Getting real-world data into clinical research

Recent advances in healthcare data access, technology and regulatory guidance have created an opportunity for the biopharmaceutical industry to use real-world data (RWD) in clinical research, with profound impact. The question is, how?

parexel®

Integrating real-world data into clinical trials can dramatically accelerate the pace of innovation in drug development. Leading life science companies are exploring ways to design hybrid trials, use RWD to complete retrospective studies on sub-cohorts of patients, and to conduct long-term safety and surveillance monitoring. While there is tremendous excitement about the potential, anyone in clinical operations can tell you one thing: it's not easy.
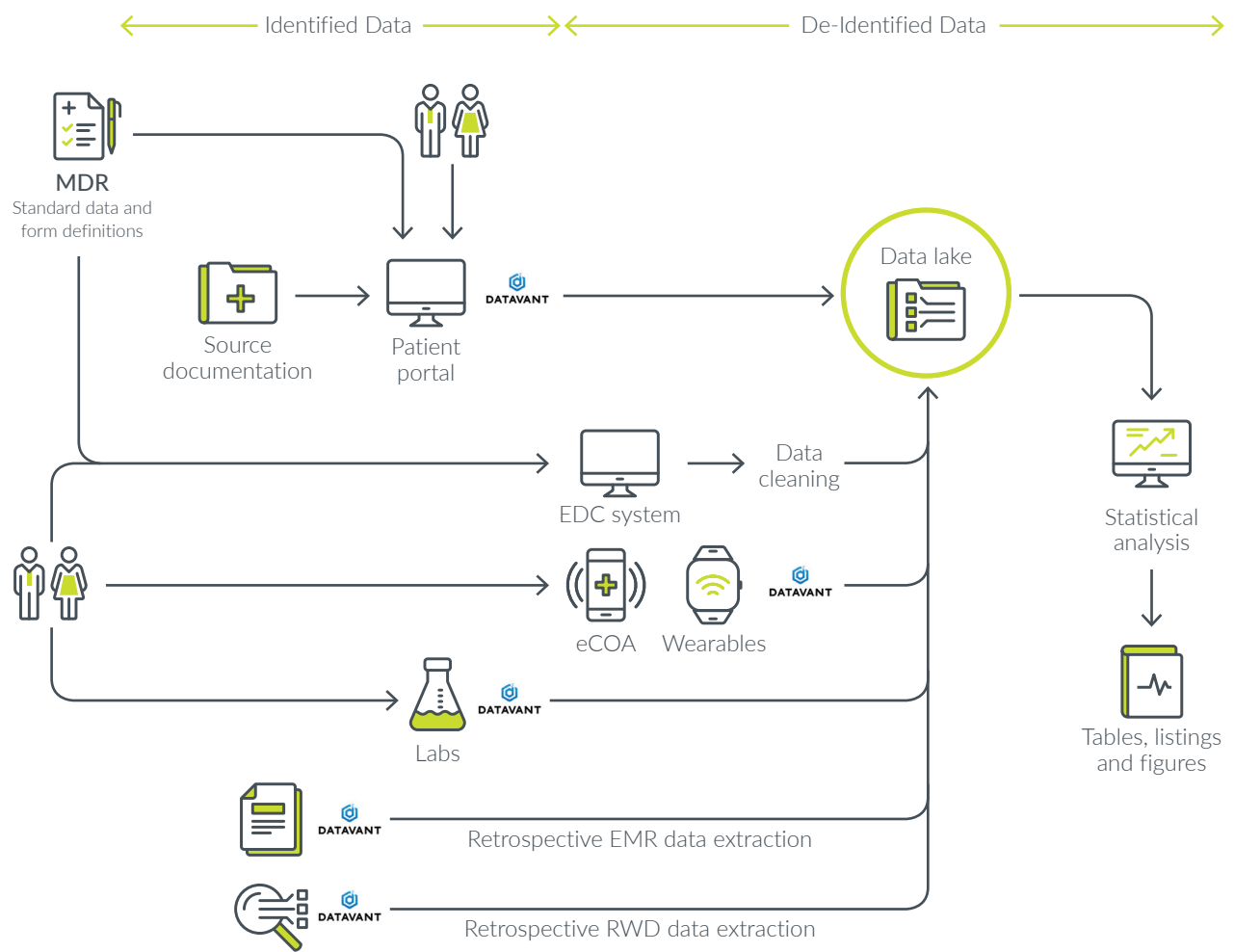
All of us in the life sciences have heard a lot about the promise of RWD, but the key question is often left unanswered: "Sounds great, how does it work?"

We set out to demonstrate the operational feasibility of a trial that incorporates RWD. Specifically, we wanted to show that we could:

> **Use multiple disparate data types:** Incorporate data beyond the electronic data capture (EDC) system used in clinical trials, including patient reported outcomes and wearables data

> **Protect patient privacy:** Connect record-level data from multiple systems while protecting patient privacy

> **Support scientific inference:** Ensure the data was gathered and connected in a way that could support scientific inferences in appropriately powered studies

Data has utility beyond the completion of a study to answer an array of clinical and commercial questions. With our operational feasibility study, we achieve an integrated view of the patient journey that can extend post-endpoint collection.

An Integrated Patient Journey - A Proof of Concept Study

Identified Data ⟷ De-Identified Data

- MDR — Standard data and form definitions
- Source documentation
- Patient portal — DATAVANT
- Data lake
- EDC system → Data cleaning
- eCOA, Wearables — DATAVANT
- Labs — DATAVANT
- Retrospective EMR data extraction — DATAVANT
- Retrospective RWD data extraction — DATAVANT
- Statistical analysis
- Tables, listings and figures

Designing the real-world data study

Real-world data (RWD) refers to data collected in the ordinary course of care rather than in the context of a clinical trial. RWD is heterogeneous and can come from multiple disparate sources. Getting these data into a usable format – accessing, integrating, cleaning and standardizing the data – is challenging. However, the payoff is significant. Incorporating RWD allows leading life science companies to develop a more holistic view of the patients they study.

We designed a study that would incorporate record-level data from diverse sources:

> **Electronic data capture system (EDC)**, the system through which data is ordinarily captured at clinical trial sites

> **Electronic health records (EHR)**, which includes data gathered in the ordinary course of care

> **Electronic clinical outcomes assessments (eCOA)**, which allow patients to report outcomes directly

> **Patient wearables and sensors**, specifically a continuous glucose monitor

> **Claims aggregators**, which are useful for capturing both billing information and patient outcomes beyond the clinical trial

> **Labs**, which contain valuable diagnostic information

> **Consumer datasets**, which contain valuable detail on the social determinants of health

With a focus on proving feasibility, we enrolled eight individuals into the study and incorporated data on those patients from all these different sources.

Successfully operationalizing this RWD study requires a number of key steps:

1. Clear RWD standards

2. Mapping of EHR data to EDC data

3. De-identification of RWD

4. Record-level linking of de-identified RWD

5. Data ingestion and storage of diverse RWD datasets

Defining data standards and mapping to those standards

The first step to conducting an effective RWD study is to implement clear RWD standards.

We leveraged Parexel's Metadata Repository (MDR) to enable a RWD platform for transforming our data sources to a common data model. Utilizing Parexel data standards, we then mapped the data to the CDISC compliant Study Data Tabulation Model (SDTM). The CDISC SDTM defines a standard structure for submitting clinical and nonclinical study data to the FDA (or appropriate agency) as part of a new product application.

We leveraged SUMMA™, Precision Digital Health's regulatory compliant RWD Platform. As the RWD platform transforms data into an integrated model,

it maintains complete traceability, as required by Good Clinical Practice (GCP). The platform also supports non-EDC sources, such as sensor, health & prescription claims, lab and EHR data, and would apply the established transformation rules as the data is ingested and characterized.

The importance of data standards only grows as the number of data sources increases. And because RWD is not collected with clinical research in mind, standardizing and mapping the data can require significantly more effort than a typical EDC system.

In our feasibility study, we used Parexel's MDR to create data specifications for RWD, EDC, EHR and Lab data. This created an information bridge between RWD, EDC Data, Lab Data and standards (Pharmaceutical, Industrial, Clinical). The MDR holds all mapping instructions from these components into industry standards such as CDISC, long before the first data is collected. The MDR houses reusable implementations and instructions to build a bridge from study implementation to analysis enabling faster study deployment and innovation.

Linking RWD while protecting patient privacy

After first defining the data standards and second, mapping the different RWD sources to those standards, the third step is to de-identify and connect the data at the record level.

In other words, you need to know that John Smith's EHR record and John Smith's patient-reported outcome in an eCOA system correspond to the same individual, without having any way to trace that record back to John Smith. This may sound like a simple problem, but it is technically complex,
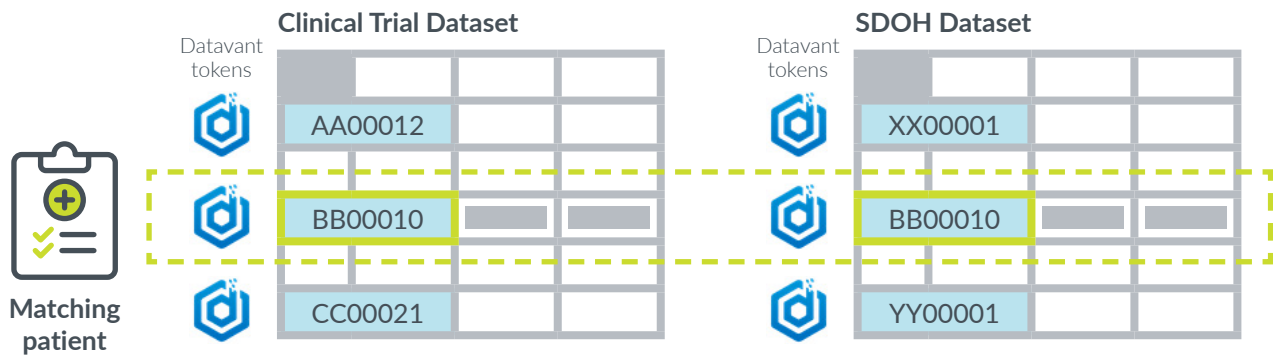
and healthcare and life science companies have struggled to do this well for decades.

Connecting de-identified records requires transforming identifying elements like first and last name into a form where it is impossible to re-identify them, but where those elements can still be used to connect records across data types. This can be done by hashing and encrypting the information to create a "token" that can be used to link data across datasets. Tokenizing patient data is an industry best practice, and Datavant's de-identification and tokenization software can be used across all data types to support this linkage.

Here is a simplified view of what linking two de-

identified patient records might look like:

Clinical Trial Dataset

Datavant tokens: AA00012, BB00010, CC00021

SDOH Dataset

Datavant tokens: XX00001, BB00010, YY00001

Matching patient

Patient records are first assigned anonymous IDs ("tokens"). The tokens can then be used to find relevant records for the same patient across multiple datasets. The sophistication of matching logic can be scaled up or down based on your use case.

Data ingestion and storage to support real-time data access

In contrast to traditional studies, where data is often collected at discrete points in time at clinical trial sites, some RWD contains data that is generated and updated much more frequently.

This is particularly apparent with connected sensors and medical devices, which can generate massive amounts of data on an ongoing basis. Participants in our study wore a continuous glucose monitor, the Abbott Freestyle Libre Pro, for five days.

Continuous data can be extremely valuable, allowing investigators to react to rapidly changing real-world conditions, such as potential adverse events. However, there is a catch: to benefit from continuously created data, you must also be able to ingest, process and store that data to support near real-time access.

All of the pieces matter

Our feasibility study was designed to show how to address all of the operational challenges that come

with RWD, allowing you to focus on the opportunity that comes with being able to map an integrated patient journey.

The success of this study depends on a few key factors:

1. Data standards and mapping

2. The ability to make record-level connections across datasets while protecting patient privacy

3. Data infrastructure to support near real-time ingestion, processing and storage

4. Selecting world-class partners

There is more health data available than ever before. As an industry, we are focused on only a small fraction of the data that will determine health outcomes.

Randomized controlled trials will remain the gold standard in scientific research. But as both the agencies that govern product development and the industry itself evolves, telling a strong value and safety story will soon come to require the effective use of RWD in clinical development.

Please contact us if you would like to learn more.

About Parexel, Datavant, and Precision Digital Health

Parexel

Parexel is focused on supporting the development of innovative new therapies to improve patient health. We do this through a suite of services that help life science and biopharmaceutical customers across the globe transform scientific discoveries into new treatments for patients.

Datavant

Datavant works to reduce the friction of data sharing across the healthcare industry by building technology that protects the privacy of patients while supporting the linkage of de-identified patient records across datasets.

Precision Digital Health

Precision Digital Health accelerates the adoption of digital health for researchers by providing a real-world evidence solution. Our SUMMA™ platform provides a highly configurable end-to-end solution with interoperability to integrate and link clinical data for advanced analytical research and clinical trials.

With Heart

parexel®