The Evolving Data-Driven Strategic Feasibility Model

How democratization of data is powering a new wave of opportunities

Chris Komelasky

Senior Director, Strategic Feasibility
Site and Patient Access

Gaurav Bhatnagar

Vice President, Strategic Feasibility
Site and Patient Access

PPD

Key Challenges

Today's clinical research environment is more competitively challenging, has higher costs and takes longer to bring a drug to market than at any other time in history. According to the 2015 Tufts Center for the Study of Drug Development research, from 2005-2015 the average Phase III study saw an increase in the complexity of protocols, driven by the ever-increasing set of inclusion and exclusion criteria (+61%), number of clinical endpoints (+25%), and number of study visits (+25%) and procedures (+70%), are driving more and more protocol amendments and adding to that cost and time. At the same time, the percentage of patients taking part in clinical studies remains at historically low rates, estimated to be less than five percent of the patient population. This contributes to low and non-enrolling sites, further contributing to the cost and time.

One of the common threads across all these challenges, is that data exists that could help address the underlying needs, but that data is often siloed across people, teams and companies, preventing it from helping to drive the right insights at the right time. By breaking down barriers and democratizing data, we are moving towards a future where that data and powerful analytics will come together to help design and execute clinical trials, bending the time cost curve of drug development.

Feasibility in an Evolving Data Landscape

We continue to navigate the evolving ecosystem that connects the most relevant patient and site/investigator data with targeted processes, analytic tools and expertise to optimize study design, selection of sites and recruitment of patients.

The traditional paradigm of feasibility was tied to a site-centric approach, where we would use historical site performance to drive analysis while also relying on feedback from sites on their patient populations through surveys, often presenting feedback that may not truly reflect an accessible patient population that could enroll in a study. That paradigm has shifted as more and more patient data has become available, allowing a more patient-focused approach to come forward.

For years, sponsors and CROs have augmented that site view with de-identified patient data through prescription sales, or pharmacy/medical claims data sets. While these data sets provided more accurate views of the patient populations in a given area and a directional improvement in showing the patient density, they have continued to have significant limitations. Those limitations start with the global coverage of those data sets and depth of data attributes that are available to match all the inclusion and exclusion criteria.

More importantly, the data privacy requirements for these data sets keeps the specific patient information blinded so that any particular patient can never be matched and connected to a clinical trial.

The next silo of patient data exists in medical practice Electronic Medical Record (EMR) systems that capture patient information at specific points in time. These structured data sets have the ability to capture rich data about the patient, but lack standardization across health systems and global geographies and often cannot provide a longitudinal view of a patient's medical history. Because these data sets are often captured at the site of care, the health care professionals at these locations can both search these systems to identify patients and communicate with that patient about how they may be a fit for a clinical trial. The latest technology improvements in areas like tokenization, are allowing a bridge between previously siloed EMR and other deidentified data sets (e.g. claims), combining the available data and giving a much deeper view of a patient's medical history.

In the last few years, federated networks of EMR systems have launched to allow this aggregated EMR data across health care organizations to be centrally searched and analyzed. While data access through these platforms is de-identified when centrally queried, the local site teams have the ability to see the resulting patient matches and recruit these patients to their studies. The industry has coined the term "pre-identifiable data" to convey the fact that, while the sponsor/CRO cannot see any identifiable patient data, the site teams or their health care professionals can, addressing one of the main gaps with the other deidentified data. Another example of this occurred in a study for a rare pulmonary condition, where a site had not recruited any patients, and once we helped identify a pool of patients using this technology, they went on to have 3X the average enrollment rate for the study. In an era of rapidly advancing precision medicine and advanced therapies, similar pre-identifiable data coming from lab and genomic testing partners are providing a compliant pathway to match down to specific gene mutations.

A key evolution in the data landscape is the introduction of natural language processing (NLP) which allows unstructured data and documents, such as free-text physician notes or a hospital discharge statement to be analyzed. This unstructured data can then be parsed and matched using machine learning and advanced analytics so that key data points can be put into a more structured format. This advancement, while still coming forward, is opening up another previously siloed data area and allowing more data to be combined in our medical records.

The ideal data set to both match patients and have an ability to directly communicate with them is a fully identified patient data set. While it comes with a host of data security and privacy challenges, it gets us closer to enabling and empowering consumers to be informed and drive decision making about their medical treatment options. Today, we are able to leverage patient consented databases to model trial recruitment and directly connect patients to potential studies.

Democratization of Data

This space is at a tipping point where over the next few years, we will only see an acceleration of the availability of new data sets that will prompt teams to evaluate their thinking of how to best access and use the data to drive new insights.

According to Forbes' Bernard Marr,[1] "data democratization means that everybody has access to data and there are no gatekeepers that create a bottleneck at the gateway to the data. It requires that we accompany the access with an easy way for people to understand the data so that they can use it to expedite decision-making and uncover opportunities for an organization. The goal is to have anybody use data at any time to make decisions with no barriers to access or understanding." We believe in this view of democratization and think the healthcare space and clinical research industry is moving down this path.

The first step towards this vision involves breaking down internal data silos as an organization and aligning that with syndicated external data sources to ensure teams have access to the right data, at the right time, in the right form to make decisions.

This has involved investment tools and capabilities that has enabled this progress:

- A platform to ingest, transform, standardize and aggregate data sets
- A focus on master data management to support how common data domains are aligned and consumed across systems
- A strong data governance team and set of processes to guide the business rules that drive the data collection, curation, stewardship, and retirement
- A data structure, utilizing data lakes to make the data accessible to both internal systems and external partners
- An analytics platform and set of models that brings the key insights to drive decision making

The next step in the journey towards data democratization involves sharing of data across partners that may have previously closely guarded that data. When we look at the rich set of clinical trial study and performance data, we have access to, it is still only a sliver of the total available data out there with sponsor partners and other service providers having complimentary pieces. To get to a more holistic view of the landscape and remove the inefficiencies and limitations that exist, we need to bring those data sets together in new ways. By coming together in this precompetitive space, various healthcare stakeholders will be investing in an approach that lets us face our common challenge of the data limitations.

Data sharing across partners will introduce some new complexities that need to be addressed in the areas of data privacy and security, data portability and data-use rights, and in the data standardization across partners. Consortiums are coming forward to help tackle these regulatory, compliance and operational challenges but more work needs to be done to accelerate the collaborations.

The last step in the democratization of data in the space will involve bringing the patient and consumer forward in a central way and engaging them. The foundation for this is already happening in clinical trials as wearables, electronic patient reported outcomes and electronic clinical outcome assessment tools are increasingly becoming part of trial designs. As key technology companies like Apple partner with Allscripts/Veridigm to bring electronic health records into the consumer's hands, and Google's Verily progresses its Project Baseline registry and consortium to bridge the gap between clinical research and clinical care, we are entering a space where the blend between data, technology and healthcare will get blurred. The key will be taking the next steps to ensure the patients have their own data, know what it means, how to use it, and how to engage the broader healthcare system to improve the outcomes they prioritize.

Artificial Intelligence and Advanced Analytics Paving the Path Forward

Now that we have data and are finding more and more ways to share data across teams, companies and with individuals, the next question to address is, 'what can we do with it all?'. The answer lies in the advanced analytics powered by machine learning and artificial intelligence. In the past our teams have taken data sets, explored relationships across various attributes and built statistical models, algorithms or simulations to help predict certain outcomes. This can be applied to predict overall study outcomes, patient responses to interventions, study timelines and key milestones, likelihood of a site to be a top enroller, and how to best recruit a specific patient population.

These use cases in drug discovery, study design and planning, site identification and selection and patient recruitment offer the most tangible examples of where these additional data sets and analytical capabilities will create near term impact. Increased global data assets that have real depth across therapeutics areas and indications, will give these analytics the power they need to increase the accuracy and precision of predictions. Collaboration across companies to access those shared data sets and jointly develop and refine the data models will create a unique ecosystem of players that are all able to get better answers to their questions than if they go at it alone.

Today, as an industry, we're taking these data inputs and layering the different dimensions on top of each other to help craft and align our site and patient strategies and also help communicate the rationale for our recommendations to our sponsors. We have a clear line of site to those dimensions that can get tracked back to the underlying data. An important part of moving into an advanced analytics space will be the change management that needs to come forward with it. It will take careful thought and support across the business and technology functions to enable the right mindset. We can't jump from the current linear thinking and approach into a black box recommendation that some of these modeling tools can provide. Instead we will need to ensure we are bringing our internal teams along the journey to give the confidence that these new tools are able to surface previously unidentified correlations amongst the data that help lead to better results. A stepwise introduction and evidence of impact to support the new way of working will go a long way to winning the hearts and minds of internal teams. In our strategic feasibility world, that data-driven approach will mean faster study delivery at lower costs, truly bending the time and cost curve of drug development. That is a game changer and that is where we see the future taking us.

References

1. Marr, B. (2017, July 24). What Is Data Democratization? A Super Simple Explanation and The Key Pros and Cons. Forbes.