# Project Proposal: A Deployed Dashboard Using Clinical Trial Data

Team: Mission Possible

Xiao Wu, Can Wang, Xindi Shan

2024-11-13

## Table of contents



# 1 Introduction

Clinical trials are important to the development of medical knowledge and therapies. Understanding the trends and characteristics of clinical trials can help make decisions in research, policy, and healthcare investments. Our project, "A Deployed Dashboard Using Clinical Trial Data", aims to explore the global distribution of clinical trials by analyzing patterns in trial registration data and predicting future trends. We will focus on identifying the types of clinical trials being conducted

across different countries, examining their key characteristics (e.g., disease type, study phase, and recruitment status), and predicting how these trends might evolve in the next few years.

## 2 Research Question and Importance

The main research question of our project is: What are the trends in the number and types of clinical trials conducted across different countries, what are their characteristics, and how do they change over time? This question is important because understanding global clinical trial activity can provide information to the allocation of resources, policy decisions, and future research priorities. By analyzing these trends, we can gain insights into the geographical distribution of research efforts. Also, predicting future clinical trial activity will provide foresight for researchers, drug developers, and healthcare providers.

## 3 Existing Work

Previous research and dashboards on clinical trials have focused on various aspects of trial data, such as geographical distribution, disease-specific trials, and trial phases. For example, the World Health Organization's International Clinical Trials Registry Platform (ICTRP) (World Health Organization 2024) aggregates data from multiple sources, providing a global view of clinical trials.

However, many existing tools don't offer an interactive, predictive interface that synthesizes historical data and forecasts future trends. Our project aims to fill this gap by providing a user-friendly, interactive dashboard that not only visualizes historical trends but also offers predictions for future clinical trial activity.

## 4 Method

In this project, we aim to develop an interactive dashboard that will summarize, visualize, and predict trends in registered clinical trials. Specifically, we hope to track trends in the number of trials across different countries, disease types, study phases, and other key characteristics over time, while also predicting future trends.

## 4.1 Data Collection and Pre-processing

The primary data source for this project will be ClinicalTrials.gov (National Library of Medicine 2024), an online database of clinical studies conducted worldwide. ClinicalTrials.gov provides detailed metadata on clinical trials, including information on disease type, geographic location, intervention details, study phase, recruitment status, and sample size. This data is accessible via the ClinicalTrials.gov API, through which we will submit requests to retrieve the data needed for dashboard construction and trend analysis.

Data Extraction through a systematic pipeline:

- Key variables to be collected include: NCT identifier, trial phase, conditions studied, enrollment count, start/completion dates, locations, intervention types, and recruitment status

Data extracted from the API will be stored locally, ensuring convenient access for analysis and future project updates. The data will then be cleaned using R packages such as `tidyverse`, `dplyr`, and `tidyr`. To visualize trends in clinical trial data, we will use `ggplot2` to create a variety of plots, including scatter plots, bar charts, and line graphs, that highlight patterns over time. A predictive model will be built to predict the number of clinical trials by country, disease type, and phase over the next three years. The interactive dashboard will be built using R Shiny with the `shiny` and `shinydashboard` packages, and will be published using Shinyapps.io.

## 4.2 Prediction Modeling

To predict future trends in clinical trials, we will use both time series analysis and machine learning approaches. Our analysis will be implemented in R.

Time Series Analysis:

Given that clinical trial data changes over time, we will use:

- Prophet models (using `prophet` package) for trend analysis and forecasting

Machine Laerning Approaches:

We might use several machine learning models (not all, conditional on our cleaned data):

1. Linear Methods:
    - Multiple Linear Regression (baseline model)
    - LASSO Regression (for selecting important features)
2. More Advanced Methods:
    - Random Forest (using `randomForest` package)
    - Support Vector Machines (using `e1071` package)

Based on the cleaned data, we will first split the dataset into training (80%) and testing (20%) sets. The models will be trained using cross-validation for parameter tuning, and their performance will be evaluated using error metrics (RMSE, MAE) on the testing data. The best performing model will then be used to generate 3-year forecasts for the number of trials by country, trial distribution by disease type, and trial phase patterns.

## 4.3 Programming Implementation and Collaboration

Throughout this project, programming in the command line and machine learning paradigms will be extensively used. For better collaboration within the team, a shared GitHub repository will be created, where command-line programming will be used for file management and version control (`Git`). Machine learning paradigms are also essential to this project, as they will enable us to generate reliable predictions and thus add a forecasting section to our dashboard.

## 4.4 Final Product: Interactive Dashboard

The final dashboard will provide users with a clear view of clinical trial data over the past 20 years and forecasts for the upcoming three years. Users will have the option to explore the data by country, disease type, and study phase. This tool will serve as a comprehensive resource for stakeholders seeking to understand the landscape of clinical trial activity globally, making it valuable to researchers, sponsors, and policy makers.

# 5 Tasks and Timeline

| Tasks | Assigned to | Timeline |
|---|---|---|
| Data collection and pre-processing | Xiao Wu | November 20 - November 24 |
| Prediction modeling | Xindi Shan | November 28 - December 2 |
| Dashboard development | Can Wang | December 3 - December 7 |
| Presentation | All members will contribute. | December 8 - December 11 |
| Final project write-up | All members will contribute. | December 12 - December 19 |

# References

National Library of Medicine. 2024. “ClinicalTrials.gov.” https://clinicaltrials.gov.

World Health Organization. 2024. “WHO ICTRP Search Portal.” https://trialsearch.who.int/.