WHITEPAPER

Standardizing and Linking Diverse Registry and Other Real-World Observational Data Yields Enhanced Understanding of Rare Diseases and Their Treatments

PRESENTED BY:

UBC

PATIENTS FIRST | POWERED BY EVIDENCE

PUBLISHED BY:

FierceBiotech

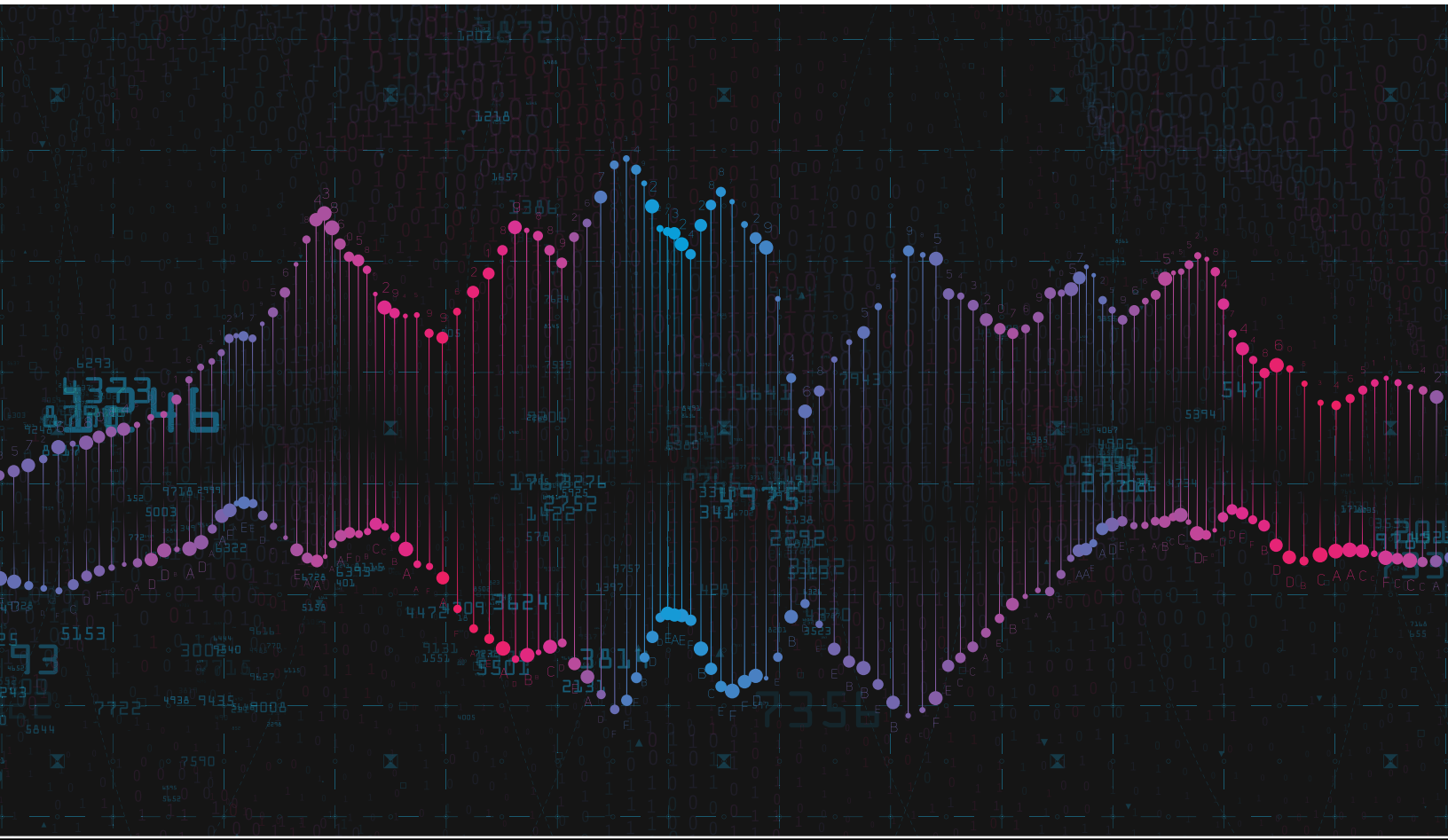
CONTENTS

INTRODUCTION

Real-world data (RWD) and the real-world evidence (RWE) generated from it are playing an increasingly central role in clinical research, healthcare management and the development of new drugs and biologics.[1] The data collected from diverse sources, such as observational studies, including registries and post-marketing cohort studies; electronic medical records; claims databases and vital statistics records, such as the US National Death Index, support advanced analytics and research, both pre- and post-marketing.

These data from sources outside a controlled clinical trial can be a rich source of insights[2] on disease history, patient outcomes, and more. Standardizing and integrating RWD from diverse data sources into a common data structure can overcome issues, such as gaps in data collection, inconsistent data formatting, and divergent reporting protocols, providing an easier and more efficient approach to assembling a relevant body of RWD. However, collecting, managing and analyzing data from an increasingly complex and widely variable array of data sources poses challenges that researchers may not have the time, resources, or expertise to address.

With a robust data architecture designed to collect, manage and automate data from a constantly expanding body of databases and registries, UBC's integrated strategies for standardizing and linking RWD sources provide comprehensive solutions to these challenges. This approach is particularly important in the study of patients with rare diseases, where no one data source is sufficiently large to provide a satisfactory patient profile or to document the patient journey and the impact of a new therapy.



REAL-WORLD DATA OFFERS NEW INSIGHTS AND CONSIDERABLE CHALLENGES

Carefully designed clinical trials remain at the core of healthcare and pharmaceutical research, but the parameters set for these studies can create artificial environments in which to understand product safety and efficacy, leaving essential questions about use in clinical practice unanswered and imposing limits that exclude some patients from the study. These limitations can result in biases that impact study outcomes in a variety of ways.

Until recently, regulatory agencies such as the FDA and the EMA were reluctant to acknowledge the value of RWD and RWE in the clinical trial process. But that stance has changed, in part due to the availability of massive amounts of high-quality patient data that can be accessed anywhere, anytime using an array of technology-driven processes. Not only does the FDA suggest[3] sponsors conduct studies to understand the natural history of disease as a complement to a clinical trial (or while planning the portfolio of trials), but both the FDA and EMA urge sponsors to understand available RWD and partner with relevant sources rather than creating a de novo dataset. These additional RWD, beyond clinical trial data, are especially important for studies involving extremely rare diseases where an investigator may only treat one or two individuals.

“Not only does the FDA suggest sponsors conduct studies to understand the natural history of disease as a complement to a clinical trial (or while planning the portfolio of trials), but both the FDA and EMA urge sponsors to understand available RWD and partner with relevant sources rather than creating a de novo dataset.”

MANAGING DIVERSE DATA SOURCES

RWD can be collected from a variety of sources, including electronic health records maintained by clinics and hospitals, insurance and medical claims, billing databases, logs, and records of patient activities in therapeutic and in-home settings, as well as patient-reported health information and data collected from various digital health monitoring devices. Patient registries for patients with specific health conditions[4] can be very important sources of RWD, particularly in cases of rare diseases.

RWD are the raw materials that must be collected, cleaned, synthesized, and interpreted to produce evidence that responds to research questions posed by a specific study. RWE derived from aggregating RWD can then provide insights on the natural history of diseases, characteristics of the patient population as a whole, and outcomes for patients treated with products under study and their competitors. In this way, RWE can inform the progress of a particular study and the approaches to managing patient care outside of it. RWD and RWE are important adjuncts to standard clinical trial protocols[5], but their use can be hampered by the sheer amount and diversity of the data involved.

The task of implementing an RWD strategy is complex, with many operational and integration processes that need to be considered, not only within a sponsor company, but also for each organization providing data. Gaining an understanding of every component such as the source of the data; how the data will be collected, structured, and exported; and work practices are imperative.

When a source of RWD is an established consortia dedicated to particular research goals, data standardization is complicated by the use of different databases by members of the consortia, typically individual clinical sites. Communication among consortia can be limited due to questions about what entity owns the relationship with the consortia, and whether the consortia is established as a legal entity. Establishing a data standardization protocol across all sites often requires the consortia to obtain individual consent from these separate clinical entities. This is best achieved through a Data Sharing Agreement. A gap analysis of all relevant databases must be completed before a Data Sharing Agreement can be implemented, since the agreement and budget are to be submitted to the consortia together.

In order to operationalize and identify gaps in data collection or inconsistencies across data sources, it is critical to obtain the following documents and information from the outset:

- Non-disclosure agreements (NDAs)
- Detailed data sharing agreements, including a commitment to utilizing the GUID (Globally Unique Identifier) process and tools for identifying unique patients
- Database structures and data collection procedures, including data collection tools, and protocols for sharing only summary data rather than individual patient data
- Data Transfer Plan and test transfers

Some projects using RWD do not attempt to supplement specific patient data with additional sources, and may view patients in each data source as unique and separate. Other programs seek to match patients in one data set with the same patient in another, in order to supplement the data available on any individual patient and create a larger base of data from which to draw inferences and identify patterns. However, these data sources are often created for differing purposes, so that, for example, the information contained in a claims database may not completely match the data in a patient's medical records. In addition, the protocols used to label and categorize data may be idiosyncratic to a particular database or registry. These issues create problems with matching patient data across different data sources and identifying areas where information is missing as well as with ensuring that all relevant data are attached to the appropriate patient.

The challenges of collecting and reconciling data from multiple databases created for different agendas[6] are further complicated by international and national differences in managing health-related databases. For example, the United States lacks a system for linking health-related databases for quick access to relevant records, but some European countries, such as Denmark, link health records, enabling multiple records to be accessed in response to a specific research question.

In the US, the task of standardizing RWD for the generation of RWE falls to companies able to employ sophisticated digital technologies. In this way, relationships between records can be established in order to produce RWE while remaining compliant with HIPAA and other regulations pertaining to patient privacy and consent.

Protocols created by the European Union's recently implemented General Data Protection Regulation, or GDPR, also affect the collection and standardization of patient data. These regulations affect data of EU citizens wherever it may be used, but they are implemented in different ways by individual member states. In France, for example, no protected health information (PHI) is collected, which makes it difficult to identify unique patients, and Germany implements more stringent protocols for protection of personal data than some other countries, which can complicate data sharing across international databases.

STANDARDIZING REGISTRY DATA STREAMLINES RESEARCH PROCESSES

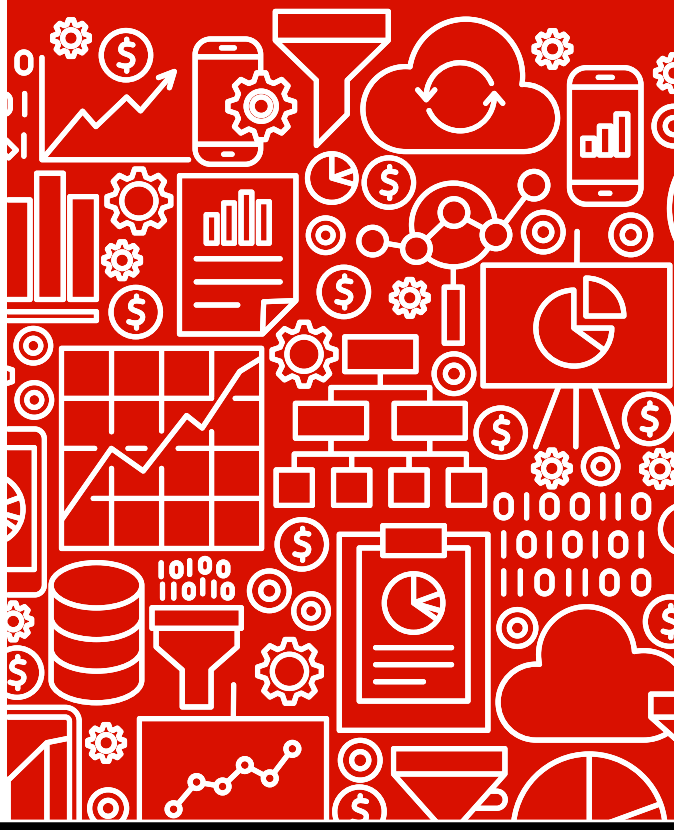
Collecting, managing, and processing RWD from its many sources in order to create a standardized body of data that can generate RWE in a usable, accessible form is a process that begins with collecting the data held in selected databases and relevant registries.

The process for standardizing diverse data starts with a protocol that presents the details of the algorithms and approach to enable merging information across datasets. Such a protocol for creating a common data structure must address the requirements of various stakeholders and regulatory bodies, including:

- Considerations related to privacy and consent
- Data ownership
- Use of digital tools for accessing and managing data
- Identifying and planning for data gaps and formatting inconsistencies
- Ownership of any products produced from the data

UBC's common data model and data processing policies/procedures address these and related requirements, using commonly defined data measuring standards applied across all project and real-world data sets. This is done in a seamless and streamlined model that moves RWD from its disparate sources to an integrated single data source to support and complement the needs of the study.

“UBC's common data model and data processing policies/ procedures address these and related requirements, using commonly defined data measuring standards applied across all project and real-world data sets.”



TOKENIZATION ANONYMIZES AND PROTECTS PATIENT PRIVACY

When data are collected from various data sources to form a single data repository, patient privacy and regulatory compliance are paramount. To ensure patient privacy and regulatory compliance when building a common data structure for a particular study, the relevant medical and other personal patient information must be anonymized. This patient anonymization process not only protects patient privacy, but also contributes to the standardization and cleaning of data sets through the ability to truly de-duplicate patients in the study.

To protect patient privacy, UBC has partnered with HealthVerity© to utilize their Census product line to identify a patient's personal information in a manner that addresses all HIPAA concerns. HealthVerity's innovative cloud model for de-identification and matching of health data assigns a unique and persistent ID to patient populations across multiple data sources, so that while each individual's identity is protected stakeholders can link to a wide array of data points in the HealthVerity Marketplace.

> When data are collected from various data sources to form a single data repository, patient privacy and regulatory compliance are paramount.

ANALYTICS AND AUTOMATION PRODUCE REAL-WORLD EVIDENCE

To maintain consistency across different registries and databases, individual data elements must be compared in a gap analysis to highlight differences in data points, formatting, and capture that might impact the integration of all data sets into a common database.

Because not all data points are a 1:1 match across registries and databases, UBC's Data GAP analysis utilizes tools including CRFs and multiple data dictionaries that define not only the contents and structures of individual databases, but also act as repositories of metadata that can expedite data matching and identification of actual gaps. In this process, data points can be categorized either as an exact, 1:1 match between data sources, or as a logical match – a match found from disparate responses to a data point, or one that could be derived from a comparison of other data points. A data point might also be categorized as omitted, a situation in which the consortia data did not capture a specific data point from a comparable registry or database and therefore could not perform a logical match.

Once all data have been analyzed, anonymized and standardized to meet the requirements of a specific study, the data can be warehoused as a common data model, ready to be interpreted, analyzed, and reformulated into a variety of structures to suit the needs of a particular requester. Those forms can include analytics, statistical models, or summaries that deliver the validated RWE to complement other observational study research or results from a controlled clinical trial setting. These processes make it possible to derive algorithms from the collected databases, and to create a mapping from consortium entities to databases.

A variety of data reporting models make collected data as accessible as possible to consortia members, study sponsors, and research teams. UBC's proprietary RWD tools are designed to automate analytics and visualizations. With a patient or data owner consent, sponsors can aggregate and analyze standardized data sets to meet project objectives and potentially find new therapies or treatments, or identify trends that could potentially facilitate disease detection. Once a study has concluded, common data models are locked and data are unavailable without the consent of the data owner(s).

CONCLUSION

Standardizing RWD into a common data model can enhance an ad hoc study, or the aggregated data can be used to create a cohort for analysis of disease natural history or a product's safety[7]. With the ever-increasing flow of data and an abundance of disparate data sources, such as registries with widely divergent protocols, electronic health records, or claims databases, selecting the optimal data and managing the information can pose challenges to researchers. With state-of-the-art data management systems incorporating industry-standard tokenization protocols, blockchain technology, and automated systems for extracting insights relevant to the scope of specific studies, UBC's model for standardizing registry and other observational data can streamline the research process, support the development of new treatments, and provide crucial insights for patient care.

Notes:

1 Sherman, Rachel et al. "Real-World Evidence – What Is It and What Can It Tell Us?" New England Journal of Medicine. 8 Dec 2016. https://www.nejm.org/doi/full/10.1056/NEJMsb1609216

2 Jadhav, Sujay. "Using Real World Data to Enhance Clinical Trials." Clinical Research News. 13 Jan 2017. http://www.clinicalinformaticsnews.com/2017/1/13/using-real-world-data-to-enhance-clinical-trials.aspx

3 "Real World Evidence." US Food and Drug Administration. 9 May 2019. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

4 "3 Huge Ways Registries Can Benefit Rare Disease Patients." Fibrous Dysplasia Foundation. 14 Jul 2016. https://fibrousdysplasia.org/3-huge-ways-registries-can-benefit-rare-disease-patients/

5 Nelson, Eugene et al. "Patient Focused Registries Can Improve Health, Care, and Science." BMJ . 1 Jul 2016.https://www.bmj.com/content/354/bmj.i3319

6 Spitzer, Ginger. "A Collaborative Approach to Real World Data Collection." Pharmaceutical Outsourcing." 20 Mar 2018. https://www.pharmoutsourcing.com/Featured-Articles/348139-A-Collaborative-Approach-to-Real-World-Data-Collection/

7 Fleurence, Rachel et al. "The Future of Registries in the Era of Real-World Evidence for Medical Devices." JAMA Cardiology. 2019(4). https://jamanetwork.com/journals/jamacardiology/article-abstract/2724987