CASE STUDY

Enriching clinical studies with longitudinal real-world data

Aaron Berger
SVP, Head of Evidence Development

Judy Lytle, PHD, MBEE, PMP
Executive Director, Evidence Development Study Solutions

Jeff Lowry
Senior Director, Technology Solution Services

Executive summary

Real-world data (RWD) is healthcare data that is collected during routine clinical care, such as regular doctor visits, in-patient and out-patient procedures, and prescriptions. Clinical trials are outside of routine care, allowing participants to receive experimental treatments for disease. Linking clinical study data to real-world data (RWD) allows researchers to understand participants' pre-trial healthcare status, contextualize adverse events and discontinuation during the trial, and follow patients post-trial, illuminating a rich longitudinal healthcare journey for clinical study patient populations. UBC partnered with a European study sponsor to design and implement a robust data strategy to link RWD to ongoing clinical trial data to produce a longitudinal real-world patient healthcare dataset, looking back three years pre-trial, and extending three years post-trial. Patients were consented and tokenized, which de-identifies patient healthcare data by replacing the patient's personal identifiable information (PII) with an encrypted, patient-specific "token," protecting the patient's confidentiality. Data interoperability protocols and standards were applied to acquire and standardize RWD for linkage to a clinical study database. We report here a case study on this effort, detailing the value, operationalization, and outcomes of enriching clinical trials with longitudinal real-world data.

Introduction: Gain deeper insights with real-world data

Randomized clinical trials (RCTs) are the gold standard for the controlled evaluation of safety and efficacy of new medical treatments.[1] But, they are also limited by their narrowly defined study populations that often do not reflect the broader heterogeneity of disease and physiology found in the general clinical population. RCTs are further limited by their highly structured clinical visits and treatment applications, allowing only for the capture of the study population's healthcare journey within the tightly bounded timeframe of the trial.[2] As such, deeper insights that might be gained during the clinical study, including appropriate assessment of treatment safety and effectiveness that may rely on long-term evaluation, aren't realized.

One powerful way to overcome the limitations of traditional RCTs is the strategic use of RWD to extend and enrich the data capture of the clinical study. RWD is routinely collected from a variety of sources related to patient health and the delivery of healthcare, including electronic medical records (EMRs), insurance claims, pharmacy records, lab results, product or disease registries, provider notes, point of sale data, patient-generated data such as wearable fitness trackers, and even social media. The majority of health information that flows through a healthcare system is gathered as some form of RWD.[3]

RWD is often employed to optimize a clinical trial's study design during planning stages, to establish external controls for trials, and during post-market monitoring for data collection within the general clinical population to improve and refine safety and efficacy profiles, uncovering signals that will generally only emerge with a large enough sample size. This usually involves analysis of all patients in a real-world database with a particular disease or drug regimen. Taking these common applications of RWD a step further, researchers can investigate the real-world healthcare journey of just the participants in a clinical study, linking participant RWD with their corresponding clinical trial data. In doing so, RWD can help overcome many of the limitations of traditional RCTs by expanding the longitudinal reach of clinical study data and enriching the insights and assessments that emerge. With the proper data interoperability strategy, linkage of diverse RWD sources with clinical trial data can extend the view of the patient's healthcare journey outside the clinical setting and timespan of the study. This enables a deeper understanding of the patient's diagnostic and treatment journey beyond the limited context of temporally bounded, site-based research, and amplifies the power and precision of RCTs. Furthermore, linking RWD to clinical trial data can result in greater insights during the trial, including identification of additional markers of disease progression and surrogate endpoints, and contextualization of adverse events and participant discontinuation.

Strategically incorporating RWD into evidence generation planning activities requires a wide range of experience and expertise to implement it effectively. Data must be assessed against trial needs to ensure that the data are fit-for-purpose, but the lack of data standardization in RWD sources can make this truly difficult, requiring expertise in data science, data governance and curation, and data privacy. Properly applied modern approaches in feasibility, tokenization, linkage, data standardization, and analytics allow us to significantly enhance clinical trial results to better assess drug effectiveness, safety, and beyond.

Here we present a real-world case study describing RWD enrichment of a clinical trial, allowing us to analyze the clinical trial participant population in the three years leading up to the trial, and for three years after the trial ended. We discuss several associated topics, including:

- Overview of the clinical trial involved in the case study
- Insights that can be gained from linking longitudinal RWD to clinical study data
- Patient tokenization and data linkage
- Effective data governance and privacy compliance.

Case study: Linking RWD to clinical trial data

Overview

UBC's Evidence Development experts were engaged by a large EU therapeutics developer to assist in the design and implementation of a strategy to support and enhance RWD and real-word evidence (RWE) generation for their randomized, two-arm, open-label, Phase 3b clinical study evaluating a lipid-lowering therapy (LLT) strategy in adult patients with atherosclerotic cardiovascular disease (ASCVD). Specifically, UBC was engaged to assist in characterizing pre-trial healthcare journeys for participants in the study, linking real-world healthcare data with clinical trial data, allowing researchers to look back three years prior to the start of the trial, and extend data collection for three years following the trial.

LLT is used for the treatment of high low-density lipoprotein cholesterol (LDL-C) in patients with ASCVD.

This clinical study was designed to evaluate an LLT-first strategy versus usual standard of care. The study included 450 patients spanning 45 sites across 20 states in the United States and demonstrated that an LLT-first strategy led to significantly greater reductions in LDL-C from baseline.

Methods

The study sponsor wanted to understand these study results in the broader context of the pre- and post-trial healthcare journeys for the study participants. (Figure 1) Consenting trial participants were tokenized by a data vendor allowing RWD — collected from EMR, open and closed insurance claims, or laboratory records — to be linked to clinical trial data, allowing researchers to gain a more comprehensive and longitudinal view of patient health and healthcare. (Figure 2) Of the 450 trial participants, 89% were successfully tokenized, and 85% of tokens were successfully matched to one or more pre-trial data sources.

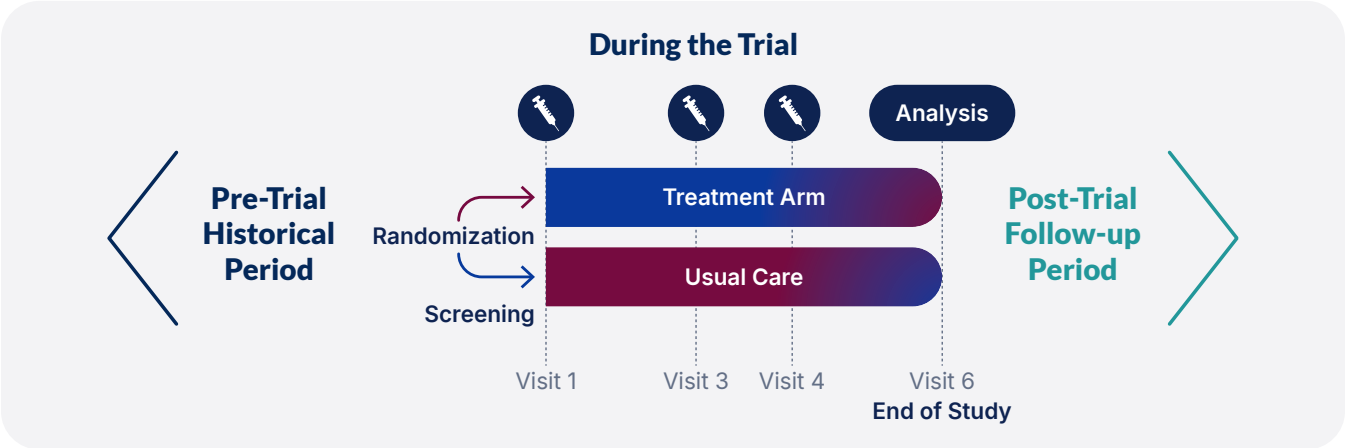

Figure 1. Linking pre- and post-trial longitudinal RWD to clinical trial data

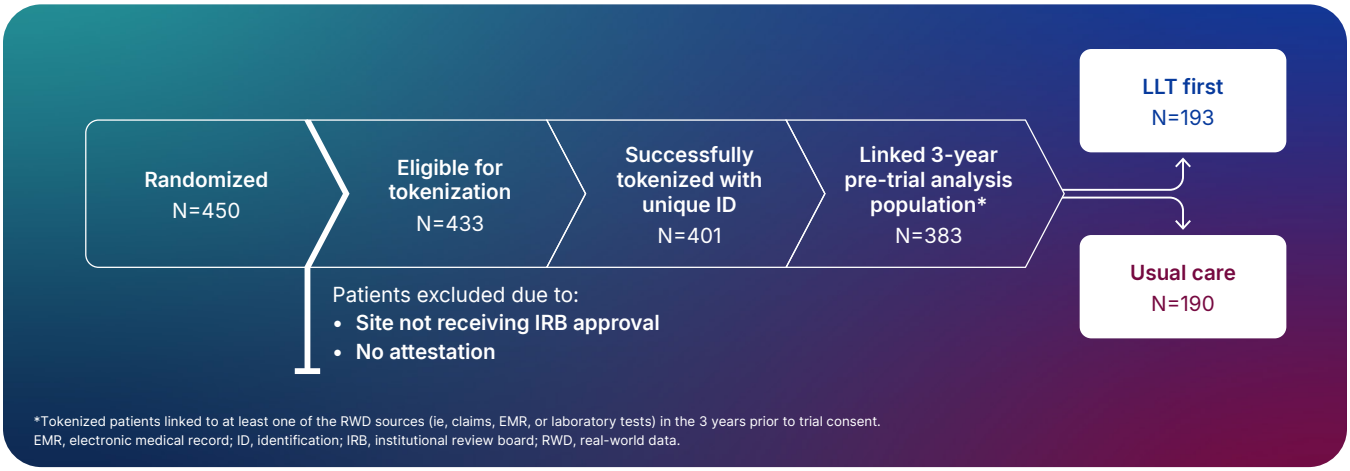

Figure 2. Tokenization of patients

Results

Tokenization enabled the study sponsor to link various, distinct RWD sources from trial participants' medical histories with clinical trial data, providing the capability to understand the pre-trial and post-trial healthcare journey. In the pre-trial period, treatments administered (e.g. LLT versus statin monotherapy, monoclonal antibodies, or selective cholesterol absorption inhibitor), the number of years when LDL-C was measured, and the number of patients receiving LLT treatment were evaluated pre-index ASCVD, post-index ASCVD, or pre- and post-index ASCVD. During the post-trial period, safety and efficacy/effectiveness can be monitored in the same clinical trial population, producing an extensive longitudinal dataset.

Linking RWD and clinical trial data has great value for drug development

This combination of RWE generated from pre- and post-trial RWD, with evidence emerging from clinical study data, can provide deeper insights that help characterize a treatment's safety profile and effectiveness:

Enriched view of pre-trial medical history

Enriching trial data with real-world healthcare journeys in the pre-trial period uncovers valuable insights into key healthcare parameters, such as previous healthcare resource utilization (HCRU) treatment patterns, standards of care for the study population, and potential markers for earlier diagnosis.

Contextualization of patient events

Adverse events and study drop out or discontinuation of treatment can be evaluated by investigating healthcare journeys before or during the clinical trial.

Clinical development

Enriched longitudinal trial data can affect clinical development in several ways, such as potentially identifying additional markers of disease progression, or identifying surrogate endpoints.

Closing the efficacy-effectiveness gap

The difference in drug efficacy for clinical study populations and its broader effectiveness in larger real-world populations is an ongoing issue for drug development.[4] Linking RWD to clinical study data in an extended longitudinal data set has the potential to benchmark the RCT population against larger RWD populations, to better understand clinical trial outcomes versus real-world outcomes.[4]

Identifying post-market patterns

Post-trial RWD can provide an understanding of the real-world administration and dosing regimen of the treatment being analyzed or other relevant treatments, treatment adherence and switching patterns, adverse events, long-term safety outcomes linked to a drug (e.g. hypertension in the case study above), emerging comorbidities (such as obesity or diabetes mellitus), and other basic health measures. For example, if a therapeutic is under scrutiny for safety signals that arise in the real world, investigating the larger real-world population could reveal important differences from the clinical trial population, where the signal was not observed. This could result in a label change based on patient sub-populations, versus a regulatory recall.

Enriching clinical trials with distinct RWD sources from patients' medical histories provides unique opportunities to gain a more comprehensive and longitudinal view of patient health and healthcare, and drug safety and effectiveness in a clinical trial population outside of the timeframe of the clinical trial.

Rich longitudinal data capture requires a strong data interoperability strategy

Producing a more comprehensive, longitudinal dataset that links RWD with clinical study data requires a well-designed and executed data strategy, including patient consent and tokenization, well thought out data sourcing and standardization, and robust data management.

Patient tokenization

Tokenization de-identifies patient healthcare data by replacing the patient's personal identifiable information (PII) with a patient-specific encrypted "token" that protects patient confidentiality. This allows their real-world healthcare data, which is recorded during routine care, to be collected and linked to clinical trial data while preserving privacy under the Health Insurance Portability and Accountability Act (HIPAA) and other standards.[5] The 'token' stands in for the individual identity related to a patient's medical information. Any data associated with the study participant can be linked across databases where their information is present — insurance claims, pharmacy visits, lab activity, hospitalization records, and other forms of EMR — without disclosing their identity.

Tokens are typically derived from a combination of PII, often including first and last name, gender, date of birth, and zip code.[6] Study participants must consent to having their identity tokenized by an approved data vendor. After tokenization, medical data for an individual can be kept up to date without the need for tracking. Tokens provide the linkage between separate RWD sources and clinical trial data, enabling the fuller picture of a participant's health journey.

Patient consent of tokenization

Patients choose to participate in clinical studies for a myriad of reasons, typically centered around their personal healthcare status. As such, it is critical that privacy concerns be assuaged with carefully crafted consent language that clearly describes the reasons for tokenized data capture, the benefits that will come from the process, and the safeguards that will be consistently applied to ensure study participants do not have to worry about data privacy.

Data sourcing, standardization, and matching

The sourcing and use of RWD comprise an emerging, quickly evolving field, and there are hurdles to overcome during implementation. Merging and analyzing vast stores of data kept in databases separated by time, region, collection methods, and standards pose complex challenges. Whereas RCTs are carefully designed to eliminate biases and other confounding factors, RWD has the potential for bias and confounding factors which may need to be accounted for. Electronic healthcare data was not initially developed for research purposes, and the use of RWD involves databases of varied provenance, quality, and structure, requiring diverse expertise and careful analysis to be effectively leveraged. This lack of standardization in RWD must be considered and addressed during study design, and protocols must be established to standardize RWD so that it conforms to regulatory guidelines for data structure.

There are also challenges in bringing RWD into alignment with data standards set forth by regulatory bodies, due to issues such as inconsistent coding or miscoding of drugs or diagnoses, changes in data collection or coding practices, and missing information from inconsistent data entry. In December 2023, the FDA recognized the importance of RWD standardization and finalized the guidance on Data Standards for Drug and Biological Product Submissions Containing RWD to help sponsors who plan to use RWD in submission packages.[7]

Each data source has strengths and limitations with respect to the suitability of the data elements to answer the research questions of the given study. Historical knowledge of RWD sources optimizes the data selection process to ensure fit-for-purpose data is being used to achieve research objectives. It is also critical to choose data sources that are well-established over time and large enough to ensure a high degree of data matching. Ensuring robust data matching typically also requires curating multiple datasets simultaneously. As such, the unification of data from disparate data sources requires integrated expertise in healthcare data science, clinical trial operations, and interoperability technology.

Data matching involves the confirmation that data connected to each patient token can be identified and curated from at least one, although ideally more than one, of the chosen data sources. A high degree of data matching is enabled by effective validation methodologies that avoid data entry errors and address inaccuracies (e.g., false names, missing details, duplicate patient entries). Data matching against multiple data sources helps create deeper datasets and more confidence in the resulting aggregate dataset, and robust data matching also ensures there is as strong an overlap as possible.

Data management instead of site management

Traditionally, collection of pre-trial patient data would entail a manual process that shifts a large part of the burden to the patient – asking them to recall data from their healthcare journey or manually track down and deliver their healthcare records. Such a process is not only manually challenging but is also likely to introduce more errors. Tokenization and data matching allows this process to be shifted from patients and site managers to the information technology infrastructure. This results in a streamlined data acquisition process that effectively shifts an RWD-enriched clinical study from a posture of site management to data management.

Data management to support the capture and operationalization of longitudinal datasets must ensure that all data is collected into a common data structure that is unified, standardized, and useful. Effective data management involves a series of continuous checks that ensure that RWD is collected and standardized appropriately (Figure 3):

- Verification of tokens with each data capture
- Verification of token accuracy
- Verifying the integrity of the matching process within the patient identity module
- Optimization of the percent overlap and utility during data matching
- Optimization of data continuity and data gap reduction

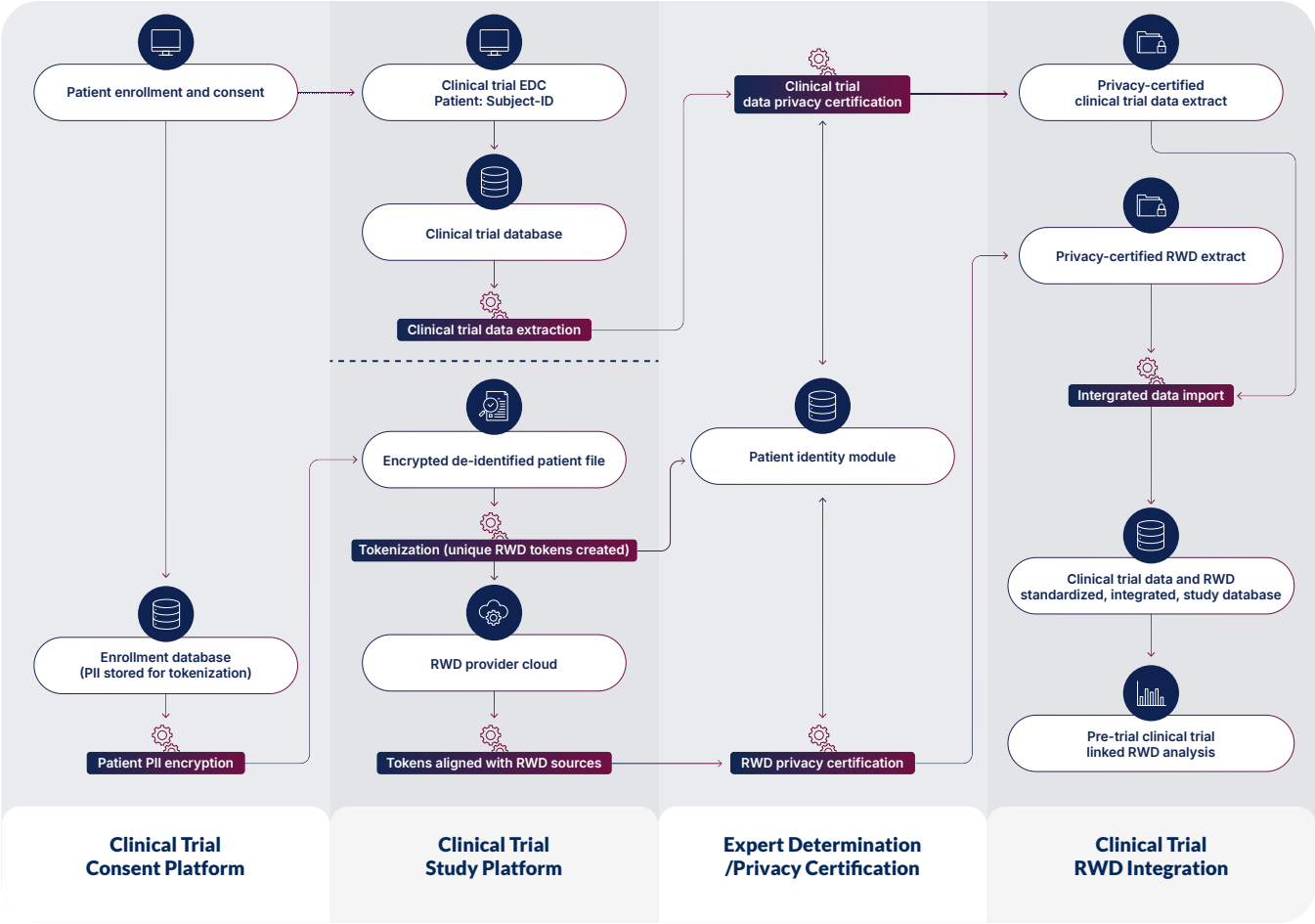

Figure 3. Deidentification, tokenization, and data management process for linking patient RWD to clinical trial data.

Conclusion: Strategic differentiation in RWD/RWE development

By leveraging a strong RWD/RWE strategy and incorporating leading technology to acquire and standardize longitudinal RWD linked to ongoing trial data, high-quality healthcare insights can be generated at a fraction of the cost of obtaining the same data exclusively through site based manual data capture. Where appropriate, this is a more efficient way to do longitudinal and follow-up studies, with the added value of potentially lowering the expense for payments to healthcare providers who participate in studies.

There is a broad and growing consensus that RWD is a powerful tool. RWE generated by RWD can support regulatory approval, and help sponsors better meet medical needs. Major shifts in the clinical trial landscape are impacting how studies are conducted, including the rising cost of clinical studies, the growing need for a more holistic picture of patient health, and the emergence of new technologies that make vast pools of medically relevant data accessible and useful. In meeting these emerging challenges, the curation, collection, and analysis of longitudinal real-world datasets represent an exciting means of enriching clinical studies.

The case study outlined here provides a roadmap for efficient, cost-effective operationalization of RWD capture and integration that preserves the integrity of patient data privacy, while establishing a dynamic infrastructure that can continue to provide real-world insights post-approval. UBC continues to execute innovative, efficient, and compliant studies while delivering high-value insights for clients. Robust infrastructure, strategic RWD/RWE expertise, and operational excellence position UBC as a key partner. The UBC team's deep understanding of RWD regulatory guidance, integrated evidence planning, and technical requirements that preserve data privacy protections not only facilitates study approval but also helps navigate the complex compliance landscape of large, regionally distributed clinical trials, such as the case study outlined above. UBC's ability to build privacy-preserving, scalable infrastructure supports a shift from traditional clinical study methods to more innovative, cost-effective approaches.

About UBC

United BioSource LLC (UBC) connects specialty therapies to the patients who need them most by delivering modern, customized solutions in access, safety, and evidence generation. Bringing over 30 years of experience, UBC provides expert-driven real-world evidence to uncover more valuable insights, optimize regulatory approval, and maximize commercial positioning for the long-term value of your therapy. We take a thoughtful, collaborative approach – guiding you through every step of program design and delivery. Leveraging real-world data and the latest technologies, we create solutions that are innovative as well as grounded in real-world patient experience. Our focus is on helping you navigate the complexities of regulatory approval, successful commercialization, and delivery of long-term outcomes – ensuring success every step of the way.

To learn more about how UBC can help you enrich your clinical studies or develop an integrated, real-world evidence strategy, reach out to us at **contact@ubc.com**

References

1. Meldrum, M. L. (2000). A BRIEF HISTORY OF THE RANDOMIZED CONTROLLED TRIAL. Hematology/Oncology Clinics of North America, 14(4), 745–760. https://doi.org/10.1016/s0889-8588(05)70309-9

2. Corrigan-Curay, J., Sacks, L., & Woodcock, J. (2018). Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness. JAMA, 320(9), 867. https://doi.org/10.1001/jama.2018.10136

3. U.S. Food and Drug Administration. (2018). Framework for FDA's Real-World Evidence Program. 2018. Accessed January 30, 2025. http://www.fda.gov/media/120060/download Accessed January 30, 2025.

4. Rudrapatna, V. A., & Butte, A. J. (2020). Opportunities and challenges in using real-world data for health care. The Journal of clinical investigation, 130(2), 565–574. https://doi.org/10.1172/JCI129197

5. U.S. Department of Health and Human Services. (2022, October 19). Summary of the HIPAA privacy rule. HHS.gov; U.S. Department of Health and Human Services. Accessed January 29, 2025. https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

6. Eckrote, M. J., Nielson, C. M., Lu, M., Alexander, T., Gupta, R. S., Low, K. W., Zhang, Z., Eliazar, A., Klesh, R., Kress, A., Bryant, M., Asiimwe, A., Gatto, N. M., & Dreyer, N. A. (2024). Linking clinical trial participants to their U.S. real-world data through tokenization: A practical guide. Contemporary Clinical Trials Communications, 41, 101354. https://doi.org/10.1016/j.conctc.2024.101354

7. Center for Drug Evaluation and Research and Center for Biologics Evaluation and Research. (2021, October 21). Data Standards for Drug and Biological Product Submissions Containing Real-World Data. U.S. Food and Drug Administration. Accessed January 29, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-drug-and-biological-product-submissions-containing-real-world-data

ubc.com
contact@ubc.com