Real-World Data for R&D

Developing a much-needed new therapeutic, maybe even for Covid 19? Seeking approval to expand labeling for an existing treatment? Wanting to understand the standard of care, design a trial, create a comparator cohort, identify a new biomarker, or model patients most likely to benefit from a particular therapy? From drug development to personalized treatment, high-quality, real-world clinical data is rapidly changing the playing field for research and development teams.

Why Now?

First, over the last decade the availability of electronic healthcare data has grown exponentially. The American Recovery and Reinvestment Act of 2009 (ARRA) and the Health Information Technology for Economic and Clinical Health Act (HITECH) incentivized providers and hospitals to implement electronic health record systems and forever changed how medicine is practiced and documented. However, those same record systems were adopted without sufficient data standards and generated frustration among many researchers who have tried to clean, normalize, combine, and use these real-world data (RWD) sources.

Second, while early frustrations with using clinical data from these systems were very real, a parallel wave of technological innovation in big data systems and machine learning have significantly improved capabilities to better process and clean the data. While imperfect and still costly to some extent, the same technologies being utilized in other industries can now be applied. Initial efforts largely focused on processing relatively small cohorts (primarily in oncology), leveraging distributed manual abstraction to generate research-quality data. However, as described below, the application of AI and other technologies are achieving the quality of manual abstraction for significantly larger datasets in such areas as immunology, cardiology and respiratory diseases to name a few.

Third, the 21st Century Cures Act changed the opportunity landscape for how RWD and resulting real-world evidence (RWE) might be used and the value that could be generated. It provided a path for the use of RWD for specific regulatory purposes, such as comparator cohorts, label expansions, and post-marketing commitments. For example, the 2019 approval of IBRANCE (palbociclib) for male breast cancer patients was largely based on RWD.[i]

What is RWD & RWE?

The FDA, in its Framework for FDA's Real-World Evidence Program, December 2018,[ii] defines RWD as "the data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources." RWD can come from a number of different sources that would be considered systems of record, including electronic health records, claims and billing data, product and disease registries, patient generated data, and other experimental forms of information coming from sources such as mobile devices.

The FDA defines real-world evidence as "the clinical evidence regarding the usage and potential benefits and risks of a medical product that is derived from analysis of RWD." Driving home once again that evidence begins with the data.

Quality for Research and Regulatory Purposes

For R&D teams to use RWD, they need confidence in its quality. Before defining how best to achieve high quality with RWD, it's critical to recognize the limitations of RWD and to use it where it is fit for purpose. RWD is different from data from clinical trials. While normalizing heterogeneous data or obtaining verifiable structured information from unstructured clinical notes present addressable challenges, the nature of RWD itself cannot be modified with technology. RWD reflects routine clinical practice. Standard of care does not follow rigid visit schedules or have absolute requirements for specific tests. Researchers working with RWD should expect varying degrees of missing data to be common. That said, there are tremendous advantages to using RWD for R&D for the right use cases but that depends on following some key principles to ensure its quality.[iii]

1. Get the data that matters. Intentionality is a major driver for quality in RWD. That means that quality data collection begins with a research plan and a planned data collection

Richard Gliklich, M.D.
CEO
OM1

network specific to the condition of interest. Registry-like efforts to acquire RWD cohorts in specific therapeutic areas are increasing and, with automation, these data collection efforts can reach significant scale. Resist the common tendency to acquire large sets of data collected for other purposes that are simply mixed and matched together. They are rarely going to answer rigorous research questions.

2. Systematically process and assess the data for research uses. Modern, large-scale data processing is performed using carefully constructed and monitored data pipelines. Quality needs to be baked into the processes in a consistent, reliable, and verifiable way. This includes normalizing heterogeneous data from multiple sources, mapping it into a common data model, maintaining its provenance, deriving additional information from both structured and unstructured data, and determining condition-specific, patient-level outcomes. Quality cannot be added later.

Additional steps are required when using data for regulatory purposes. The FDA RWE Framework[iv] advises that RWD should be fit for purpose and that study conduct must meet FDA regulatory requirements. In practice, however, there are many different relevant FDA requirements that must be met depending on the particular field and use for the real-world evidence. From an RWD collection perspective, the most consistent principles are those of relevance, reliability and traceability. Relevance is having data that is sufficiently detailed to capture exposures and outcomes of interest in an appropriate population; and that the data elements enable the evidence to address the specified question. Reliability refers to the procedural and data quality factors involved in the collection and processing of data. How was the data collected? Were the sites appropriately trained? Were human sub-

> From drug development to personalized treatment, high-quality, real-world clinical data is rapidly changing the playing field for research and development teams.

jects protected? Was there adequate assurance that errors were minimized? How was missing data handled? Further, and very important from an auditing perspective, the chain of data transmissions and transformations needs to be understood and verifiable. In other words, the data should ultimately be traceable back to an individual patient.[v]

Challenges and Opportunities to Add Value

Despite the virtues of the digitization of data, there are very few standards in terms of how data is collected and stored in various electronic health record (EHR) systems. There have been movements towards interoperability but mostly around the exchange of information instead of the storage of the raw data. Adding to the chaos, most EHRs use their own data models and nomenclature around certain types of information. Processing and normalizing the data at scale is costly and resource intensive. Some data and technology companies have developed systems that alleviate this burden from life sciences manufacturers. For example, at OM1 we've built the OM1 Engine, a platform that processes and normalizes large datasets at scale, with the focus on delivering research and regulatory-grade data.

Another consideration is what additional data may be needed to fill in gaps and provide a more complete view of the patient journey. Consider if patient reported information is needed, which will vary by the nature of the disease. In autoimmune conditions for example, patient reported outcomes may represent key clinical endpoints. Rapid3 scores in rheumatoid arthritis (RA) and SLEDAI scores in lupus help measure disease activity and progression.

Another source of condition-specific clinical data may come from existing registries, which may already capture a wealth of data in certain condition areas. For example, the American Academy of Otolaryngology-Head and Neck Surgery Foundation (AAO-HNSF) Reg-ent registry collects data from early 3,000 specialists in conditions ranging from nasal polyps to head and neck cancer and is now being made available for life sciences research purposes.[vi]

Regardless of the source, key data for R&D use cases often resides as unstructured data in many RWD sources because clinicians largely document their findings through dictated encounter notes, pathology, imaging and other reports. These unstructured sources are critical for understanding key R&D questions, such as whether new lesions are present on an MRI in a multiple sclerosis (MS) patient or if a diagnosis has been confirmed by biopsy in a patient with NASH. Further, key metrics such as EDSS scores in MS or CDAI in RA may only be obtainable from the clinical notes. While extraction of information from unstructured data to structured data points has traditionally been managed by human abstraction, medical language processing and machine learning have enabled the ability to obtain this data at scale in much larger data sources and potentially in a more complete manner than through abstraction. The caveat is that different technologies and implementations have different levels of accuracy. Therefore, when structured variables are extracted or derived from unstructured data, the validation and performance characteristics of those derivations must be known in order to understand for which use cases those derivations may reliably be used. Ultimately the goal is to maximize the outcome variables of interest in the data. Fortunately, there are increasing efforts to standardize outcome measures such as the government sponsored Outcomes Measures Framework.[vii, viii]

Use Cases for RWD

Clinical data use cases for R&D are growing. For some uses, real-world clinical data are further linked to additional data sources, such as medical and pharmacy claims or social information, to fill in gaps or to add variables for analyses. The growing interest in personalized medicine requires subgrouping disease entities using both phenotypic and genotypic information, necessitating increasingly large and clinically deep datasets for modelers. Several use cases for RWD in R&D are listed below.

Regulatory:
▶ Post-Marketing Commitments
▶ Label expansions
▶ New drug approvals and comparator cohorts

Other R&D Uses:
▶ Identifying unmet need
▶ Applying AI to better target drug development
▶ Identifying subtypes / clusters of patients
▶ Improving clinical trial design and execution (protocol planning and testing, patient identification, or targeting responders)
▶ Understanding standard of care and value
▶ Developing new biomarkers leveraging phenotypic and omics data that will enhance clinical decision-making

Conclusion

Real-world data is having an increasingly important role in research, regulatory, and clinical-decision-making. Automation of data collection creates opportunities for more significant datasets, less bias and more use cases for research and development. Understanding emerging quality standards are key to generating research and regulatory quality RWD. Regulatory submission of RWD requires the data meet rising standards for relevance and reliability, but the opportunity for reducing time and costs to reach approval or to meet post-marketing commitments are huge. RWD use cases continue to expand across drug development making access to deep clinical and compliant RWD in each therapeutic area an increasingly strategic imperative for many R&D teams. ⓅⓋ

[i] Pfizer. FDA Approves IbrancePalbociclib for the Treatment of Men with HR HER2 Metastic Breast Cancer. April 4, 2019.
[ii] U.S. FDA. Framework for FDA's Real-World Evidence Program. December 2018.
[iii] Gliklich R, Leavy M. Therapeutic Innovation & Regulatory Science. Assessing Real-World Data Quality: The Application of Patient Registry Quality Criteria to Real-World Data and Real-World Evidence. March 27, 2019.
[iv] U.S. FDA. Framework for FDA's Real-World Evidence Program. December 2018.
[v] U.S FDA, Centers for Devices and Radiological Health. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices. August 31, 2017.
[vi] OM1. AAO-HNSF Partners with OM1 to Empower More Measured & Precise Care and Treatments for ENT. May 13, 2020.
[vii] U.S. Department of Health and Human Services. Outcome Measures Framework. December 2019.
[viii] Gliklich, et al. Annals of Internal Medicine. Harmonized Outcome Measures for Use in Depression Patient Registries and Clinical Practice. May 12, 2020.

A leading real-world outcomes and technology company, **OM1's** *RWE platform, clinical registries, data networks, and AI technologies accelerate research, measure, and benchmark health outcomes and personalize patient care.*

For more information, visit om1.com.