WHITE PAPER

A Guide to Navigating Today's Real World Data Landscape

PRESENTED BY:

PicnicHealth

PUBLISHED BY:

FIERCE Pharma

2023

PRESENTED BY: PicnicHealth

PUBLISHED BY: FIERCE Pharma

Contents

The Expanding Role of Real-World Data and Evidence

Historically, the life sciences industry has leveraged a limited set of real-world data (RWD) and real-world evidence (RWE) options to support a narrower scope of typically enabling access and post-marketing support. Recently, the landscape has changed. The industry now benefits from a growing number of RWD sources – from structured electronic health records (EHR) and claims, to narrative clinical data, patient reported outcomes (PROs), and device data – coupled with increasingly sophisticated technology to harmonize and analyze RWD. Greater acceptance of RWD in trial, approval, and post-launch monitoring processes has led life science organizations to explore additional use cases for RWD further upstream in the clinical development lifecycle. This acceptance can be seen in recent regulatory approvals: 90% of all new drug approvals in the United States in 2020 included RWE as part of their submissions.[1]

However, not all data is created equal. The ever-increasing number of RWD sources is certainly a blessing, but incomplete or even wrong data can cause costly yet preventable delays to a company's regulatory and access strategies. To meet these critical business needs, it is imperative for organizations to ensure that data is fit-for-purpose and can paint a complete picture of the patient journey.

Navigating the New Era of Patient-Centric RWD and RWE

A deeper understanding of individual health journeys represents a paradigm shift for the industry. To unearth these insights, research teams must obtain critically important but traditionally hard to capture data that is highly siloed in individual healthcare systems, non-coded (i.e., key information only available in narrative text) or requires access to patients themselves (i.e., PROs). This is only possible with patient-centric approaches to RWD collection that leverage the patients' unique role as the only through line in their journey through the healthcare system. To that end, organizations must be mindful of the evolving terminology that surrounds patient-centricity, what it means in the context of RWD collection, and where it may be missing in traditional data models. (See sidebar, **Understanding Patient-Focused Terminology**.)

1. The Role of Real-World Evidence in FDA-Approved New Drug and Biologics License Applications. Clinical Pharmacology & Therapeutics. November 2, 2021. Understanding Patient-Focused Terminology

The ever-increasing number of RWD sources is certainly a blessing, but incomplete or even wrong data can cause costly yet preventable delays to a company's regulatory and access strategies.

Understanding Patient-Focused Terminology

Patient-mediated refers to a model in which patients own and have governance over their data. Organizations must obtain patients' consent to use their data.

Patient-generated is a model under which consent has been obtained and organizations are now able to obtain data directly from patients, primarily in the form of PROs or medical/wearable device data.

Patient-centric is an end-to-end process for considering the patients' role in how data is collected and used. Patient-centric practices may include but are not limited to the following:

- Expand the ability for more patients to participate in research by engaging patients within their communities instead of requiring them to come to a physical trial location, as 70% of Americans live more than two hours from a trial site.[2]
- Listen to the patient voice and reflect what is important to patients, such as quality of life or non-clinical health and wellness goals.
- Reduce the undue burdens placed upon patients enrolled in RCTs or post-trial studies while also undergoing treatment for rare or chronic diseases.
- Develop a seamless experience using a portal or other data-sharing platform so patients do not need to enter data manually and/or attach it in emails to medical affairs teams.
- Provide patients' data back to them, reflecting their role as a valued participant in the trial and care process and improving their understanding of their condition.

According to ISPOR, the professional society for health economics and outcomes research, patient-centered research is defined as "The active, meaningful, and collaborative interaction between patients and researchers across all stages of the research process, where research decision making is guided by patients' contributions as partners, recognizing their specific experiences, values, and expertise."[3]

2. Sanofi launches new virtual trials offering with Science 37. Fierce Biotech. March 2, 2017.
3. Patient-Centered Research. ISPOR. 2023.

Patient-centric data is essential when your research questions require: truly longitudinal data, access to individual patient insights and assessments not captured routinely in clinical care, and access to deep clinical phenotypes and outcomes that cannot be determined from coded data alone.

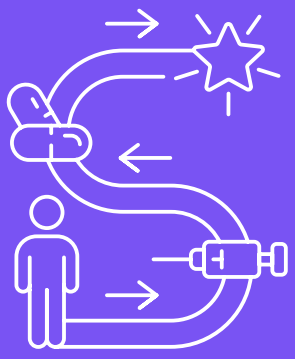
Patient-centric data is essential when your research questions require: truly longitudinal data, access to individual patient insights and assessments not captured routinely in clinical care, and access to deep clinical phenotypes and outcomes that cannot be determined from coded data alone. There are three considerations when determining if a real-world dataset meets your needs.

Data Depth

Deep clinical data means that a dataset provides a 360-degree longitudinal view of the patient journey with access to complete EHR data including the narrative text of provider notes and reports.

Knowing whether a dataset is truly deep requires asking several key questions:

- ***Does it include all patient visits of interest?*** A dataset should be site-agnostic (not relying on a single EHR or other standalone clinical system), should be updated continuously (to support analysis within reliable, defined observation windows), and should be longitudinal (to enable longer observation windows and perspective on additional outcomes).
- ***Is the population representative?*** The FDA and patients have rightfully raised concerns about improving representation. A dataset that does not reflect the relevant patient population raises questions about generalizability of findings and may impact approval and access.
- ***Do researchers have access to unstructured data?*** Many clinical insights remain in the narrative text of clinical notes and are not well represented in diagnostic coding, such as disease subtypes, markers of disease progression, or reasons for treatment switch. This information is difficult to structure and cleanse, but it is vital to understanding how a disease impacts patients and how treatments may or may not be working.
- ***Does it include multimodal data sources?*** Along with PROs, organizations benefit from access to imaging, medical and pharmacy claims, and medical device data to complement the data readily available in EHRs by more accurately illustrating resource utilization, for example. It is important to note that some data sources may need to be linked via tokenization, a strategy that comes with benefits and drawbacks (see sidebar, **The Role of Tokenization in Building RWD Sets** on page 7).



Deep clinical data means that a dataset provides a 360-degree longitudinal view of the patient journey with access to complete EHR data including the narrative text of provider notes and reports.

Data Source Characteristics

Answering the following questions can help an organization determine where data comes from, and whether it can support fit-for-purpose research.

- ***How was it sourced?*** Data sourced from a single system is rarely fully complete or longitudinal. That makes it important to consider how well the data supports an organization's specific needs and complements whatever data is already in hand. Be sure to evaluate the trustworthiness of the data set based on factors such as what variables are missing and what inherent biases may be present.
- ***What variables are available?*** The most relevant data points vary significantly between diseases. In addition, individual diseases tend to be multifaceted, as one patient may present very different symptoms than another patient. Finally, certain signs and symptoms - such as fatigue, shortness of breath, and elevated blood pressure - may be associated with any number of conditions. Looking at a limited number of variables in turn limits the research questions an organization can address, could drive the organization to incorrect answers, and limits the benefits that patients receive from research.

Tunability & Scalability

Partners engaged in complex research efforts recognize that data needs vary between different use cases. In order to ensure that data is fit-for-purpose to address a specific research need, a data platform must be flexible, scalable and tunable. This requires use of the latest technology, including machine learning (ML) to structure data, but also requires the ability to leverage expert human abstractors to further validate model output when the use case requires it. In some high leverage use cases, including regulatory submissions, the additional transparency and validation provided by human review is warranted. Fine-tuning large language models (LLMs) is a particularly promising ML approach, layering expertise in medical data structuring on top of the natural language understanding trained into base LLMs. A data platform source that can be tuned to specific use cases, including the ability to process data through a combination of machine learning and human curation is more flexible and is more likely to empower a broad range of an organization's research use cases.

Partners engaged in complex research efforts recognize that data needs vary between different use cases. In order to ensure that data is fit-for-purpose to address a specific research need, a data platform must be flexible, scalable and tunable.

The Role of Tokenization in Building RWD Sets

While there is much excitement around the use of tokenization to protect against patient reidentification when using RWD, there are important limitations to take into account:

- Matching is generally probabilistic, meaning that you cannot guarantee true matches between data sources.
- Tokenization is typically applied to large but shallow and therefore incomplete data sets. As a result of these data gaps, there may not be enough data fields to match patients.
- In cases where matches can be made, the temporal coverage may be too short to warrant meaningful comparisons. This is a critical consideration for rare diseases that can take several years (or more) to properly diagnose.
- These shortcomings are compounded when attempting to make matches across three or more data sets, further limiting the value of the cost-prohibitive tokenization process for conducting meaningful patient-centric research.

Exploring the Patient-Centered Approach Through RWD Use Cases

PicnicHealth utilizes patient-centric processes to collect, abstract, and link regulatory-grade RWD and develops data models that enable life science organizations to generate evidence throughout clinical development and commercialization. Additionally, with the PicnicHealth platform, patients can regain control of their health through better insight into their medical journey.

Case Study: Disease Burden And Unmet Need

PicnicHealth enrolled over 450 patients with a rare blood disorder and collected longitudinal EMR, labs, and patient reported outcomes to better understand disease burden and unmet need with current management options. For all participants, key outcomes were captured from longitudinal EMR, labs, and patient surveys, including subtypes, treatment use, reasons for discontinuation, and specific aspects of disease burden. The results were published and launched into an educational campaign to better educate providers and patients on persisting disease management challenges.

PicnicHealth utilizes patient-centric processes to collect, abstract, and link regulatory-grade, RWD and develops data models that enable life science organizations to generate evidence throughout clinical development and commercialization.

Case Study: Post-Marketing Study With Real-World Data

PicnicHealth was included in a 1,000+ patient post-marketing study for a cardiovascular disease treatment, which traditionally only collects clinical data from onsite visits. Enabling electronic collection of PROs and coupling this with historical data and prospective data reduces gaps in understanding the therapy's impact on healthcare resource utilization and disease outcomes, which can help address payer requests and post-marketing commitments.

Conclusion

Whether researchers are gathering RWD, or decision-makers are assessing RWE, healthcare and life sciences are navigating the new era of data and evidence together. To maximize the availability and capability of RWD and RWE, organizations need to ensure that their approach is truly patient-centric, consider how RWD and RWE will be used throughout the organization, and have a framework in place to determine the value of a data set before it is put to use. Contact PicnicHealth to learn how we can help power your research and evidence generation through patient centric RWD.

PicnicHealth

PicnicHealth is a healthcare technology company that partners directly with patients to build deep real-world datasets. The company leverages state-of-the-art machine learning, combined with human curation, to port complete medical records into an easy-to-use online application. The platform gives patients unprecedented access to and control over their medical records and, with their consent, the opportunity to contribute this valuable data to further scientific research. Founded in 2014 by Noga Leviner and Troy Astorino, the company partners with dozens of the world's largest biopharma companies and academic research institutions. Learn more at PicnicHealth.com and picnichealth.com/research-platform.