

# Why Data Science is Key to Improving Medical Research Processes

---

The need for digital transformation has led to widespread adoption of big data and data analytics across businesses, including the pharmaceutical industry. Increasingly, drug companies are leveraging data science insights to drive innovations in research and development, especially when it comes to improving existing research processes.

Data science is unlocking more efficient ways of determining which compounds and molecules to use for drug development. With Artificial Intelligence (AI) and Machine Learning (ML) based tools, data scientists can use predictive modeling to accelerate drug discovery, optimize clinical trial design, and target specific patient populations for new drugs. Here's why data science is key to improving medical research processes:

## Data Science Accelerates Drug Discovery

Only about [12 percent](#) of drugs entering clinical trials ultimately gain FDA approval. Recent studies also found the average R&D cost of a new drug range from less than \$1 billion to more than \$2 billion.

During the process of drug discovery, pharmaceutical companies spend millions of dollars screening compounds to test in preclinical trials. Until recently, drug discovery was also a time-consuming process that could take a decade or longer. However, using data science and machine learning algorithms can drastically shorten the amount of time needed through predictive modeling.

Predictive modeling can be useful during all parts of drug discovery, from the initial screening of drug compounds to forecasting the drug's success rate based on patients' biological factors. Predictive algorithms can sift through past clinical trial data, like drugs with similar molecules and compounds or drugs created to treat the same illness. This allows researchers to identify and isolate specific molecules for testing at a quicker rate.

Using data science, researchers can generate new insights based on enormous datasets of drugs and assets that have undergone pre-clinical testing. These insights allow organizations to better prioritize which experiments to run and better understand the potential impact of a new drug. More insight and perspective during the drug discovery process helps avoid negative outcomes and potential disasters from patient risk factors.

## Data Science Enables Better Clinical Trial Planning

Drug companies must file an Investigation New Drug (IND) application with the FDA before any clinical trial can begin. This application includes key information, like the candidate drug's molecular structure, the results of preclinical work, how the experimental drug is predicted to work in the body, and a listing

of any potential side effects. An IND also needs to provide a clinical trial plan that describes how, where and by whom the studies will be conducted.

The IND application process can be complex, especially when data is collected at multiple locations involving many investigators. However, data science initiatives can help organizations breakdown data silos and leverage that data to design more optimized clinical trials. For example, data scientists can use an AI software solution to analyze patient profiles along with their medical histories to pick patients who will respond best to the drug being tested. The result is valuable time saved when trying to find the suitable patients for clinical trials.

Through predictive analytics, pharmaceutical companies can use their past trials or a database of trial data from other companies to find and establish best practices for upcoming clinical trials. Having information about procedures, clinical trial operations, and the trial's relative success rate will help researchers plan trials with ideal conditions and avoid past mistakes.

---

*Modern data science... will accelerate the creation of smarter studies, with fewer protocol amendments and greater confidence that trials will more efficiently understand the efficacy and safety of new medicines.*

*Craig Lipset, former head of clinical innovation at Pfizer*

---

## Data Science Targets Specific Patient Populations

With any disease or illness, different patients will respond differently to treatments for various reasons. But regardless of the condition, there is massive amounts of patient data being collected. Electronic medical records, medical sensor data, and genomic sequencing data are all becoming widely available. Combing through all the data from these different sources is a task greatly simplified by modern data analysis software.

Trained data scientists can use the latest data analytic tools to spot trends and patterns that enable drug companies to create more targeted medications for patients with common features. Large pharmaceutical corporations, such as Pfizer, are combining data from a patient's sequenced genome, clinical trials, and electronic medical records to find more treatment opportunities for specific patient populations.

Taking an analytical approach to testing compounds allow companies to identify a subset of patients with a specific gene mutation who are lacking treatment options. Cancer patients with specific gene mutations, for example, are one subset of patients who have benefited from this targeted approach to drug development and research.

---

*Starting with protocol design, we can better utilize observational, epidemiological and safety data for the identification of the right patient population, inclusion and exclusion criteria, and sample size.*

*Victor Lobanov, VP, informatics solution development at Covance*

---

## Leveraging Data Science Requires Tools and Training

Microsoft Excel is a tool that medical researchers are familiar with. Many have been using it for years to track data and prepare data for analysis. However, while Excel can handle simple data manipulation, it is focused on presentation and ease of use rather than providing deep analysis. With medical research using increasingly complex statistical models, Excel is proving impractical for deep data science.

Data science mostly involve regression, classification, and complex statistical models that Excel is not optimized for. As a result, modern pharmaceutical companies are turning to open-source languages like R and Python for these functionalities. However, modern data science tools require skilled data scientists to use to generate actionable insights.

Data science training can be one of the best decisions a drug company can make for its long-term success. By investing in data science initiatives, you can shorten the biopharmaceutical research and development process, which typically requires billions of dollars in investment and many years of testing. From drug discovery to preclinical tests to clinical trials, data science can speed up researching and developing new medicines.

## Conclusion

From a quicker drug discovery process to better optimized clinical trials, data science is driving industry-wide innovations when it comes to medical research. Thanks to accurate predictive modeling and machine learning algorithms, researchers can efficiently develop new drugs under time pressure and identify new uses for existing drugs. However, you need trained data scientists that understand algorithms, data collection, storage, validation, and visualization to kick off the process.

Data science and data analytics is redefining how drugs are made and tested. However, drug companies need trained data scientists to parse through the data and turn it into actionable insights. With effective training, your organization can upskill existing employees into a skilled data science team capable of using programming languages like R and Python. Data science training can prepare your organization to harness Big Data, creating substantial competitive advantages during the research process and beyond.