Machine Learning and AI in Clinical Trials Recruitment – Potentially Cost-saving

Ebenezer Adjei Armah, BS Computer Science and Mathematics

MSE Candidate, Center for Bioengineering Innovation and Design, Biomedical Engineering Department, Johns Hopkins University

Keywords: Machine Learning; Artificial Intelligence; Clinical Trials; Recruitment for Clinical Trials; Electronic health records; Computational Phenotypes; Natural Language Processing

Abbreviations: Electronic Health Records (EHR); Machine Learning (ML); Artificial Intelligence (AI); Natural Language Processing (NLP)

Introduction

The US healthcare system is burdened financially. The overall cost of healthcare in the US was $3.7 trillion in the year 2017 [1], a 4.6% increase from the previous year, and $4.4 trillion in 2018 [2]. With this alarming rate of spending, sustainable measures must be taken to ensure that expenditure is kept under control.

One of the contributing factors to this problem is the cost of pharmaceutical drugs [3]. Drugs can sometimes be very pricey and usually pharmaceutical companies get bad media coverage for raising the price of life-saving medicines. The case of the price hike of the EpiPen in 2016 [4] is an example of the vilification of pharmaceutical companies who charge high prices for certain medicines. However pharmaceutical companies are not the only ones to blame for this situation. For example, it takes between 10 - 15 years and about $1 billion for a company to bring a drug to market [5]. As such, the usual strategy for many companies to make up for the Research and Development (R&D) costs and eventually profits, is to assign a high price to their drugs, especially while it is protected under patent. If pharmaceutical companies could reduce the costs and time associated with bringing a drug to the market, there would be a substantial decrease in the price charged for a drug which has the potential downstream effect of reducing the overall costs to the US healthcare system.

The US healthcare system comprises patients, providers (physicians, physician offices, health systems), payers (insurance companies) and regulatory bodies. Regulators ensure that healthcare products in the market have a reasonable level of safety and efficacy. As such, there are certain standards that must be met by all manufacturers in order for their products to make it to the market. This ensures that consumers of healthcare products are safe. However, these standards come with many unintended consequences.

Firstly, it drives up the cost of product development for manufacturers in this domain. The increased cost affects the prices of these products on the market leading to an overall increase in expenditure. Secondly, it slows down the time to market for potential life-saving products. Also, the longer a drug takes to get to the market, the higher the opportunity costs for the producers and the poorer the health outcomes for patients who would benefit immensely from this drug.

Organizations Involved and Drivers of Cost

Organizations involved are the Food and Drugs Administration (FDA) and the Pharmaceutical Research and Manufacturers of America (PhRMA). The FDA formed with the passage of the Pure Foods and Drugs Act in 1906 is responsible for ensuring the safety and efficacy of drugs and devices in the US. The Center for Drug Evaluation and Research (CDER) is the specific department of the FDA that evaluates clinical trial data and ensures only qualified drugs make it to the market. PhRMA represents biopharmaceutical and biotechnology companies involved in research and production of drugs. PhRMA companies have to go through a regulatory process in order to get their drugs cleared or approved.

The main challenge that these organizations face is to find the most cost-effective way of doing clinical trials that would lead to maximum assurance of safety and efficacy. The key driver is that policy makers are seeking to reduce the overall costs of healthcare in the US and the price of drugs play a very important role in this cost [6]. The amount of money spent during clinical trials are factored into the prices of drugs and there could be drastic decrease in price if there was a way to decrease cost of clinical trials.

This article takes the perspective of a pharmaceutical company whose goal is to reduce the amount of money spent during clinical trials. For every drug that makes it to the market, PhRMA (aggregated) spends about $2.7 billion [7]. This is because about 90% of drugs fail in the clinical trial phase and do not make it to the market. What accounts for this figure includes the cost not only of these failures, but also of not putting the money spent on them into something that would give a more reliable return [7].

Also, about 29% of clinical trial costs could be attributed to the clinical trial enrollment process [8]. Roughly 80% of clinical trials fail to meet enrollment timelines, and approximately one-third of Phase III clinical study terminations are due to enrollment difficulties, according to a Cognizant report on recruitment forecasts [9]. Difficulties in recruitment causes delays in clinical trials which in turn lead to overbudgeting. If there could be a way to optimize the clinical trial enrollment process, the right patients could be determined in a short amount of time and could drastically reduce costs and certainly reduce the chance of a potential life-saving drug failing during the clinical trial stage.

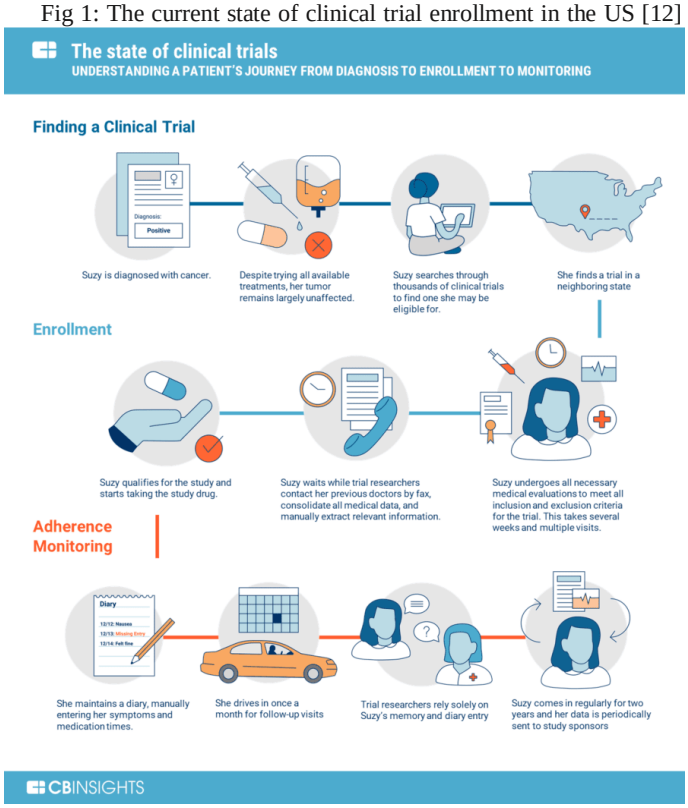
Current Recruitment Workflow

The current workflow relies heavily on patients finding out whether a clinical trial is ongoing or not. Healthcare providers (HCPs) are the point of contact for enrollment — they are familiar with their patients' demographics and ideal for helping patients get accepted. Usually, drug companies have to make the physical effort of door-knocking through physician office visits in order to enroll patients. Patients who are not in close proximity to sites where these trials are being conducted do not get to participate in this process. It has been estimated that less than 5% of adult cancer patients are enrolled in clinical trials [10]. In fact, according to a White House report in 2018, only 3% of cancer patients in America were involved in any form of clinical trial [11]. This underscores the fact the current task of recruiting new patients for a clinical trial is handled inefficiently.

Sometimes patients who are technologically savvy try to work hard to find clinical trials that they may be eligible for. However, the arduous task of finding clinical trials drains some patients and their families.

Fig 1 shows the current typical pathway for a determined patient to enroll in a clinical trial

Fig 1: The current state of clinical trial enrollment in the US [12]



Current Information Systems in Clinical Trial Recruitment

The most popular information system to help patients know what kind of trials are ongoing is the clinicaltrials.gov database [13]. This resource is up to date with all the valid clinical trials ongoing in the US and globally. When the patient visits the homepage of the website, they have the option to enter keywords that would be mapped to concepts that would direct them to the right clinical trial. The user can also search by country and find which trials are close to them. This system is not only designed for patients in search of clinical trials but also researchers who are seeking to enroll patients in a clinical trial. There is an advanced search which gives a user the opportunity to filter the number of results using demographics and the status of the trial. This system however does not communicate with any other system nor does it take any signs and symptoms of a patient in order to guess the disease.

The data in the current system is text data that is written to a database or read from a database. It has pertinent information about a clinical trial including but not limited to the location, timeline, status, eligibility criteria, exclusion criteria, funders, and statistical plans. With this information, a healthcare provider or a patient decides on their own as to whether they or their patients qualify for a trial.

For clinicaltrials.gov, the only decision support is the results from query using keywords and filters. There is no other decision support. The algorithm used is just a database query and is not very sophisticated. There is no encryption of data because no personal information is entered in the search bars and the system does not pull information from any electronic health record (EHR). The current information system is web-based and can be viewed from any device that supports a browser. As such, computers, phones and tablets could be used to access this resource

Information System Proposal

It can be seen that the current information system is very passive and creates room for many errors associated with recruitment. Patients or their healthcare providers have to use keywords that match the concepts used in the database. This system does not actively find patients who could be potential candidates for a trial. Both patients or researchers do not gain the maximum benefits and as such the current system leaves much room for improvement. This article proposes the use or application of Machine Learning in the recruitment process.

- **Overview**

Machine learning (ML) is an umbrella term that refers to a broad range of algorithms that perform intelligent predictions based on a data set. These data sets are often large, perhaps consisting of millions of unique data points. There are two main types of ML concepts namely supervised learning and unsupervised learning. Supervised learning involves teaching the model with a collection of input data that has the correct output called labels already associated with it. Unsupervised learning however involved the feeding of data to a model without labels and the model trains itself using features it learns on its own. Examples of supervised learning include predictions based on images and their names, features and a label. Unsupervised learning however involves clustering, and various forms of pattern recognition.

A machine learning model works by extracting important features from a set of data. For example, there could be a significant amount of data on patients' characteristics and specific outcomes. The patients' characteristics could be used as input and the outcome as output. As more and more of this data is fed to the model, the model over time "learns" important features from the characteristics that lead to specific outcomes. With time the model is able to predict outcomes with only characteristics fed to it.

An information system based on this model could be connected to an electronic health / medical record (EHR / EMR). This system could then mine the EMR for the right patients for a clinical trial. This will ensure that the system is not passive but active.

- **Data Information Knowledge Models**

Machine learning concepts applied to EHR could optimize the clinical trial enrollment process. There is enormous data yet to be tapped that exists in EHRs. EHRs have data about a patient's demographics, medical history, medications and social determinants of health. All these data points could be fed to a machine learning model and as the model learns over time, it will be able to predict patients who are eligible for certain trials. Most of the data in EHRs are text data. As such a Natural Language Processor (NLP) could be trained to make sense of the text data. NLP is a branch of artificial intelligence that deals with analyzing, understanding and generating the languages that humans use naturally in order to interface with computers in both written and spoken contexts using natural human languages instead of computer languages. An NLP can glean insights regarding context and could be used in written text.

The algorithms that could be used in the proposed system involve a series of steps targeted at creating computable phenotypes. A computable phenotype is the ability to first define a condition, disease, patient characteristic or clinical event using only data processed by computer and secondly, to use a model or algorithm to identify a population of patients with a condition of interest.

It should be noted that not all data in an EHR are in the form of text. There are images and sometimes sounds. Computer Vision algorithms could be applied to these images and the sounds could be processed using Recurrent Neural Networks. These algorithms combined could create a very robust computable phenotype therefore optimizing the clinical trial enrollment process.

This system after being integrated into EHRs could make suggestions to physicians and other users of the EHR that their patients qualify for certain clinical trials. This system could also be bidirectional and as such alert researchers who are seeking patients to enroll in their clinical trials of potential candidates.

With such a system fully deployed clinical trials delays and drug failures could be drastically reduced thereby affecting the pricing of drugs.

- **Strengths and Limitations**

The system proves superior to the current system in that it actively finds patients who would not be found using the current methods. Using the predictive capability of ML models, patients who are not in close proximity to the trial site could be spotted and could be enrolled accordingly. The time for enrollment could be drastically reduced as enrollment could happen at multiple sites simultaneously.

The proposed model comes with these limitations. Firstly, there is a possibility of a HIPAA violation which could cost health systems so much money. Also, EHRs in different health systems are usually not interoperable and could be a barrier to realizing the full potential of this system

Conclusion

AI and ML models have the potential to drastically US expenditure on healthcare by optimizing the clinical trial process. The advantages of a system based on these models outweigh the minor limitations that these systems pose.

References

1. US Healthcare Trends and Contradictions in 2019," accessed May 6, 2020, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6404804
2. The Healthcare Imperative - The Healthcare Imperative - NCBI Bookshelf," accessed May 2, 2020, https://www.ncbi.nlm.nih.gov/books/NBK53944/.
3. Are Hospital Prices a Bigger Problem than Drug Prices?" accessed May 2, 2020, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6487976/.
4. Why Did Mylan Hike EpiPen Prices 400%? Because They Could," accessed May 2,2020, https://www.forbes.com/sites/emilywillingham/2016/08/21/why-did-mylan-hike-epipen-prices-400-because-they-could/#5d1bd323280c.
5. The Antibiotics Business Is Broken—But There's a Fix | WIRED," accessed May 9, 2020, https://www.wired.com/story/the-antibiotics-business-is-broken-but-theres-a-fix/.
6. "Policy Proposals for Reducing Health Care Costs - National Health Council," accessed May 2, 2020, https://nationalhealthcouncil.org/policy-proposals-for-reducing-health-care-costs/.
7. "The Cost Of Developing Drugs Is Insane. That Paper That Says Otherwise Is Insanely Bad, "accessed May 2, 2020, https://www.forbes.com/sites/matthewherper/2017/10/16/the-cost-of-developing-drugs-is-insane-a-paper-that-argued-otherwise-was-insanely-bad/#3203ee9a2d45
8. New Tufts Research Explores Barriers to Clinical Trial Recruitment Success," accessed May 2, 2020, https://www.patientcentra.com/patient-recruitment-insights/tufts-research-barriers-clinical-trial-recruitment-success.
9. Improving Clinical Trial Enrollment Forecasts Using SORM," accessed May 2, 2020, http://connection.ebscohost.com/c/articles/87731977/improving-clinical-trial-enrollment-forecasts-using-sorm
10. Role of Clinical Trial Participation in Cancer Research: Barriers, Evidence, and Strategies," accessed May 2, 2020, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5495113/.
11. President Donald J. Trump to Sign Right to Try Legislation Fulfilling the Promise He Made to Expand Healthcare Options for Terminal Americans | The White House," accessed May 2, 2020, https://www.whitehouse.gov/briefings-statements/president-donald-j-trump-sign-right-try-legislation-fulfilling-promise-made-expand-healthcare-options-terminal-americans/
12. The Future Of Clinical Trials: The Promise Of AI And The Role Of Big Tech," accessed May 2, 2020, https://www.cbinsights.com/research/clinical-trials-ai-tech-disruption/.
13. Home - ClinicalTrials.Gov," accessed May 2, 2020, https://clinicaltrials.gov/.