

Clinical trials are the backbone of drug development and diagnostic methods. Clinical trials are essential because they inform researchers about diseases and treatments in a way that cannot be learned in the laboratory.

Clinical trials are also resource intensive - they are expensive, time consuming and require trained workers. They are often burdensome on patients. They have multiple phases which can last months. The investments of resources, time, and funding grow with successive stages, from pre-clinical through phase 3. Thus, the cost of a failed phase 3 trial is not just the cost associated with the trial itself but the cost of all prior trials as well as the opportunity cost of lost time pursuing a potentially viable alternative. Focusing on phase 3 trials, the Pharmaceutical Research and Manufacturers of America estimated the cost at **\$42,000** per patient in 2013, with **\$10 billion** spent on 1680 phase 3 clinical trials comprising over **600,000** patients.

The purpose of our project was to analyze clinical trial data and determine the most important features contributing to clinical trial phase failure. After modeling the data from NIH U.S National Library of Medicine's website - <https://clinicaltrials.gov/>, we concluded that the enrollment fraction i.e. the actual enrollment divided by the desired enrollment is the major predictor for successful completion of clinical trials. Our logistic regression model, if implemented, would save **24%** of the money spent.

It is not the length of the trial, nor how invasive the trial is, or which age group it targets which make a difference. So clinical trials should not be given undue priority on the basis of these factors. Instead, there should be an increased effort to increase enrollment. This could be via increased marketing, more monetary and non-monetary incentives for the participants, decreased wait times, etc to name a few factors.

The National Institutes of Health (NIH) invests about **\$41.7 billion annually** in medical research. More than 80 percent of NIH's funding is awarded for extramural research, largely through almost 50,000 competitive grants to more than 300,000 researchers at more than 2,500 universities, medical schools, and other research institutions in every state. Using our logistic prediction model, NIH could save **24%** of the **30 billion** it spends on clinical trials i.e approximately **7.2 billion** dollars of taxpayer money.

We suggest that data related to factors which could influence the enrollment factor of a clinical trial be recorded so as to facilitate a deeper analysis into how to increase enrollment and predict successful completion. This could save the NIH and the pharmaceutical companies billions of dollars.