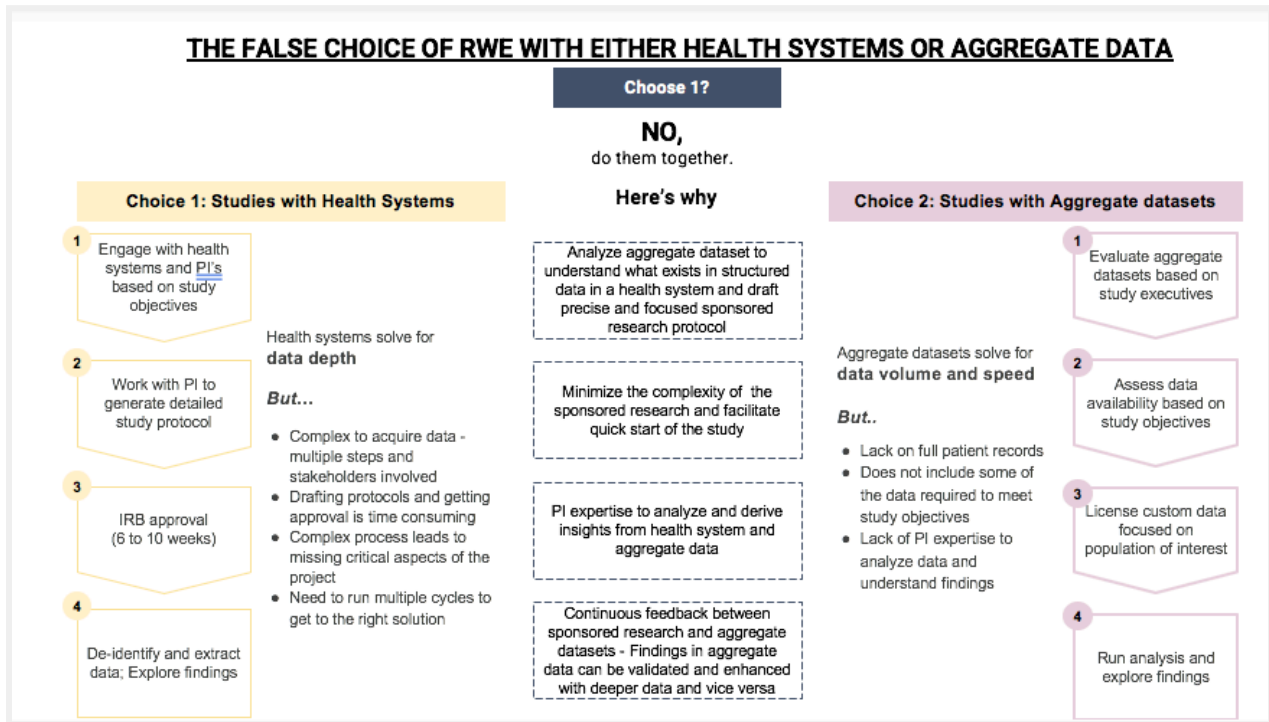


Bundling Big Data and Deep Data to Solve Hard Problems in RWD



The hard-to-reach data is normally of central concern in our dialogues with clients. For example, the questions for a dermatology product indicated for Prurigo Nodularis turn out to be critical to know whether patients are itchy. A celiac study needs extensive labs and longitudinal histology data to understand gluten-free exposure and progression. Diabetic retinopathy studies need ophthalmology images. Cancer studies want CT scans. An enzyme replacement therapy on the market for 30 years wants to know whether patients with their enzyme have fewer side effects. A narcolepsy drug developer wants to know about daytime sleepiness and cataplexy episodes. These pieces of data are not easy to get because, if they are documented in the EHR, they are normally in free-text notes and rarely extracted into structured data. If they do exist, they often are not aggregated into scale data sets.

Because it is so hard to find a matching data set and we like to make good decisions, we have set-up choices about how to obtain retrospective data needed for research. The typical choices framed by clients are to decide whether to license an aggregated data set or to work through sponsored research with health systems. This is a false choice, or at least it should be. No one



can find the perfect data set for your research. That is a fool's errand. However, you can be strategic in how to approach data sets to be more efficient to get more answers faster. The two methods of engagement – aggregation and sponsored studies – are complementary and should be executed as a joint approach to get the most value out of either investment. RWD acquisition can be a 1+1=3 situation.

Health systems solve the age-old problem of getting to the full patient record. They can do so because, whether it has ever been made available or not in aggregate forms such as a data warehouse, they do have the data from all of the different systems that collect information about patients. Beyond that, they also employ physicians, informatics staff, and data stewards who know where the data is and how to extract it. Here is the challenge with the deep data sets with health systems: even with all the work we put into streamlining the process, it is complex to acquire data because there are many steps to plan and execute a study. A principal investigator needs to generate a protocol, which needs to be priced by the IT group, and that needs to be approved through the IRB process. Then the IT group needs to source data. When all of those steps are part of a single cycle, you may miss critical aspects of the request; so the cycle needs to be iterative 2-3 times after the initial data request to get to the right solution.

Why did Graticule partner with IBM Watson Health™? The answer is because Watson Health has made it possible to analyze large-scale (50M+ patients) aggregated, deidentified patient data sets. This includes both a leading claims data set, MarketScan®, and a leading EHR dataset, Explorys®. These are complementary to the approach we are taking with advanced datasets through health system engagement.

In contrast with sponsored research, data aggregators offer a licensing model for RWD like MarketScan or Explorys that allows for transactions. You can ask for a quote, get the cost for licensing data for a study, and then license the data for a fee as soon as you have budget approval and get the research started. There are no personalities to manage, no protocols to carefully craft and propose to the IRB at 5 institutions, no de-identification logic to review. If we could do every study with aggregate data, we would. Furthermore, the licensing can scale up or down to the specific project needs. For example, a study can license a small cut of Explorys, MarketScan, or a combined data set focused on just the narrow population of affected patients needed for the study. Because the total volume of data is so large, it is more likely to get to the critical mass of what you need to power the statistics, but some data we need is not there.

The compromise in the 1+1=3 situation is to work in parallel. The aggregate data sets provide breadth and speed. Many questions can be answered with that data that may never be answered with a health system, such as 'how much does this type of patient cost in claims data in the year following their index diagnosis?'. You also can begin to build models with this data knowing that certain fields are missing that are important to the model. That modeling can



provide important insights into how to design the data extractions and how to focus areas in sponsored research. How can you know what you are missing until you know what is in the structured data in a typical health system? The answer is to analyze the EHR data from Explorys and use it to understand what can and cannot be determined without the extended data. Then you can make intelligent decisions in designing the study and addressing data needs with a health system partner.

The health system partner model can also add significant value to analyzing aggregate data. Because the sponsored research study always includes a principal investigator who is likely a clinician working with the patients and data, they know how they code patient records in EHRs and how the billing represents their practice patterns. They also know what and where to look for insights about care because they have intuition based on real knowledge of how patients are treated, their journey in the health system, and where to look for hypotheses of unmet need. A solution that leverages both the breadth of the aggregate data and depth of the health system data enables engagement with clinical experts that drives insight generation and the potential for translation into care.

As we continue to approach clients with these hard problems, we are increasingly offering the option to bundle the two different approaches together to meet client needs. We can build a budget that includes both the IBM Watson Health data and a feasibility study or full research study at a partnering health system into a single project with an integrated plan and team focused on answering the highest value questions for the client.

Interested in learning more?

Reach out to us at info@graticule.life