about RWD assets, ask yourself whether your company's data is of high enough quality to meet its needs. If not, perhaps it's time to shop around to see how to improve your RWD/RWE strategy with the best underlying foundation, the data.