Overview
Over the past several years a new breed of real-world evidence (RWE) analytic platforms has emerged, enabling researchers to submit analysis parameters through a user-friendly interface, execute modular analytic programs across a variety of data sources, and produce results rapidly without custom programming. These RWE platforms emphasize shortened analysis cycle times, generating results in minutes or hours for analyses that used to take days, weeks, or longer. There is no denying that these rapid cycle analytic platforms have had a meaningful impact on analytic efficiency. However, many are proprietary – forcing RWE organizations to choose a single technology solution, or manually mediate workflows and analysis results across multiple solutions. As these platforms increase in number and availability, they also contribute to an increasingly fragmented and disorganized RWE environment.

At the same time, a significant body of research highlights the benefits of data standardization as a way to help manage issues caused by the growing volume and heterogeneity of real-world data. A standards-based ecosystem supports interoperability among diverse RWE stakeholders, enabling RWE analytic platforms that “plug into” the standard to deliver not only analytic efficiency, but also reuse and reproducibility. The greatest benefits of data standardization will only be achieved with the adoption and use of a single standard across an entire RWE ecosystem, and those benefits are substantial. Use of a single standard extends the capabilities of RWE analytic platforms beyond the current focus of rapid cycle analytics towards an environment of true analytic democratization.

The Current RWE Environment: Fragmented and Rapidly Evolving
The growing need for evidence generated from real-world data is disrupting the product development lifecycle. Real-world evidence is essential at every stage – from understanding product value and achieving favorable market access, to preserving and enhancing product positioning. At the same time, the RWE environment is evolving rapidly. Several current trends have the potential to significantly impact the way in which evidence is generated in the future.

Data, Data, and More Data
While secondary analysis of health system data has been used in epidemiological research for decades',
the volume and variety of secondary data available for real-world evidence generation is exploding, both in the U.S. and across the globe. In the U.S., the availability of patient electronic health record (EHR) data for medical research is expanding, partially catalyzed by the widespread implementation of “meaningful use” standards that provide incentives for providers and hospitals when they use EHRs to achieve specified improvements in care delivery. Unlike many other secondary data sources, EHR data sources provide patient data collected at the point of care in near real time, blurring the boundary between retrospective and prospective research. In addition, many EHR systems also include previously untapped sources of real-world data, such as free-form text found in physician notes and imaging data in therapeutic areas where images are used in diagnosis and treatment.

Beyond health systems, data generated by individuals – mobile health data and social media, for example – are growing exponentially. Consider the following:

- Individuals generate 70 percent of all available data worldwide. More data has been created in the past two years than in the entire previous history of the human race.
- Within five years there will be over 50 billion smart connected devices in the world, all developed to collect, analyze and share data.
- By the year 2020, about 1.7 megabytes of new information will be created every second for every human being on the planet. Our accumulated digital universe of data will grow from 4.4 zettabytes today to an estimated 44 zettabytes in 2020.

Yet, less than 0.5 percent of data currently generated by individuals are analyzed or used today.

A Patchwork of Analytic Capabilities

There are a growing number of both commercially available and internally developed RWE analytic platforms, providing a wide variety of analytic capabilities. For instance, many data providers include “query tools” to analyze their own proprietary data. Some analytic platforms target a specific type of analysis (e.g., pharmacovigilance), while others focus on analysis of a particular type of data (e.g., observational data, social media) or a specialized technical issue (e.g., natural language processing, data visualization). Overall, these RWE analytic platforms have had a significant, positive impact on the speed and efficiency of analysis execution. However, many also include proprietary technologies, methods, and/or data, forcing RWE organizations to manually address gaps, overlaps, and inconsistent workflows among platforms, and to mediate conflicting sources of evidentiary “truth” when different analytic platforms produce conflicting results.

Evidence Generation as a “Shared Service”

Until recently, real-world evidence generation has mainly been a specialized function – confined to teams with custom resources, data, and technologies to address the specific evidence generation needs of that team. However, this silo mentality is beginning to change as enterprises recognize that the same real-world data, technologies, and resources can support evidence generation needs across an entire organization. RWE Shared Service Centers are growing in popularity, providing centralized evidence generation capabilities across a broad and diverse population of evidence consumers. The main objective for providing evidence generation as a shared service is to increase efficiency and cost savings through the sharing and reuse of data, technology, resources, and analytic expertise.

Real-World Data Standards: Ready for Prime Time?

Standardized approaches to real-world data analysis have been widely studied as a means to cope with growing volume and heterogeneity of real-world data. Analysis standardization relies on the “harmonization” of the data – that is, the use of common words (data elements and terminology), structures, and data organization across disparate data sources. This is often accomplished through the use of a Common Data Model (CDM).

Several research networks have developed and implemented CDMs into their clinical research infrastructure as a means to promote efficiency in evidence generation practices and to provide better interoperability among diverse research partners. These networks, briefly described below, provide a public forum for advancing the science of analysis standardization, while highlighting the benefits (and drawbacks) of such approaches.

- The Observational Health Data Sciences and Informatics Program (OHDSI)
  OHDSI (pronounced Odyssey) is a multi-stakeholder, open source collaborative with an established international network of researchers and observational health databases. OHDSI research is focused on the development standardized methods and tools for large-scale analytics of health data using the Observational Medical Outcomes Partnership (OMOP) CDM.

- The FDA Mini-Sentinel Program
  Mini-Sentinel is a pilot project sponsored by the U.S. Food and Drug Administration (FDA) to create an active surveillance system to monitor the safety of
The Future of Evidence Generation: Standards-Based, Efficient, and Democratized

As the need for real-world evidence intensifies, RWE organizations who are the most proficient in evidence generation will have a significant competitive advantage over their peers. However, these organizations are currently struggling with an explosion of disparate data and technologies, coupled with a lack of standards, all of which contribute to an increasingly fragmented and inefficient environment for evidence generation.

Data Standardization using a CDM has been widely studied as a way to enable agile research, providing a framework for rapid and transparent analyses across heterogeneous databases in support of research-related questions. Several research communities have developed CDMs for real-world data analysis; however, differences in these CDMs can lead to inconsistent analysis results. In order to fully realize the benefits of using a CDM beyond individual research communities, the adoption of a single, universal data standard will be necessary. This requires the convergence of existing CDM research into one uniform set of data standards that are accepted and implemented across diverse RWE researchers, communities, and stakeholders. While alignment of existing CDMs may be difficult to achieve, it provides the greatest potential for expanding CDM research and use beyond the boundaries of individual research communities into the broader RWE ecosystem. The main benefits of an industry-wide, uniform data standard are summarized below.

Interoperable and Reproducible

Adoption of an industry-wide data standard would provide an interoperable evidence-generation infrastructure supporting diverse organizations – researchers, RWE organizations, data and technology vendors, etc. Any analytic program or technology could “plug into” the standard and analyze any data source conforming to that same standard. Moreover, analysis results produced from a CDM are easily reproducible and meaningfully comparable across disparate data sources.

This is substantiated by the results of several studies. In one recent study, OHDSI researchers replicated an analysis across six disparate databases in the OMOP CDM format using one analytic routine, efficiently producing a consistent set of results. Without the CDM, independent programs would be required for each database and results may not have been directly comparable due to differences in the data structure, source vocabulary, and analytic module customization.

Open, Transparent, and Collaborative

The use of a CDM facilitates standardization across the evidence generation lifecycle, making it possible for RWE stakeholders from diverse backgrounds (clinical, epidemiology, data science, technology, etc.) to work together. For instance, modular analysis programs written for CDM-format data can be developed and distributed for use across an organization. Parameters for selecting patient cohorts, healthcare events, and covariates can be defined and stored in a library for reuse and sharing. The industry-wide adoption of a common data standard also supports the creation of transparent, open source repositories of cohort and event definitions, and modular analysis programs. Interested stakeholders from across the entire industry could develop, publish, share, and reuse cohort and event definitions and analysis modules, moving away from a fragmented environment of custom analytic programs and technologies towards an environment of collaboration and analytic democratization.
Getting from Here to There

Rapid-cycle analytic technologies represent an important advance in the pursuit of improved evidence generation practices - but this is only the first step. A growing body of research substantiates the notion that adoption of data standards through the use of a CDM provides the foundation for an interoperable, transparent evidence generation ecosystem. Data standards not only support efficient evidence generation, but they also promote sharing, reuse, interoperability, and reproducibility across diverse RWE stakeholders. In this environment, data providers, technology providers, researchers, and other RWE stakeholders can continue to innovate and develop new analytic offerings, while increasing the value of those offerings by plugging into a standardized infrastructure that supports interoperability and integration with the rest of the RWE ecosystem.

There are a handful of similar CDM standards available today, each making a significant impact within a confined community of researchers and users. Convergence of these standards into a single, universal CDM could potentially spread the benefits of standardization beyond research communities and into the RWE ecosystem. Moreover, adoption of a uniform standard could facilitate better harmonization of the research being done by these communities. Future research could be directed towards extending the universal standard to include additional types and sources of real-world data such as images and free-form text as well as data created by individuals and prospectively captured real-world data.

For more information, please contact Stephanie.Reisinger@evidera.com.

REFERENCES


