



Linked Data Sources Can Elevate Answers from a Data-Heavy Environment to Peak Insight

Enhancing Real World Evidence through Combined Registry and Claims Data

Without Context, Data Drowns Out Insight

Bringing together data collectors, data analysts and physicians using the data offers a peek at how to make sense of our data overload. In a recent conversation, these panelists discussed the benefits of combining deep clinical data with claims data to answer key questions across HEOR studies and medical research in complex patient disorders and shared their perspectives about the evolution of linked data:

- **Su Huang**, Head of Data Strategy, Datavant
- **Scott Robinson**, Vice President of Health Economics and Outcomes Research, Data Solutions team of Inovalon Insights
- **Scott Henderson**, Senior Director of Linked Data Analysis, CorEvitas
- **Dr. Jeffery Curtis**, MD, MS, MPH, rheumatologist and a professor of medicine, Epidemiology and Computer Science at the University of Alabama at Birmingham

They each shared insights and case studies to illustrate how data source linkage has evolved.

Data Rich, but Insight Poor

Many sources of data available, but it's challenging to find the answer to the question with any certainty when you have different restrictions for each. While health care professionals don't generally have access to the same sources, here are some common sources of data:

- large open-source data
- closed claims data
- electronic medical records, or clinically based information from patient management systems (which is at various stages of integration in different organizations.)
- hospital data (can be combined with EMR but also multiple systems for quality, safety, claims, survey data, etc.)
- registry data (which varies over time and topic)
- social media data (also varies per channel, i.e. Facebook versus a patient community)



The Bad News: History of Data is Jumbled, Overprotective and Redundant

Historically, every real-world data company used their own proprietary de-identification and encryption methods. Disparate data sources, ranging from claims and from pharmacy or medical claims electronic medical records, hospital data, registry data, and now even social media, have been utilized in observational research and for analytics.

Over time, joining disparate data sources allowed for the combination of data, with some challenges, and progress was slow due to the number of sources and the intricacy of their data structures and proprietary methodologies. Due to more recent demand for the combination of these data sources outside of those disparate data sources, companies such as Corevitas, Inovalon, and Datavant have created new ways to combine datasets for a real-world perspective. Using tokenization services and secure token exchange is a way to allow easier transmission of information to connect these sources together across different organizations.

The Good News: The Future of Data is Streamlined, Open and Sortable

While large open-source data can be information-rich, these frequently capture only data on a specific focus area. For example, only prescription claims, office claims, or consumer data, which are useful for answering very broad-based questions, but do not provide a comprehensive picture and can potentially be misleading if key information is missing.

“Adding in closed claims data is going to be a little bit more robust than larger open-source claims, and cover activity for a particular patient population as long as they're insured. However it doesn't necessarily have clinically-based information, and that information within patient management systems might be limited to individual physician groups or healthcare systems, which may take some processing to access,” said Scott Robinson (Inovalon).

Even adding in specialized systems and registry data may not allow for the right information to be captured, as data might be collected prospectively using standardized questions by clinicians and patients to collect data for a specific purpose.

“And as an example, Corevitas, they have data on comprehensive granular information around safety data that might also include data on serious adverse events and data

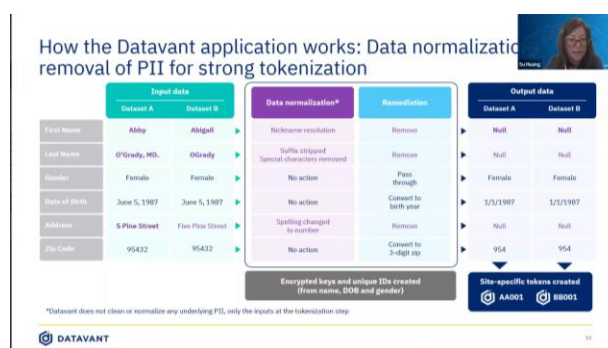
that's captured on a routine basis, based upon clinical encounters with a healthcare provider, more specifically a specialist,” said Robinson.

Privacy, Security and Connectivity – Can You Have It All?

Historically, data sources were not linkable because data companies used proprietary de-identification and encryption methods.

“At Datavant, our focus is on being a neutral and trusted platform focus just on data connectivity. We do not aggregate data, and we do not conduct analytics. Instead, we focus on connectivity so that we can be the ubiquitous enabling platform used by everyone in the health data ecosystem, both data sources and data users,” said Su Huang (Datavant).

Huang described how the database application works to de-identify patient information and then apply certain normalization rules to those data elements, such as nickname resolution, stripping suffixes, or special characters. These input elements are encrypted to create unique keys or tokens that represent a person, with each organization having unique tokens created on their site as an important security feature.



This process has high precision and reliability, with Datavant’s latest cloud match solution resulting in 99.4% precision and 95.1% of recall in performance testing using high-quality data.

“Overall, the mission is to enable easier data exchange and connectivity for more comprehensive data about your patient journey, including use cases from outcome studies to clinical trial protocol design, and rare disease identification,” Huang said.

Designing Real World Evidence Studies using Linked Data: The Key is in the Questions!

Types of questions that can be addressed by the unique combination of administrative claims data and clinical registry data

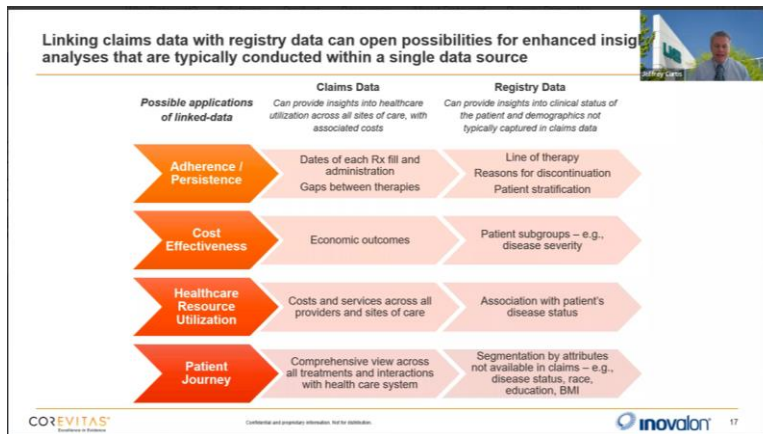
- Are clinical and economic outcomes associated with disease status?
- Does a lower disease activity state impact other outcomes including healthcare services?
- Is lower disease activity associated with fewer ED visits, a lower likelihood of all-cause hospitalization, a lower risk for all-cause mortality, and lower healthcare costs?
- What are the merits of achieving low disease activity targets?
- What is the economic impact of more aggressive RA treatments to achieve more intensive disease activity targets?

Dr. Jeffery Curtis said that the most important consideration in thinking through how to select the optimal data sources is to focus on what question you need answered.

“Not every question needs linked data and for some questions, registry data alone, or administrative claims data alone might be sufficient. However, there are many questions where having linked data sources may be indispensable,” said Curtis. “Specifically the need for clinically rich data suggests that registry data is likely to be valuable. But conversely, administrative claims data may help one supplement registry or another clinical data source to allow for filling in the gaps about what’s happening between registry visits.”

Methodological considerations must be taken into account in using linked data, however, particularly the impact of linkage on sample size, relative size and generalizability, or representativeness. Also, when combining two different types of data, Curtis cautions careful resolution on variables that might violate privacy standards, geography-based variables and data sources that may disagree, such as one indicating a treatment was given while another indicates it was not.

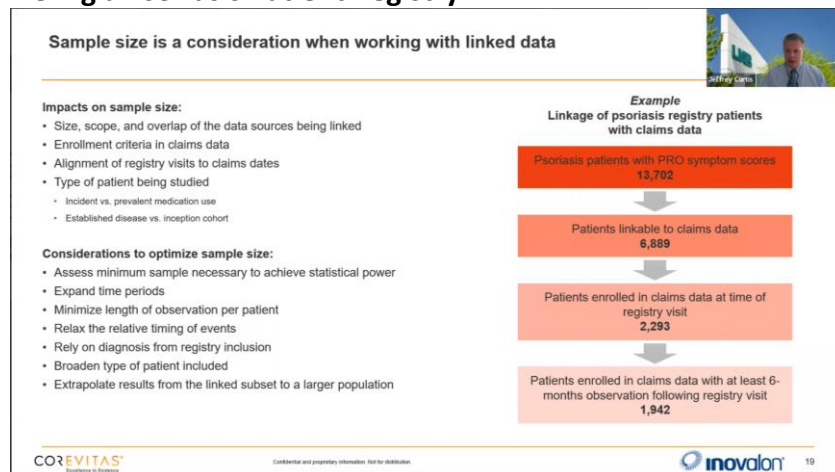
“You have all the richness of both, but you’re going to have to decide how to resolve overlap or even potential discordance between them and to harmonize discrepancies and what your data sources tell you,” Curtis said.



Dr. Curtis points out that frequently people seeking a broad view of linked data may have a preferred source, due to familiarity, but that the benefits of adding other types of datasets can increase insights. Adding claims data to pharmacy refill data allows for an understanding of treatment pathways, including medication possession ratios, and registry data can capture the lifetime history of past treatment. Registry data can provide disease severity, biomarker results, phenotype lab results, and also insight into the reasons for discontinuation provided by doctors and patients, which is often lacking from pharmacy data alone. Health plan data is excellent for deriving insights into economic outcomes and health care utilization, and claims data can complement it.

“Linked data gives you a comprehensive view across all provider types and points of contact with the healthcare system, not just the specialists who contribute to the registry information. As you can see across these several examples, linked data provides an excellent opportunity to augment your analysis conducted with one data type and enrich it with the unique information present in another data type,” said Curtis.

Case Study: Funneling a Psoriasis Patient Registry



Dr. Curtis explained that the filtering of patients by criteria allows flexibility in the time periods as well as the disease inclusions because the data is coming from specialists, not generalists.

The specificity of the data sources can offer more reliability, and both data sources may not need follow up simultaneously which can be more efficient.

"The point is, is that what some of you might often require for a claim all the analysis could perhaps be relaxed once you join it with the registry data which allows you more flexibility and increases the sample size and person time for follow up once you perform," said Curtis.

Case Study: Rheumatoid Arthritis Clinical Application

In an effort to understand whether disease activity in rheumatoid arthritis was associated with adverse events, a question was designed to consider hospitalization, emergency department visits, mortality rates and lower healthcare costs. Data was linked from a CorEvitas registry, Medicare Fee for Service claims data and disease activity measured using the clinical disease activity index (CDAI). The CDAI is a variable unique to rheumatoid arthritis, which measures disease activity in RA, and would not be found in claims data alone.

"In fact, there's no clinical disease activity measure I would ever expect to find in claims data by itself. And so this analysis wouldn't have even been possible without the use of links to clinical plus administrative data," said Curtis.

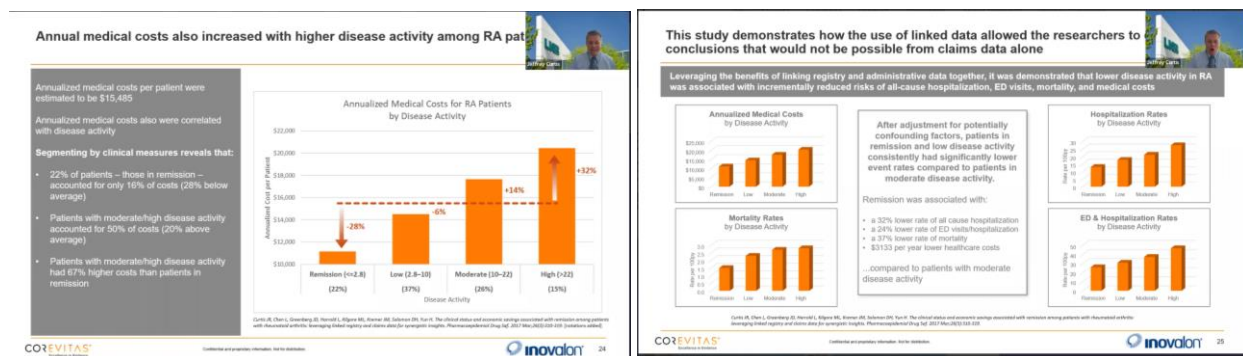


Figure 1 (left): The rate of hospitalization for these patients had a strong relationship with RA disease activity, with a 44% increased rate of hospitalization compared to the average, with those in remission had a 31% lower rates of hospitalization. **Figure 2 (right):** Similarly, the medical costs, which were obtained from the Medicare data were strongly associated with disease activity, with 32% higher costs while patients whose RA was in remission had to cost 28% lower compared to the annualized average cost per patient.

"This difference between patients in the highest versus the lowest disease activity categories with approximately \$10,000 per year, and you would have entirely missed this if you didn't have the clinical data available from the CorEvitas registry about their disease activity, which was represented in a time-varying fashion, and updated at every registry visit," said Curtis. "We could not have tied RA disease activity to the healthcare and utilization and the economic outcomes that we examined couldn't have been studied in this way without the length registry data, given that they would otherwise lack clinical information."

The importance of keeping RA patients in remission rather than allowing for increasing disease activity was proven by [this study](#) published in Pharmacoepidemiology Drug Safety, thanks to the linked data available by leveraging linked registry and claims data.

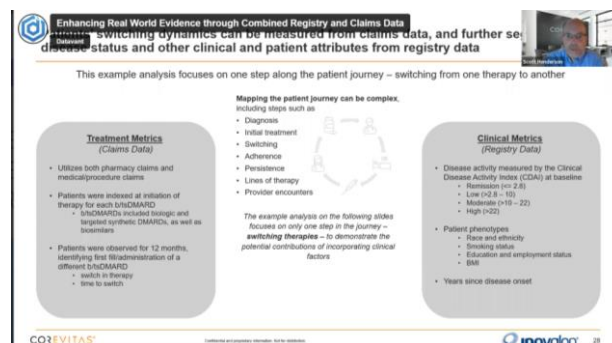
New Possibilities through Data: A Patient Journey Case Study

A patient journey can be complex from initial diagnosis to initial treatment, changes in therapy, appearance, and other interactions with the healthcare system. Not all patients follow the same pathway. With more information, we can better understand the different pathways patients follow. Claims data, and in particular, closed claims data, can be good sources to understand patients' interactions with the healthcare system.

"Claims data often are also very large and often projectable to a broader population, but lack clinical insights about the patient. Registry data, on the other hand, is a rich source of clinical information captured about patients, including patients' disease severity, their outcomes and other demographics," said Henderson. "Registry data captures information about patients' treatments, but can lack visibility to patients' interaction with other sites throughout the healthcare system outside the registry site."

The value of syncing the data together is that information not typically found in claims data can be collected, including pharmacy and medical claims, demographics, and disease activity.

"For RA, some therapies are administered or infused, in a medical office, not at a pharmacy, which would show up as a medical claim. So it was important to ensure that we had pharmacy and medical plans and that our patients who were enrolled in plan types we then link that claims data to the CorEvitas registry data to capture the patient's disease activity at baseline. Things like race, smoking status, education and employment, BMI or body mass index, things that would help us further segment and subset our patient billboard through their insights, were included," said Henderson.



This study identified patients who started on biologic therapy, and indexed them on their first claim for that therapy. Those patients were followed for 12 months to see if and when they switched to another therapy using claims data. A typical lookback period (or washout period) was used to ensure that they were new to that therapy, and claims about therapy previous to

that effect were not included. Patients were also confirmed as enrolled in the health plan for both the lookback period and the observation on follow-up period.

See More Clearly with Registry Data

Drilling into the data with different sub-segments of patients allows understanding of the different pathways that patients are taking and why they may be taking different paths throughout their treatment.

“As you’re designing your research studies, rather than thinking and trying to understand the optimal, real-world data source that would be best suited for your hypothesis, start to think about what combinations of data sources might be even better suited to answer your questions and may be able to gain more insights into your analysis. And even better, are there questions that you can now ask that couldn’t be answered by one source alone, and can now be addressed now that we can link multiple sources together and get richer information about patients,” said Henderson.

Case Study: RA Biologic Treatment Switching Rates

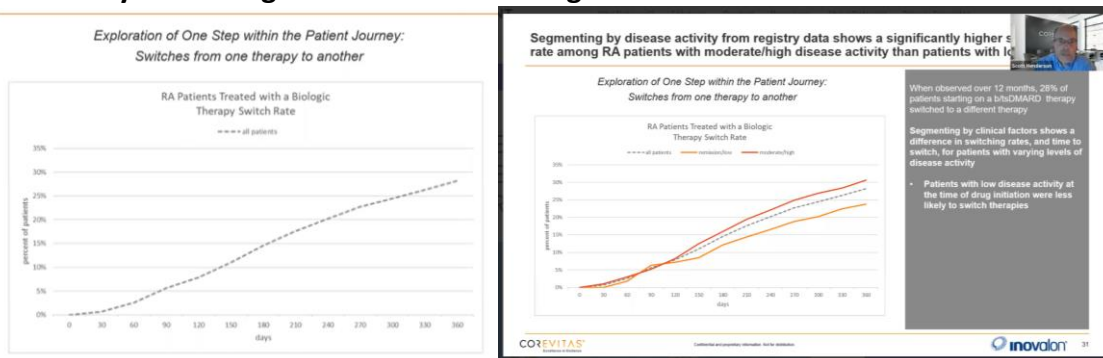


Figure 1 (left): These are the results of what was previously available, about RA patients who switched treatments prior to linked registry data, which is helpful but not nuanced. Figure 2 (right): With registry data, it indicated that RA patients who had moderate or high disease activity are more likely to switch to another therapy and to switch earlier.

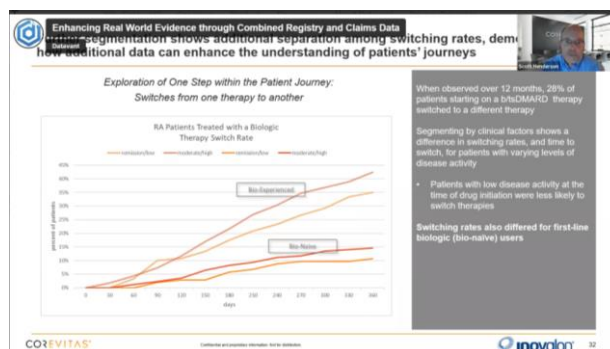


Figure 3: This is the most robust information on patient’s treatment switching behaviors. With continued exploration, separating patients based on their treatment experience indicated that those naive patients or first-line patients were less likely to switch. Patients who had

experience with other biologic therapies and needed to switch therapies again are **more likely** to switch again.

Conclusion: Linking Data Makes a Chain of Insight

Linkage between administrative and clinically rich data sources such as registries, including the CorEvitas registry, allows for a wide variety of expanded analysis that may not be possible with either of these kinds of data sources alone.

There are many HEOR research questions where having linked data sources may be valuable, and specifically, the need for clinically rich data suggests that registry data is likely to be indispensable. Adding in other types of administrative claims data or another clinical data sources may supplement in order to provide a clearer picture and a better answer to the research question. The industry is moving to make data linkage easier, more secure and able to answer the important questions for our patients, payors, clinicians and stakeholder.

If you'd like to watch the entire webinar, register for free at the link here: [Enhancing Real World Evidence through Combined Registry and Claims Data](#)

Got questions about the process? Got research questions you'd like to answer using linked data? Want to discuss gaining access to our registry?

<https://www.corevitas.com/contact>

CorEvitas is the built-for-purpose provider of gold-standard real-world evidence. CorEvitas brings 20 years of experience in specialized patient registries and curated data sources, building on our strengths in comparative effectiveness research and drug safety studies.

www.corevitas.com