COMPLETED AND TERMINATED CLINICAL TRIAL STUDIES

### PREPARED BY

Colleta Nandutu

Janet Jepkorir Cheserem

TABLE OF CONTENTS

*Title Page* [1](#_heading=h.30j0zll)

[INTRODUCTION](#_heading=h.2et92p0) [3](#_heading=h.2et92p0)

[METRIC OF SUCCESS](#_heading=h.tyjcwt) [3](#_heading=h.tyjcwt)

[RESEARCH DESIGN](#_heading=h.1t3h5sf) [4](#_heading=h.1t3h5sf)

[Business Understanding](#_heading=h.4d34og8) [4](#_heading=h.4d34og8)

[Project Objectives](#_heading=h.17dp8vu) [4](#_heading=h.17dp8vu)

[Data Mining Goals](#_heading=h.3rdcrjn) [4](#_heading=h.3rdcrjn)

[Assessing the Situation](#_heading=h.lnxbz9) [4](#_heading=h.lnxbz9)

[Data Understanding](#_heading=h.35nkun2) [5](#_heading=h.35nkun2)

[Data Collection](#_heading=h.1ksv4uv) [5](#_heading=h.1ksv4uv)

[Data Description](#_heading=h.44sinio) [5](#_heading=h.44sinio)

[Data Exploration](#_heading=h.z337ya) [6](#_heading=h.z337ya)

[Data Quality](#_heading=h.3j2qqm3) [6](#_heading=h.3j2qqm3)

[Data Preparation](#_heading=h.1y810tw) [6](#_heading=h.1y810tw)

[Data Cleaning](#_heading=h.4i7ojhp) [6](#_heading=h.4i7ojhp)

[Exploratory Analysis](#_heading=h.2xcytpi) [7](#_heading=h.2xcytpi)

[Univariate Analysis](#_heading=h.3whwml4) [7](#_heading=h.3whwml4)

[Bivariate Analysis](#_heading=h.2bn6wsx) [9](#_heading=h.2bn6wsx)

[Evaluation](#_heading=h.3as4poj) 12

A PREDICTIVE MODEL TO DETERMINE THE COMPLETION OR TERMINATION OF CLINICAL TRIAL STUDIES

# INTRODUCTION

Clinical trial studies play a pivotal role in advancing medical research and healthcare practices. These studies are structured investigations conducted on human participants to assess the safety, efficacy, and outcomes of new medical interventions, such as drugs, treatments, or procedures. The goal is to gain valuable insights into the prevention, diagnosis, or treatment of diseases and conditions.

Our dataset comprises 30 essential attributes, each offering a unique perspective on various aspects of these trials. From the distinct NCT numbers assigned to each study to the comprehensive study titles elaborating on research objectives, the dataset captures crucial details. It delves into the study status, indicating the current phase and progress of the trial, and provides concise summaries outlining the trial's purpose and methods.

Additionally, the dataset includes information about the conditions targeted by the trial, the interventions being tested, and the primary and secondary outcome measures used to evaluate the study's success. Furthermore, it outlines vital details such as the sponsors and collaborators involved, the demographic characteristics of participants (including age and gender), and the trial's start and completion dates.

Beyond these, the dataset also encompasses various identifiers, study locations, and documents associated with the trial, offering a comprehensive overview. Each attribute holds significance, contributing to a holistic understanding of the clinical trials documented within the dataset. This rich and diverse dataset forms the foundation for in-depth analysis, enabling researchers and stakeholders to draw meaningful conclusions and facilitate advancements in the field of healthcare and medicine.

The aim of this project is to come up with a model that can predict which clinical trial studies are most likely to be completed or terminated.

# METRIC OF SUCCESS

This research will be considered successful when we are able to predict the clinical trial studies that have the high chances of being completed or terminated along the way.

Being able to envision which part of the

# 

# RESEARCH DESIGN

In this project, we use the Cross Industry Standard Process for Data Mining as our research design to comprehensively dissect the problem of identifying the trial studies that will be terminated or completed, and will constitute the following steps:

* Business Understanding
* Data Understanding
* Data Preparation
* Analysis
* Evaluation

# Business Understanding

Clinical trial studies are systematic investigations conducted to assess the safety, efficacy, and outcomes of medical interventions on human subjects. These rigorous experiments are essential steps in the development of new drugs, treatments, and therapies. They play a pivotal role in advancing medical knowledge, guiding healthcare practices, and improving patient outcomes. Clinical trials follow strict protocols and are often conducted in multiple phases, each designed to answer specific research questions. Participants in these trials are carefully monitored, and the data collected undergoes rigorous analysis. The results obtained from clinical trials inform healthcare professionals, regulatory agencies, and patients about the benefits, risks, and effectiveness of medical interventions. Ultimately, these studies contribute significantly to medical innovation, enabling the development of life-saving treatments and shaping the future of healthcare.

## Project Objectives

Our research objectives include:

* To establish the type of clinical studies conducted
* To obtain the number of clinical trial studies in each clinical trial phase
* To find the most studied diseases
* To find out the top sponsors of the clinical studies
* To find out the statuses of the registered clinical studies
* To establish the most commonly used method of clinical studies

## Data Mining Goals

In order to sync our business goals with data science,we list the following goals:

* Exploratory data analysis : univariate, bivariate and analysis on the data variables.
* Supervised Machine Learning

## 

## Assessing the Situation

With clearly outlined goals,we make an assessment of the situation as well as putting together the resources that will see the completion of the project.This are personnel and data.

* Dataset
* Programming tools ; Google colab , python language and several other libraries such as matplotlib, seaborn, matplotlib, pandas and numpy.
* Personnel ; 2 Data scientists

# Data Understanding

Before working on the data,it is important to understand the contents of the dataset provided and to establish the relevance of the information contained in the dataset to our research.This therefore involved the exploration of data.

## Data Collection

The dataset was provided for this research.

## Data Description

The data contains 161,863 rows and 30 columns.The table below shows a summary of the columns:

| COLUMNS | DESCRIPTION |
| --- | --- |
| NCT Number | The unique identifier assigned to a clinical trial |
| Study Title | The clinical trial formal title |
| Study URL | The web address to the clinical trial studies |
| Acronym | Abbreviation from the initial letters of each word in study title |
| Study Status | Current status of the clinical trial; recruiting, completed etc |
| Brief Summary | Description of the trial's purpose, methods, and goals |
| Results First Posted | First time the results were made public |
| Conditions | Diseases |
| Interventions | Details of treatments, drugs & procedures being tested |
| Primary Outcome Measures | The main objectives of the study |
| Secondary Outcome Measures | Additional measurements beyond the primary objective |
| Sponsors | The entity responsible for initiating, managing, and financing the trials |
| Collaborators | Other institutions involved in the clinical trial alongside the sponsors |
| Sex | Gender of the participants |
| Age | Specifies the the age range/demographics of the participants |
| Phases | Indicates the stage of the clinical trial study |
| Enrollment | The number of the participants planned/enrolled in the studies |
| Funder Type | Type of funding source eg government, industry,academic, or NGO |
| Study Type | Specifies the type of clinical trial(Interventional or Observational) |
| Study Design | Overall plan of the clinical trial; the structure, duration, and methodology |
| Start Date | Official start date |
| Primary Completion Date | The anticipated date when the final participants’s data for the primary outcome measures are collected |
| Completion Date | The anticipated/actual date when the clinical trial concludes |
| Locations | Geographical locations where the trials are being conducted |
| Study Documents | Files related to the clinical trial |
| Other Outcome Measures | Any secondary measurements not specified as primary or secondary outcome measures |
| Other IDS | Additional identification numbers related to the trials |
| First Posted | First Date of publicly posting on the clinical trial registry |
| Last Update Posted | Most recent date when any updates to the clinical trial studies |

## 

## Data Exploration

Exploration is important in any research as it helps identify errors,checking for any unique values which helps in the preparation of the data for data cleaning.

## Data Quality

After previewing and exploring the data,data cleaning processes are required to make sure that the data being used is of quality and one that will give the most accurate results possible.

# Data Preparation

This is the process of making sure the data achieves and maintains quality for the final use in the analysis.

## Data Cleaning

In data cleaning,we go ahead to check for the uniformity of the data which was done by ensuring that all column names are in lower case.Missing values were also identified and they were dropped.We also identified any duplicated values and we established there was none.

# 

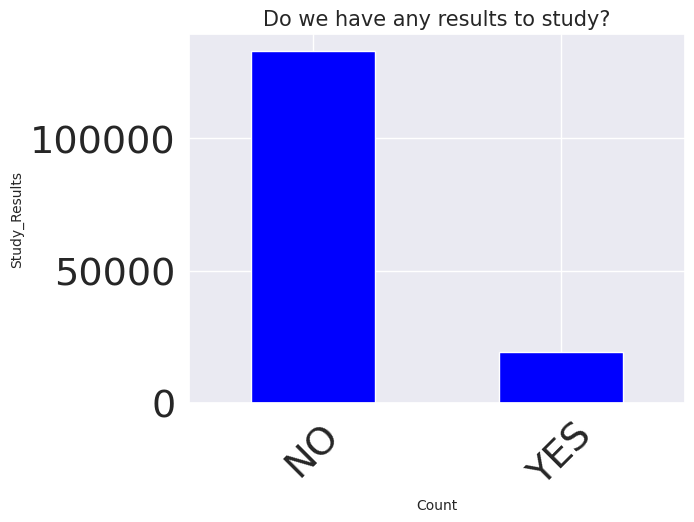
# Exploratory Analysis

Exploratory data analysis that was conducted include:univariate analysis and bivariate analysis.

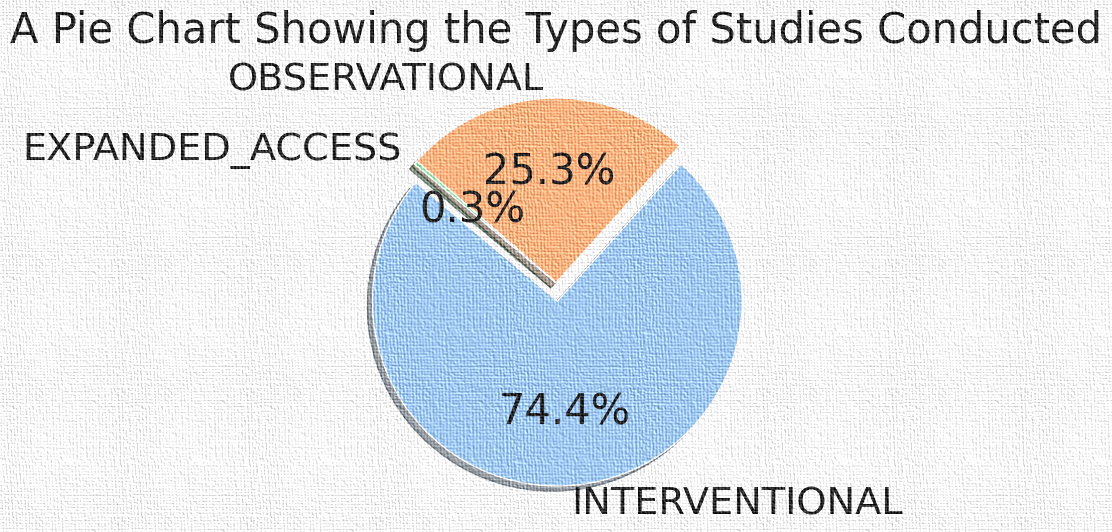
## Univariate Analysis

The following observations were made when the analysis was conducted:

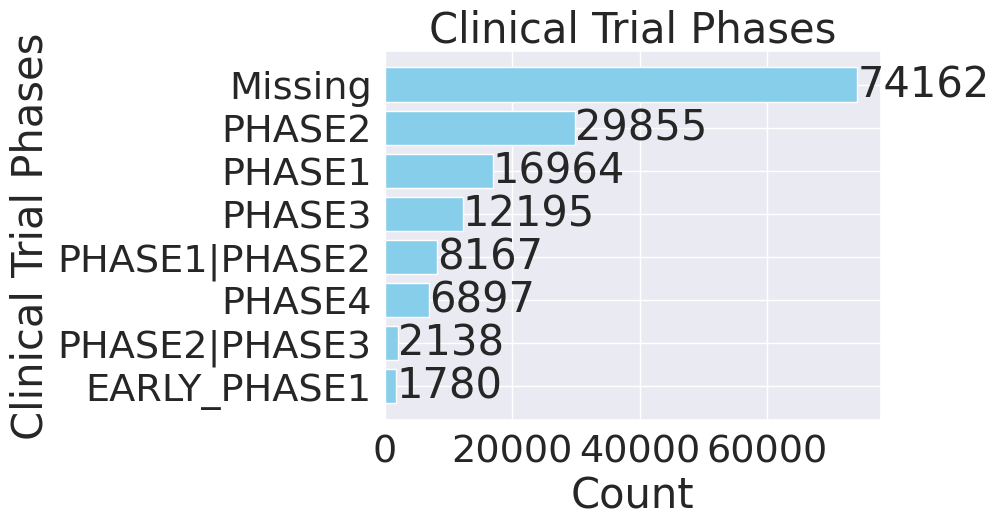
1. Only a small number of the clinical research studies(19,232) had the final study results while 132,926 did not have the final results.



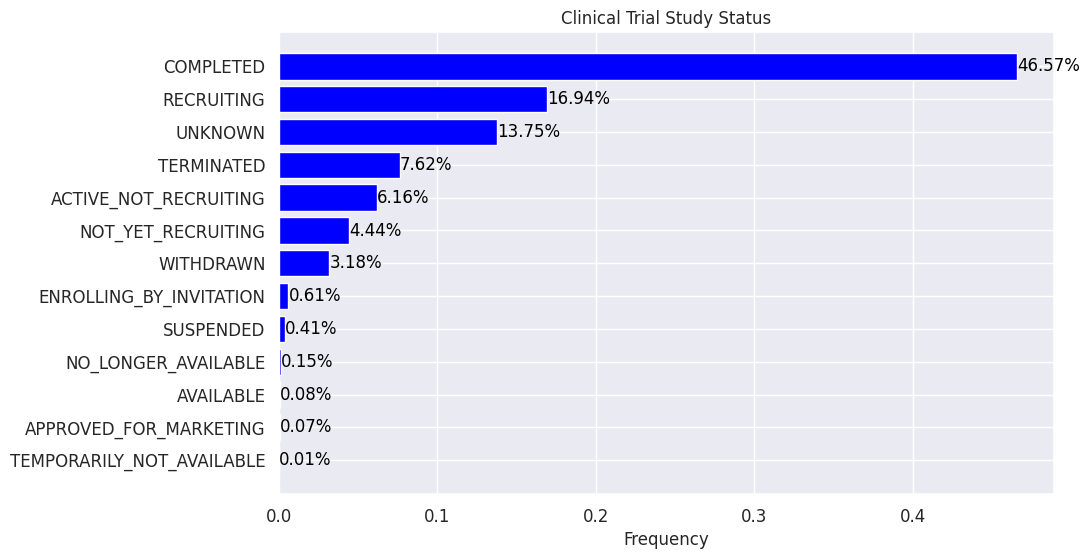
2. 74.4% of the studies were interventional, while 25.3% were observational, and only 0.3% classified as expanded access.



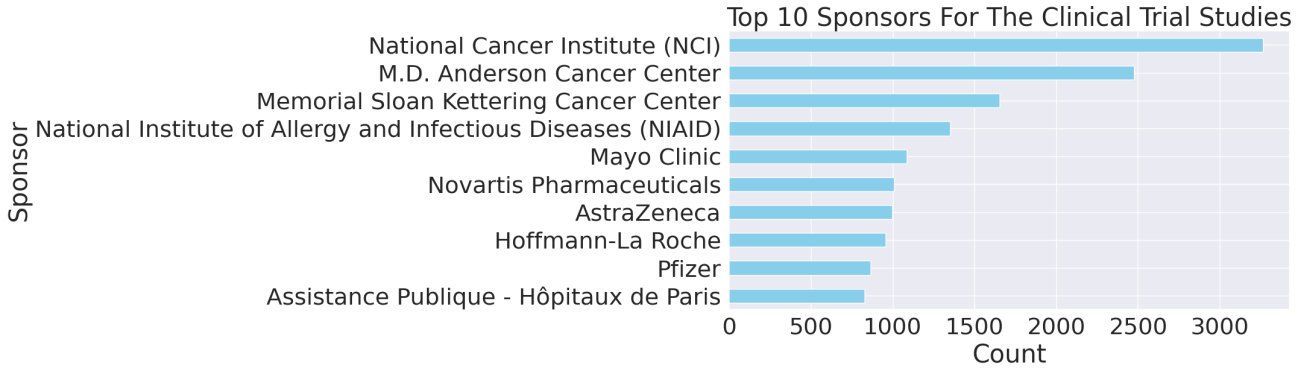
1. Most of the clinical research studies did not have the precise trial phases.



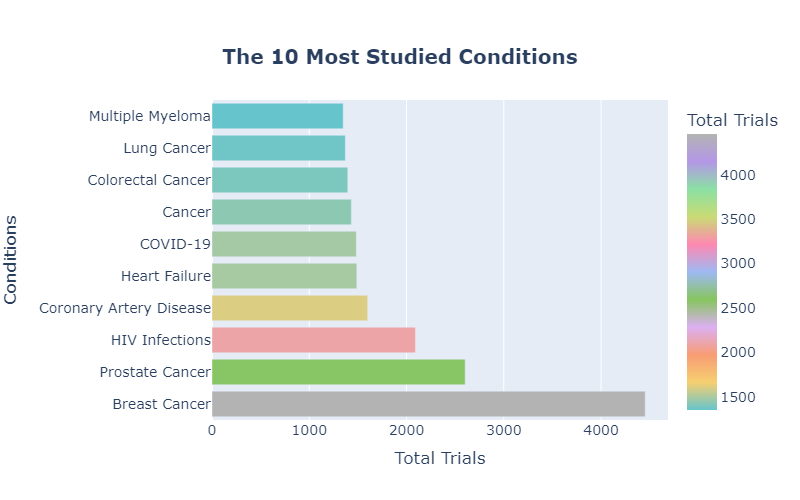
1. 47.57% of the studies were completed as per the protocols: 16.94% were still recruiting



1. The top sponsors were government agencies/hospitals from the United States of America, and pharmaceutical companies from Switzerland and UK/USA.



1. Six of the top studied conditions are cancers of the bones, lungs, breast,colorectal, and prostate.



1. Most of the participants are of the category “Adults” (18 to 64 years)and Older “Adults” (65 years and above).

## 

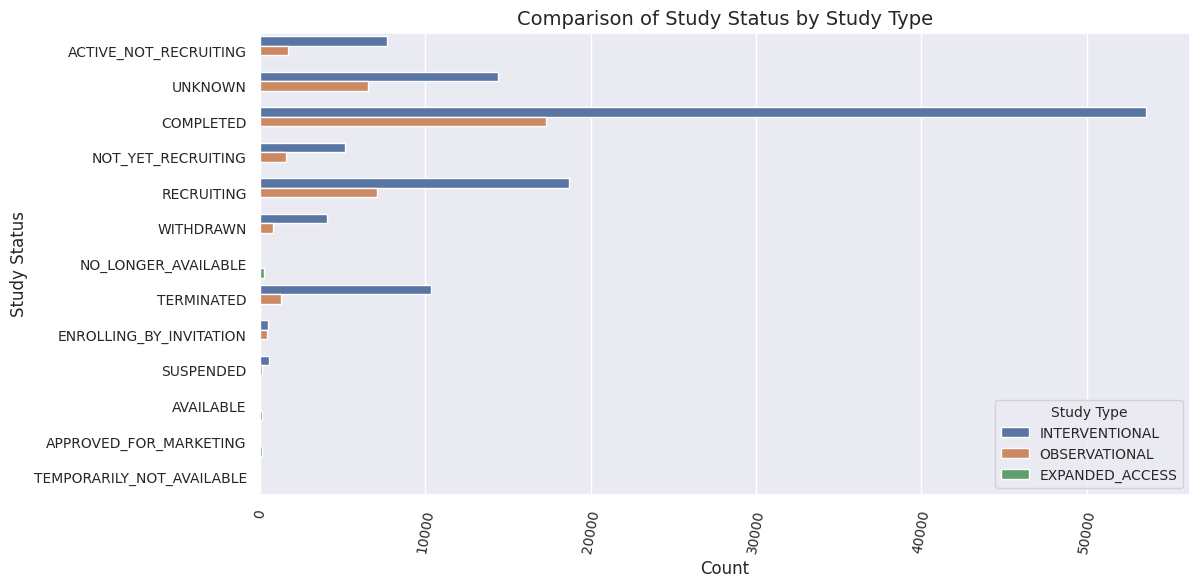
## 

## Bivariate Analysis

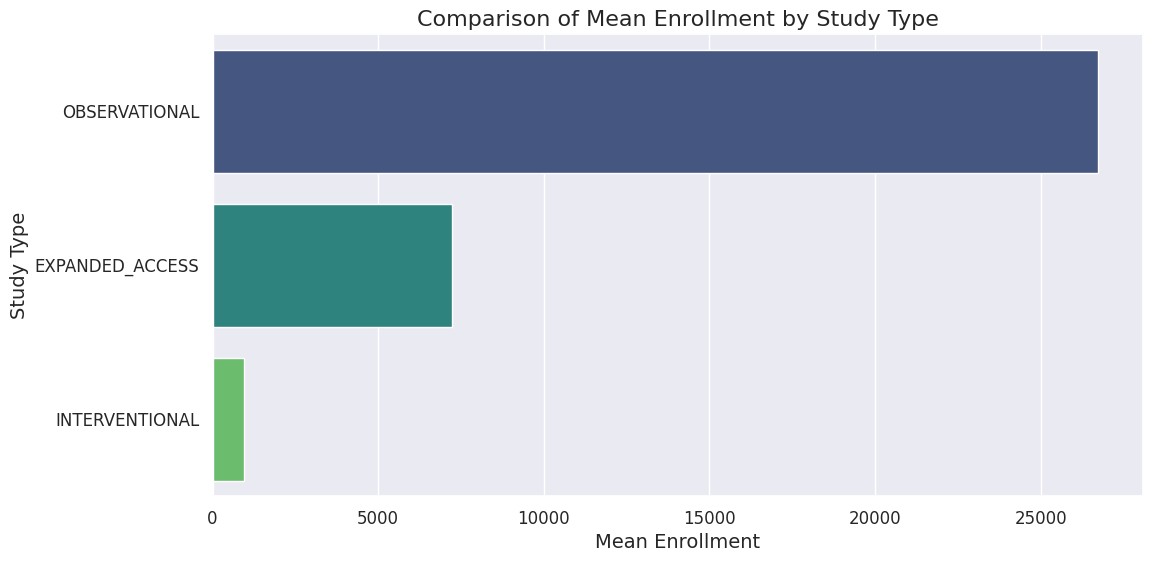
Determine if there is a statistical link between two variables.

The following observations were made from the analysis:

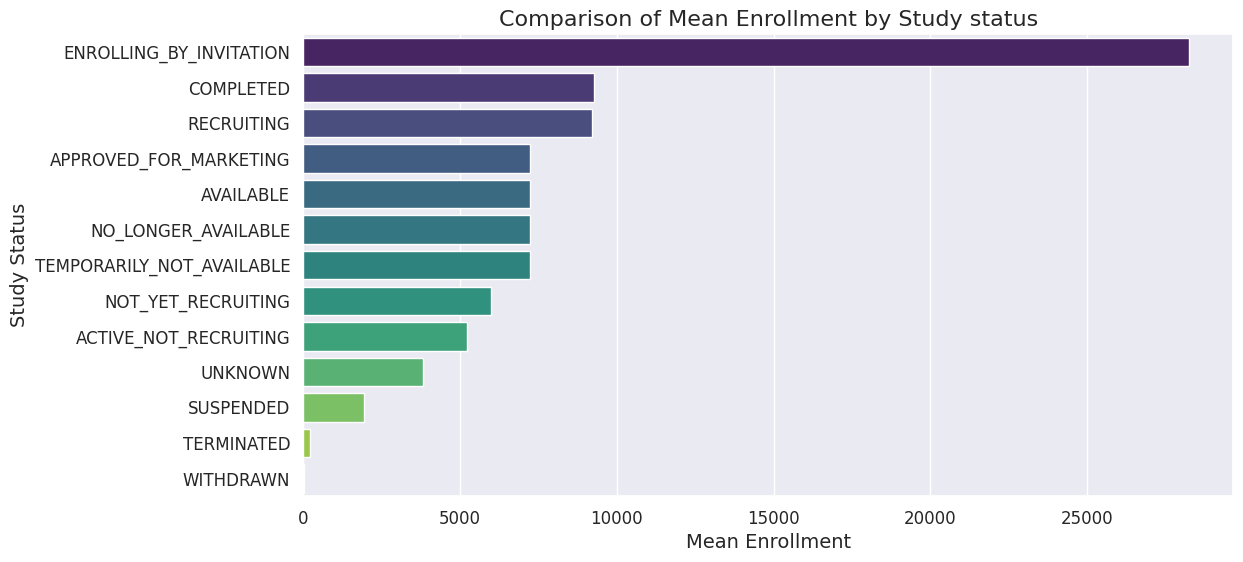
1. Most of the clinical research studies were interventional in nature closely followed by the observational ones.



1. Most of the completed and currently recruiting clinical research trials are interventional.
2. The completed interventional trials confirm the effectiveness/safety of specific interventions guiding medical practice.
3. The completed observational trials offer insights into real world health patterns.
4. The number of participants enrolled in the observational studies was higher than those enrolled in interventional and expanded access studies.



1. The withdrawn and terminated studies were characterized with the lowest enrollment.



# Evaluation

From the research findings , we were able to deduce that :

* There are various research approaches that are employed in clinical trials. The most commonly used according to the analysis is the interventional type of study.
* The data reveals a predominant focus on adult and older adult participants, with limited representation of children and a gap in older adult involvement in clinical studies.
* We identified a select group of diseases that have garnered the most attention in clinical studies. These include Cancer related diseases,HIV/AIDS,Heart disease and COVID-19.
* The analysis highlighted the key sponsors who are driving clinical research.They include, National Cancer Institute, M.D Anderson Cancer Center,Memorial Sloan Kettering Cancer Center among others. Understanding these major stakeholders is crucial for assessing the funding and resource allocation in medical research.
* Investigating the statuses of registered clinical studies, gave insights into the progress and success rates of ongoing research endeavors, which allows us to identify areas of potential improvement.
* From the analysis, the United States has emerged as a prominent hub for clinical studies.This geographic pattern reflects the presence of well-established research institutions, robust healthcare infrastructure, and patient populations conducive to clinical trials. While these hubs drive innovation, it also underscores the need for equitable distribution of research opportunities to ensure that medical advancements benefit a broader and more diverse population.
* The data shows that many studies are done, which is good for progress. Some are still looking for participants, showing ongoing research. But there are also some we don't know much about, and some that got stopped. This tells us that clinical research can be uncertain, and it's important to find out why some studies don't go as planned to improve future trials.
* The predictive model we developed to forecast whether a clinical study is likely to be completed or terminated offers valuable insights into the factors influencing the outcome of these studies. By successfully utilizing this model, we gain a better understanding of the determinants that contribute to the successful completion of trials or their premature termination. This knowledge can guide stakeholders in making informed decisions and optimizing resources to increase the chances of successful clinical studies, ultimately advancing medical research and healthcare.

# Recommendation

We there made the following recommendation:

* Given the prominence of cancer-related diseases, HIV/AIDS, heart disease, and COVID-19 in clinical studies, it is crucial to continue investing in research and allocating resources to these areas. Additionally, ensuring accessible and affordable healthcare for these conditions is essential for better health outcomes.
* There is a need for more collaboration with major sponsors in the clinical trials for impactful partnerships and funding hence successful clinical trials.
* For areas with ongoing research, maintaining a system that monitors the progress and success rates of clinical studies is essential. Periodic reviews and quality assessments can help identify and address any issues, ensuring more efficient and successful trials.
* While established research hubs are important, it's vital to promote a more balanced distribution of clinical studies to benefit a broader population. Encouraging research institutions in less prominent regions can help in achieving a more equitable spread of medical advancements.
* For studies with unknown or terminated statuses, there is a need for in-depth analysis to understand the reasons behind these outcomes. This analysis can guide future research efforts, ensuring a more efficient use of resources and better planning.
* Harness the power of predictive modeling as a pivotal tool in clinical trials. By integrating predictive modeling into the decision-making process, we can significantly enhance our ability to foresee study outcomes. This not only helps us avoid costly project stalls and untimely terminations but also enables us to allocate funds more effectively, maximizing the impact of our investments.