Is Data Science the Treatment for Inefficiencies in Clinical Trial Operations?

By Venkat Sethuraman, Aaron Mitchell, Yogesh Sharma, Sharma R D and Jessica Rine
Is Data Science the Treatment for Inefficiencies in Clinical Trial Operations?

By Venkat Sethuraman, Aaron Mitchell, Yogesh Sharma, Sharma R D and Jessica Rine
Half of the total cost of bringing a drug to market is lost to lengthy drug development timelines, and inefficiencies associated with clinical trial design and implementation make up a significant portion of those costs. One key factor that leads to delays in clinical programs is patient recruitment and enrollment. Today, nearly 80% of clinical trials fail to meet enrollment timelines, and approximately one-third (30%) of phase III study terminations are due to enrollment difficulties. Moreover, longer recruitment periods mean longer trials and increased cost. Clinical trials typically last 42% longer than expected in phase I, 31% longer in phase II, and 30% beyond planned timelines in phase III, mostly because of recruitment delays. These costly delays are getting worse. One study showed that between 2008 and 2011, the cost for a patient in a phase I trial increased 33%, phase II costs rose 75% and phase III costs rose 88%. The smallest increase occurred in phase IV, which still saw costs increase 30%.

How can pharmaceutical companies address these seemingly insurmountable clinical trial inefficiencies? A holistic, data-driven approach to clinical trial decision-making can help companies enhance the feasibility process, design trials that allow for more efficient patient enrollment, and run those trials at the right sites that treat patients who are best suited for their trials. While we’ve found that most major pharmaceutical companies are making efforts to align with this type of approach—using real-world and publicly available data to build more sophisticated analytics capabilities—these efforts are often fragmented, implemented inconsistently and difficult to scale. Companies often budget in anticipation of rescuing struggling clinical trials; however, if they applied some of that budget to the design phase, they could save considerable time and resources in preventing such failures. Here’s how a data-driven approach can help improve clinical trial design in six key areas:

**Competitive intelligence:** A thorough understanding of your patient population, and the trials that will compete with yours, is essential during the planning phases. An investigator’s current trial burden can be a strong indicator of how they will perform on your trial. Selecting investigators who are participating in complementary trials (versus competing trials) will ensure that you’re able to maximize your investment in a site. For example, an investigator who has a robust trial portfolio may have resourcing issues with your trial, but if all of the investigator’s trials are in the same disease area, there may be an opportunity to capture a complementary patient population.

Another key factor to consider is the number of trials that an investigator is participating in compared with what epidemiology data tells us about the number of patients available in that region. Using anonymized patient longitudinal data, you can learn how many patients with a specific condition reside in a geography. If the incidence of a disease is low, and there are other trials in that geography, it’s important to take a closer look at how those competing trials will affect enrollment and consider alternate regions or adjustments in timelines.
Country selection: When determining global allocation for a clinical trial, pharma companies need to consider several factors:

- **The population:** Estimates suggest that more than 35% of sites will fail to enroll the number of subjects they indicated during the site qualification process, according to the Tufts Center for the Study of Drug Development. Using electronic health records, claims data and other epidemiology data, you can determine which countries have the highest concentrations of patients eligible for your clinical trial, allowing for appropriate allocation of patients in different countries and sites.

- **Local clinical trial history:** Looking at average site initiation timelines and enrollment timelines for similar trials within your preferred countries will give you an idea of how many patients you can expect the region to deliver in the amount of time you have to enroll patients in the study.

  It’s also essential to consider the cost. Using internal and industry benchmark data on historical clinical spending, you can model the cost of running a specific trial in a country or region.

Lastly, the availability of high-performing sites and investigators is key to country selection. Building a detailed picture of the site and investigator landscape can help inform a robust country selection strategy. Using publicly available data, combined with your own internal data on site performance and industry benchmarks, you can identify the highest-performing centers across therapeutic areas. This will ensure that you’re selecting the right sites in the right countries for the right trials.

- **The commercial and regulatory environment:** Data on the utilization of currently approved products, treatment pathways and paradigms, and reimbursement should be considered when determining country selection.

  Understanding the potential for commercialization and any concerns that should be addressed during product development can eliminate the need for additional and costly country-specific studies. For example, if you were developing a new therapy for arthritis that you expected would become the third product in your class to market, would you invest in clinical trials in a region where the local health authority had not agreed to reimburse for the second product, favoring the first?

  Regulatory factors should be considered in determining country selection, prioritization and planning. For example, study timelines can be significantly impacted by how long health authorities take to review and approve a protocol, as well as the requirements and time for institutional review boards or research ethics boards to complete protocol reviews. It’s also critical to develop a thorough understanding of local regulations. For example, some countries will require that a sponsor purchase any co-therapies involved in a trial design for patients, while others may allow patients to receive the products commercially.
Site and investigator identification: The industry is experiencing a shortage of experienced investigators, and this lack of experience has consequences. According to the Tufts Center for the Study of Drug Development, about 40% of investigators each year choose not to conduct any further clinical trials at a time when typical multi-center studies require an increase of 30% more investigative sites. Moreover, it has been estimated that about 30% of investigators under-enroll trials, and about 20% fail to enroll a single patient.

Most companies have processes that involve reviewing internal data on prior performance with their own sponsored trials, but rarely do companies look beyond that at a site or investigator’s performance with other sponsors. It takes a lengthy 3.2 months, on average, to complete the site identification process for phase II and phase III studies. A combination of internal and external data can be used to speed up that process.

Companies often have their own powerful data on historical site performance that they fail to leverage in making site selection decisions. Data for average startup times can be used to develop a site list that supports an appropriate enrollment cadence. Internal quality data for protocol deviations, average query resolution times and audit findings can be factored in to establish a baseline to determine how past sites will perform on future trials. This data can also be compared with external benchmarks, such as those produced by the Metrics Champion Consortium, enabling an organization to create standards that sites can be held to throughout trial participation.

The probability of effective recruitment and a successful trial can be increased by finding sites with a demonstrated track record of good performance in similar trials and site personnel with strong credentials. This establishes the site’s expertise, their ability to execute procedures similar to what can be expected in your trial, and that they have a patient population to recruit from.

Another challenge of site selection is identifying a site’s qualifications. Many new technology tools are now available for sites to disclose their qualifications and credentials to sponsors and clinical research organizations. This technology can help select the best-suited sites for trials in a simpler and more effective way than traditional screening questionnaires.

Epidemiology: The number of available real-world data sources is rapidly growing and evolving. From commercially available sources like claims and EMR data to patient registries and even social media, most companies utilize this data well in commercial organizations but fail to consider it in the context of clinical trials.

When it comes to traditional sources of epidemiology data, patient density mapping can be performed using anonymized patient longitudinal data, as well as claims data and other sources. The first step is identifying the most robust data sources for each region where you plan to conduct a clinical trial. From there, you can create heat maps of potentially available patients.
Using these data sets, you can perform predictive analytics, which can be particularly helpful in finding undiagnosed and underdiagnosed patients. For example, if you were conducting a clinical trial in progressive supranuclear palsy, a condition frequently misdiagnosed as Alzheimer’s or Parkinson’s, you could identify potentially misdiagnosed patients using predictive analytics. By identifying a set of parameters that may indicate a misdiagnosed patient, you can identify likely candidates and help to inform targeted campaigns that help physicians promote disease awareness.

In terms of non-traditional data sources, automated social listening can help you identify regions where patients are looking for clinical trials. Automated social listening techniques can mine data on social media platforms like Facebook and Twitter, providing services that connect patients to each other. This data, coupled with patient heat mapping, can provide a robust picture of potentially eligible patients for your trial.

+ **KOL influence mapping:** Understanding key opinion leaders in a disease state is critical to the success of a clinical research program. Beyond standard KOL mapping techniques, advanced technology can be used to identify current and future networks of influencers. Through machine learning and AI, you can create a comprehensive influence map of a disease area. You also can create an in-depth picture of current and up-and-coming influencers in your therapeutic area by analyzing their papers and practice guidelines, investigator participation in clinical trials, speaking engagements at conferences, leadership positions within associations and editorial positions in key journals.

For example, clinical trials in certain disease states are particularly challenging because a variety of specialists may treat a specific condition. Finding the key influencers requires a strategy that goes beyond traditional KOL mapping to create a comprehensive network and pathways for patients to find clinical trial sites through referrals. When you’ve identified key opinion leaders in a specific geography, you can plan a referral strategy that maximizes patient enrollment. If you create a site map that identifies where the influencers are and then open sites near clusters of KOLs, you can use these influencers to work with the community to refer patients. You can also identify referral sites based on established relationships across trial sites and non-trial sites, and adjust for established competitive relationships.

+ **Enrollment forecasting:** Patient enrollment has been one of the most difficult and challenging steps in clinical trials for several years. There are a host of operational factors that negatively impact enrollment. However, the most significant problem is forecasting that differs considerably from reality. For example, the issue is not that a trial might take 30 months to enroll the required number of patients, but that it was wrongly forecasted to take only 20 months. Having realistic expectations for how quickly a trial will complete enrollment has many downstream benefits.
Enrollment forecasting can be broken down into four major steps:

1. **Assess the availability of eligible patients.** The first step of the process is to estimate the number of eligible patients based on inclusion/exclusion criteria, treatment guidelines, procedures and the competitive landscape. Widely used datasets to determine the proportion of patients that meet study criteria include integrated EHR or claims data, clinical trial registries, publications and other subscription databases such as ZS’s Feasibility Database. It’s important to geographically distribute the availability of eligible patients on a heat map.

2. **Map productive sites and investigator-to-patient ratios.** Once the patient map is built, the next step is to identify productive sites and investigators. It’s paramount to model the performance of each site and investigator based on historical data and map it back to patient distribution to build the proximity of sites, investigators and patients. Leverage an analytics platform that provides an integrated and comprehensive global view with ample data sources.

3. **Simulate enrollment rates.** There are various models used to simulate enrollment rates, but the most widely used are the Poisson-Gamma enrollment model and the Monte Carlo Simulation. These models are highly flexible and provide an opportunity to model enrollment at different geographic levels. They work by calculating the probability that a certain outcome will occur based on a set of inputs: for example, predicting the probability that a certain trial will complete enrollment in nine months. Being able to predict enrollment with this type of precision can enable more efficient spending, better use of internal resources and more accurate financial projections for commercialization. (You can also use these models to make predictions using credibility intervals to determine the probability of on-time completion and evaluate the effect of trial marketing or changes to the number of centers.)

4. **Integrate the effect of direct-to-patient campaigns.** While companies are trying to estimate and identify eligible patients, sites and investigators for a trial, it’s important to analyze and plan for recruitment campaigns as well. There are several data sets (from within your company and external) that can be used to estimate the cost and effort involved in attracting trial patients to websites or call centers. Companies can further determine how many patients would pass through different levels of recruitment, such as initial screening, on-site screening, consent signing and final randomization. Companies also can estimate how many patients could potentially be referred. This data can then be fed into the simulation model to optimize the enrollment forecasts.
As our society continues its digital transformation, and the availability of data sources increases, it’s essential that pharmaceutical companies embrace this shift and leverage both their own data and external data sources to drive decision-making and get clinical trial operations right the first time.

Thoughtful, data-driven, strategic approaches to the problems of clinical trial feasibility are an essential element of success in the world of drug development. And aside from the obvious bottom-line benefits, there is an undeniable human factor to consider. More efficient and successful clinical trials get treatments to patients faster, and provide a better experience for enrollees. Given that an upfront focus on trial operation design will save companies money in the long run as well, we see nothing but upside to making an investment in data science.
About the Authors

Venkat Sethuraman is the global clinical lead within ZS’s R&D excellence practice. He has nearly 20 years of experience in R&D drug development life cycle with deep expertise in biostatistics, clinical trial design strategy, clinical trial optimization, and regulatory approvals. Venkat has a Ph.D. from Temple University and an MBA from Wharton Business School.

Aaron Mitchell is a managing principal and the global lead for ZS’s R&D excellence practice. Aaron partners with clients to strengthen the value of their pipelines, optimize the execution of clinical trials, drive effective use of real-world evidence, and improve medical and scientific engagement.

Yogesh Sharma is an associate principal in ZS’s New Delhi office and is a leader in the firm’s advanced data science practice. Yogesh has more than 20 years of experience in analytics and data science across many industries, from life sciences to financial services, retail, utilities, airlines, casino and hospitality, telecom, and technology. Yogesh specializes in bringing clients new ideas on how to leverage big data, AI, machine learning and predictive analytics to make better business decisions across functions.

Sharma R D is a manager within ZS’s global R&D excellence practice. Sharma has more than 10 years of experience in life sciences, with a special focus in clinical trial design and operations, pre-clinical toxicology, and patient enrollment and engagement. Sharma holds Global MBA from SP Jain School of Global Management, an executive master’s in information management from Stanford University and a bachelor’s in electronics and communication engineering.

Jessica Rine is a consultant within ZS’s R&D excellence practice. She has nearly 15 years of experience in R&D, managing clinical research programs across multiple therapeutic areas. Jessica has a bachelor’s in business administration from Centenary University.
About ZS

ZS is a global management consulting firm focused exclusively on helping companies improve overall performance—from customer insights and strategy, to analytics, to operations and technology. More than 6,000 ZS professionals in 22 offices worldwide draw on deep industry and domain expertise to deliver impact for clients across multiple industries. To learn more, visit www.zs.com or follow us on Twitter and LinkedIn.