



LIISA EISENLOHR
Guidehouse Inc.

Liisa.eisenlohr@guidehouse.com

Liisa Eisenlohr is an associate director at Guidehouse within the Life Sciences Medical Affairs Center of Excellence, with more than 20 years of experience within the life sciences industry.



BILL WOYWOD
Guidehouse Inc.

William.woywod@guidehouse.com

Bill Woywod, MPH, MBA, is an associate director in the Life Sciences practice at Guidehouse, where he leads big data analytics and machine learning projects. He has played a central role in expanding Guidehouse's deep data analytics capabilities, including the development and execution of sophisticated claims data analyses and machine learning initiatives for clients in the medtech and biopharma industries.

THE POTENTIAL PITFALLS of Real-World Data Analysis

Real-world data and evidence can help medical product companies understand how products are really being used, knowledge that can feed back into many aspects of their businesses. But beware the pitfalls of amalgamating, curating, and analyzing data sets at this early stage in the field. Consultants at Guidehouse tell us what to watch out for.

>> Liisa Eisenlohr and Bill Woywod

As medical technology companies increasingly seek to harness the power of data analytics to gain a deeper understanding of their industry and its competitive landscape, a number of third-party analysts are emerging. These external partners aim to mine, curate, and analyze the growing wealth of available data to provide insights in support of such major efforts as launch planning and execution. They often tap into real-world data (RWD) to produce real-world evidence (RWE).

RWD is data relating to patient health status or the delivery of healthcare routinely collected from a variety of sources, such as registries, collections of electronic health records (EHRs), administrative and medical claims databases, and, increasingly, personalized digital health data. Analysts use RWD to derive RWE, that is, clinical evidence about the usage and potential benefits or risks of a medical product.

The analysis of real-world data can lead to improvements in patient care by helping providers correlate outcomes to various patterns of patients and treatments. It can also help medical products companies and regulators understand how products are really being used, knowledge that can improve clinical development planning (e.g., by suggesting potential additional indications), underpin dossier submissions for regulatory approval, strengthen the case for reimbursement, inform medical

communications, and bolster market share. Furthermore, as part of the 21st Century Cures Act of 2016, the US Food and Drug Administration uses RWD and RWE to monitor postmarket safety and adverse events and to make regulatory decisions.

Many companies will want to turn to third-party RWD and RWE vendors, of which there are hundreds. There are advantages to outsourcing this expertise, but also risks. At this relatively early stage in the field, many potential pitfalls lie in the data sets themselves, especially given the power of artificial intelligence (AI) and other analytic techniques to magnify errors and create biases.

We share here insights on some of the primary associated risks, as well as key advantages to enlisting third-party analysts.

Associated Risks

Crucial considerations come into play when benchmarking these types of data, from consolidating various and oftentimes unstandardized data sources to the quality of the data itself, and whether it can support defensible conclusions. In addition, there are compliance and governance requirements around the use of medical data. Let's tackle these one by one.

1 Consolidating sources and data quality

Merging datasets is more complicated than it might appear. The sets being merged must be consistent in quality, as well as cleansed and standardized. Called “deterministic matching,” this involves rigorously matching like data elements to like data elements as precisely as possible. If the quality of the data is inconsistent across sources or the data itself does not match, then the output will be error prone. This fundamental activity is easier said than done, especially when working with multiple, disparate data sets, which may be vastly different in the ways in which the data was collected, stored, and shared.

For example, let’s say that at the highest level, your company is interested in learning the prevalence of co-morbid characteristics to size the patient population that has the potential to benefit from your therapy. Your external partner starts with EHR data feeds from a variety of providers and medical claims to determine how many patients with those certain conditions were treated in the past year. Now suppose that the analyst does not have access to the source data. It seems straightforward, but for a number of reasons it’s not, including the likelihood that some records are duplicative across data sources, meaning patients are cited in both sources, thus inflating the figures.

Likewise, the findings could be missing important data elements. For example, collectors do not necessarily use the same conventions when it comes to creating categories, and what can be collected or shared varies widely in the collection entity and the governing country’s data privacy and sharing laws. Consider that some nations do not have routine, systematic ways of collecting data, and some will not share it across borders.

2 Data Compliance

Compliance is another critical issue, especially depending on the target market. Currently, about 66% of countries have data privacy laws and 19% have no legislation. For the sake of reference, to regulate patient records and sensitive information, the US has the Health Insurance Portability and Accountability Act of 1996 (HIPAA), the EU has the General Data Protection Regulation (GDPR), Canada has the Personal Information Protection and Electronic Documents Act (PIPEDA), and so on. Conversely, China does not have a single, comprehensive data protection law.

Furthermore, the process for collecting data varies widely, and some nations (about 5%) don’t even collect data. So, depending on the planned target markets, getting a clear view of the potential market opportunity through RWE analysis could range from daunting to impossible due to the lack of data or the regulations governing the data.

Consider the following challenges. In the US, the HIPAA Privacy Rule establishes national standards to protect individuals’ medical records and other personal health information. This means any identifying patient information (e.g., name, social security number, address, etc.) must be removed from data before analysts can access it. Healthcare claims are among the most common types of data available for analysis. Multiple sources generate this kind of data, including public and private payors, and usually require a license for access. So, many compliance issues can quickly arise.

As mentioned above, the amalgamation of data sets from different sources can result in duplicate records. To remove duplicates, the analyst would need to enable the program to allow for personally identifiable information. For that to happen, legal measures would

need to be taken and put in place to compliantly allow the information to be shared. For example, in the US, as noted, the data would need to comply with HIPAA, or any rules that apply to data from other countries. While this is possible, niche third-party analysts that specialize in governance and maintenance typically are relied upon for this kind of data collation.

Even with the proper legal mechanisms in place, merging data sets requires due diligence. The data sets must be put through a coding mapping process to “crosswalk,” that is, translate equivalent, identical, or similar information. Done incorrectly, the analysis will be based on pooled data sets that are at least partially mismatched or unmatched. Also, logic should be applied in the programming to establish the consistency and homogeneity of the different data fields. Without these steps, the accuracy and reliability of the analytical output is highly questionable.

3 Data Governance

Data governance brings another risk. According to the Data Governance Institute, data governance is “a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods.” In simpler terms, governance dictates who is accountable for the data assets, who can access the data, and how it can be used. Data governance establishes these “rules” to help ensure the integrity of the data and mitigate risks, such as those associated with compliance and security.

The following is a real-life case showing how these risks can come together to compromise a company’s business.

Let's say to gain more business and satisfy clients, a data vendor adds a new RWE service line. The team successfully pitches conducting a study that combines three data sets—one from a third-party claims database, one using the client's license for a healthcare provider's proprietary EHR data, and one licensed from a private insurance company. For the external analysis, they hire an outsourced subcontractor to remove the identifying patient information from the proprietary data set, and merge all three sets together, without accounting for duplicate patients across data sets. They enlist internal team members to analyze the results, and eventually share the insights gleaned with the study sponsors. Then they store those results on an unsecured server.

What's wrong with this scenario? The external analyst breached the primary use license the client had with the healthcare provider, used duplicative data, and broke both strict compliance laws and established best practices. They also based findings on methodologically unsound data aggregation approaches.

Enlisting Third-Party Analysts

Realistically, building and supporting a robust in-house data analytics team requires continuous investment, managerial oversight, and organizational responsibility. Despite the potential risks, outsourcing this function can often make the most sense from a fiscal and human resources perspective. It also offers several additional advantages, assuming you are doing business with disciplined, expert data analysts, and we offer here a list of key considerations when deciding to work with an outside data provider (see box, *Key Considerations When Outsourcing RWE*).

Most third-party analysts will develop cloud-based access for the data they house and frequently have an overriding


relational database structure, pulling from a variety of sources and types of data. This relational database construct allows easy indexing within the data, which makes it easier to select columns and filter data tables (e.g., creating a subset data table that isolates patients of interest) to run a quick analysis on those purpose-built data tables. Many also include the functionality for "point-and-click" stratifications to create subgroups based on therapeutic

area, disease, patient history, and concomitant therapies. Basically, these "workbenches" allow you to learn the same kinds of insights as a labor-intensive customized analysis. So, for example, if you need a rough estimate of customer switching behavior, you can speedily drill down to figure out how many and where.

Furthermore, in these cases, the external analyst will have addressed any licensing issues, and, if need be, can negotiate third-party access to specific data sets you might require. This saves you from managing multiple agreements.

For example, in one scenario, a global life sciences company had a team of 15 analysts. The team was responsible for all incoming data, end to end, including licensing agreements, storage and backup, governance, and compliance. The team, which was also responsible for processing and managing all analytics queries coming in from the various business units, often had a backlog, which resulted in delays to key business decisions.

This company could have safely and more easily accessed data and insights through a qualified, experienced third-party analyst. Specialized vendors have the capabilities to provide advanced analytics techniques and information management systems that companies may lack, and as such, partnerships with the right qualified vendors can offer many benefits. These benefits include avoiding the costs and hassles of staffing and supporting in-house resources; gaining access to easy-to-use, speedy platforms; and, with a well-vetted vendor, the confidence that data is handled according to the relevant regulations.

To safeguard your business, product development, and commercialization strategies, it is crucial to validate the integrity and rigor of your third-party vendor's data and analysis expertise, processes, and mechanisms. 

Key Considerations When Outsourcing RWE

- How do I develop a unified corporate data strategy that supports RWE?
- How do I select the right RWE vendor from among the large number of potential options?
- How do I ensure that I am using the correct data sources to address my business questions?
- How do I ensure and verify that my merged data sets are complete and of good quality?
- How can I make sure that the data is being collected, stored, and used in a compliant and secure manner?
- How can I best leverage RWE for HEOR (health economics and outcomes research) analyses that directly lead to positive business outcomes?
- How do I leverage RWE in innovative ways, beyond HEOR? Can it be applied to functions such as clinical development or medical affairs?

Source: Guidehouse Inc.