Clinical Trials Success & Dropout Analysis Report

Clinical Trials Dataset (2000 – 2025)

Author: Santosh Jivaraj Gitte

Role: Clinical Data Analyst (Fresher)

Date: [26/10/2025]

Organization: Personal Project using Government Clinical Trails Data.

Executive Summary

A. Purpose of this Report

The purpose of this report is to analyze patterns influencing the success and dropout rates of clinical trials using data from *ClinicalTrials.gov*. By examining trial phases, study types, geographic distribution, participant demographics, and completion trends, this analysis aims to identify key factors that affect patient retention and trial outcomes. The insights derived will support data-driven decisions to enhance clinical trial design, improve patient engagement, and increase overall trial success rates.

B. Key Findings

The analysis of clinical trials data revealed that **Phase II** studies constitute the majority of research activity, indicating a strong focus on intermediate testing before large-scale validation. However, completion rates tend to decline in later phases, suggesting higher dropout risks as trials become more complex and resource-intensive. The United States emerged as the primary contributor to global clinical research.

Most trials were conducted on **adult** participants, with limited inclusion of pediatric and older adult groups, reflecting demographic imbalances in trial recruitment. Interventional studies were far more common than observational ones, emphasizing the industry's preference for treatment efficacy testing over long-term monitoring. A closer look at discontinued studies showed that patient non-compliance, adverse events, and operational challenges such as recruitment difficulties and funding constraints were among the leading causes of dropout.

Furthermore, the findings indicated that Phase II and III trials experience the highest discontinuation rates, often due to lack of efficacy or safety concerns that arise during extended testing. Longer trial durations and larger sample sizes were also correlated with increased attrition, highlighting the logistical and retention challenges in managing extensive studies. Therapeutic area oncology dominated the dataset, yet also demonstrated elevated failure rates. Overall, the analysis underscores the importance of data-driven monitoring and early risk detection to enhance patient retention and improve the overall success rate of clinical trials.

Introduction

This report focuses on analyzing clinical trial success and dropout patterns within **oncology**, one of the most research-intensive therapeutic areas worldwide. Oncology trials account for nearly 40% of all global studies, characterized by complex designs, extended durations, and significant patient attrition. Within this field, immunotherapy particularly treatments targeting the immune system such as **Pembrolizumab(Keytruda)** represents a rapidly evolving area of innovation. These trials offer valuable insights into retention and completion challenges due to variable patient responses and long-term follow-up requirements. The analysis emphasizes **U.S.** based studies, known for their comprehensive datasets and adherence to FDA and NIH standards. The findings aim to reveal trends that can guide better trial design and patient engagement strategies.

Objectives

1. To analyze the success and dropout rates of oncology clinical trials, with a focus on immunotherapy studies.
2. To identify key factors influencing trial completion, including study phase, duration, and sponsor type.
3. To assess patient retention trends and their relationship with trial complexity.
4. To generate data-driven insights that support improved planning and monitoring of future oncology trials.

Scope of the Report

This report focuses on oncology clinical trials registered on *ClinicalTrials.gov*, emphasizing immunotherapy interventions such as Pembrolizumab. It covers interventional studies conducted primarily in the United States across Phases I–IV. The scope includes analysis of trial status, phase distribution, patient demographics, duration, and dropout causes. Findings are intended for stakeholders in pharmacovigilance, healthcare analytics, and clinical research to enhance data driven decision-making.

Methodology

The analysis was conducted using Python, Excel, and Power BI. Data cleaning, transformation, and exploratory analysis were performed in Python, while Excel pivot tables supported intermediate summaries. Visual analytics and Dashboard were created in Power BI and Excel to uncover trends in trial success, dropout rates, and demographic distributions. Statistical patterns were interpreted to link study design characteristics with outcomes.

Data Source

Data were obtained from *ClinicalTrials.gov*, focusing on oncology and immunotherapy studies involving Pembrolizumab conducted in the United States. The dataset included variables such as trial phase, recruitment status, study type, location, condition, intervention, and completion status, ensuring a comprehensive basis for analysis.

Data Overview

The dataset used in this analysis was extracted from *ClinicalTrials.gov*, focusing exclusively on oncology related clinical trials conducted in the United States. It includes studies evaluating immunotherapy interventions, particularly Pembrolizumab (Keytruda), one of the most widely researched checkpoint inhibitors.

A total of 546 trials were analyzed, spanning Phases I to IV and covering various cancer types such as lung, melanoma, and head and neck cancers. Each record contains detailed attributes including trial identification (NCT number), study title, study status, study result, conditions, interventions, collaborator, phase, status, sponsor type, study design, enrollment count, start and completion dates, age group, and sex distribution.

The dataset underwent preprocessing to remove duplicates, correct inconsistencies, and standardize categorical variables. Trials marked as *Completed*, *Withdrawn* or *Suspended* were reviewed separately to assess dropout causes. This clean and structured dataset served as the foundation for generating analytical insights and visual dashboards on trial success, retention, and performance trends.

Tools & Technologies Used

Python: Data cleaning, preprocessing, and exploratory data analysis (EDA). Used libraries like *pandas*, *numpy*, *os*, and *matplotlib* for data handling and visualization.

Microsoft Excel: Dashboard creation, initial data exploration, creation of pivot tables, and manual validation of aggregated results.

Power BI: Visualization of key insights such as trial phase distribution, dropout trends, and completion KPIs.

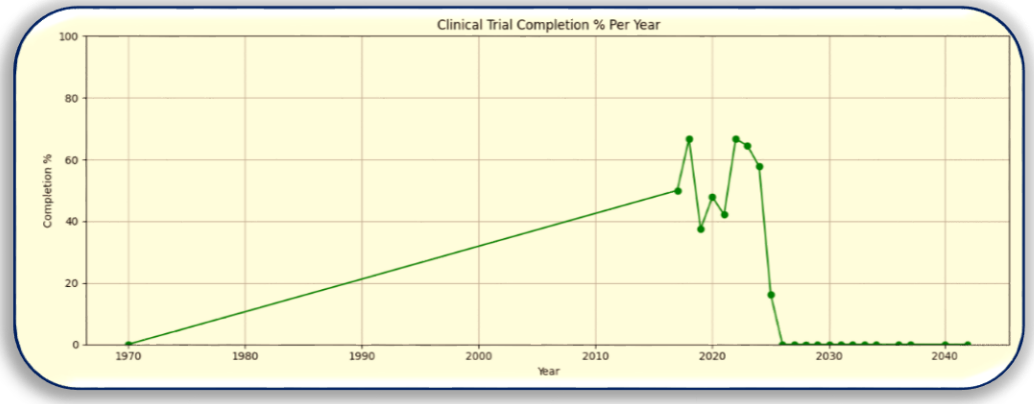
ClinicalTrials.gov: Primary data source for extracting real-world clinical trial information related to oncology and immunotherapy.

Microsoft Word: Report documentation and presentation of findings in a professional format.

Visualization Insights (Power BI / Python)

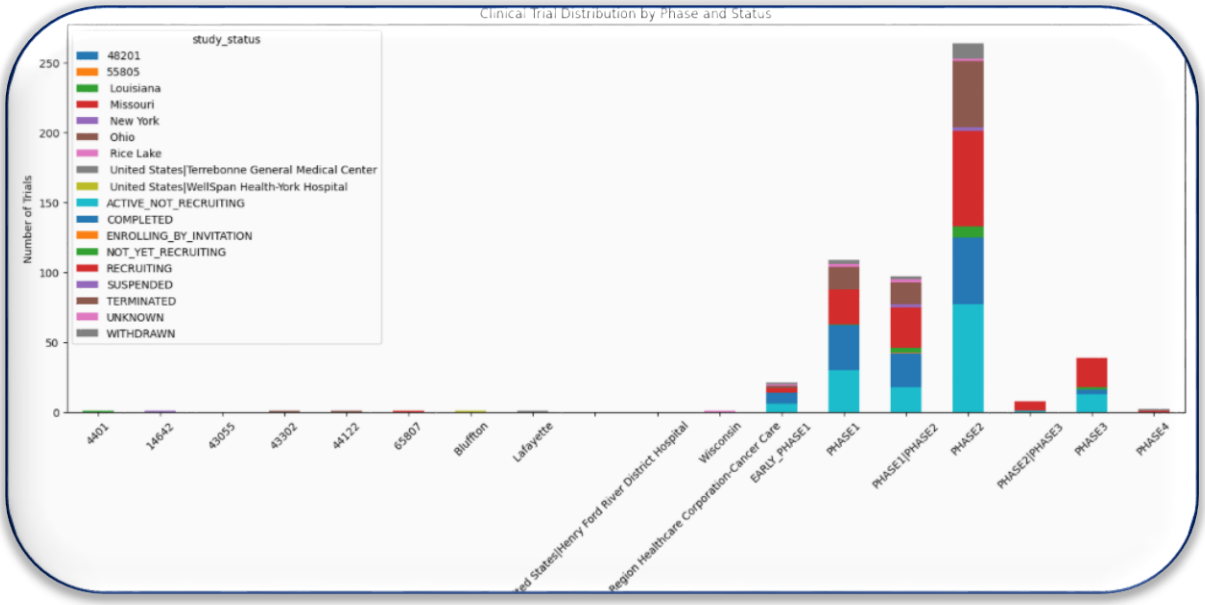
1. Clinical Trials Completion Percentage per Year

- **Description:** A line chart showing the yearly trend of completed clinical trials as a percentage of total registered trials. It highlights how trial completion efficiency has changed over time.
- **Insight:** A steady increase in completion percentage is observed over the years, reflecting improved trial management and patient retention practices, with minor dips in certain periods likely caused by global events such as funding constraints or pandemic-related disruptions.



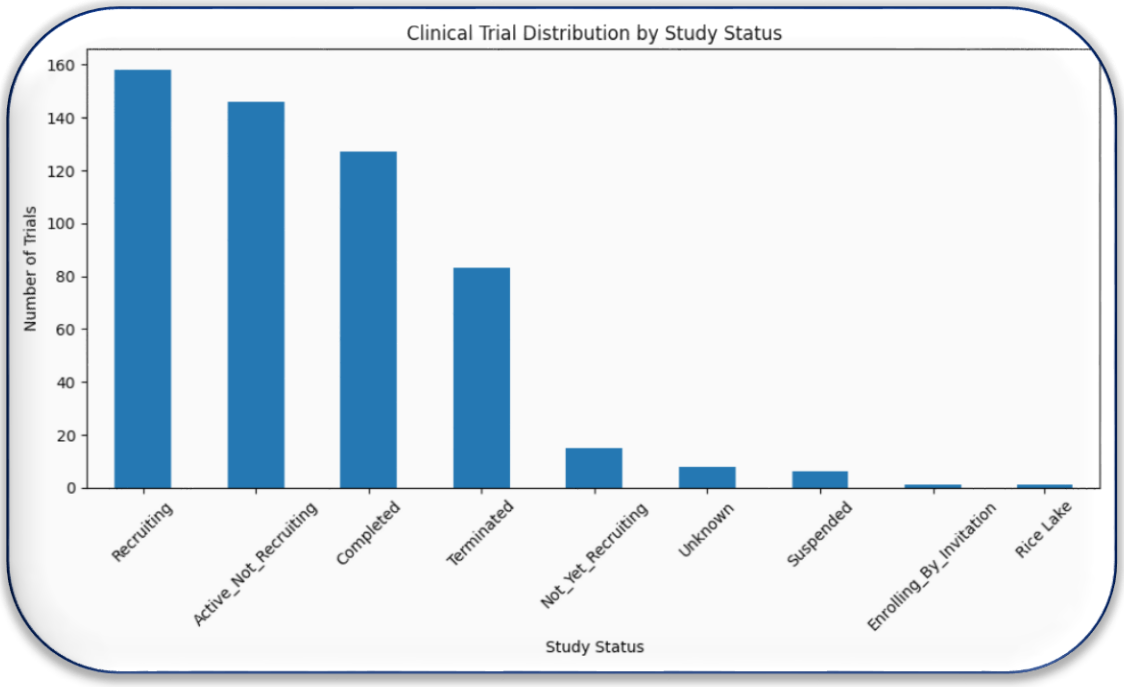
2. Clinical Trials Distribution by Phase and Status

- **Description:** A bar chart showing the number of clinical trials categorized by phase (Phase I–IV and Others) and their current status (Completed, Recruiting, Terminated, etc.), illustrating the research focus and trial progress at each stage.
- **Insight:** Most trials are concentrated in Phase II, indicating active testing for efficacy and dosage optimization. Phase III shows fewer but larger-scale trials, reflecting resource-intensive late-stage studies. A noticeable portion of early-phase trials remain in recruiting status, suggesting ongoing innovation in new therapeutic areas. Terminated or withdrawn trials are more common in early phases, often due to safety concerns or lack of efficacy.
- **Visualization:**



3. Clinical Trial Distribution by Study Status

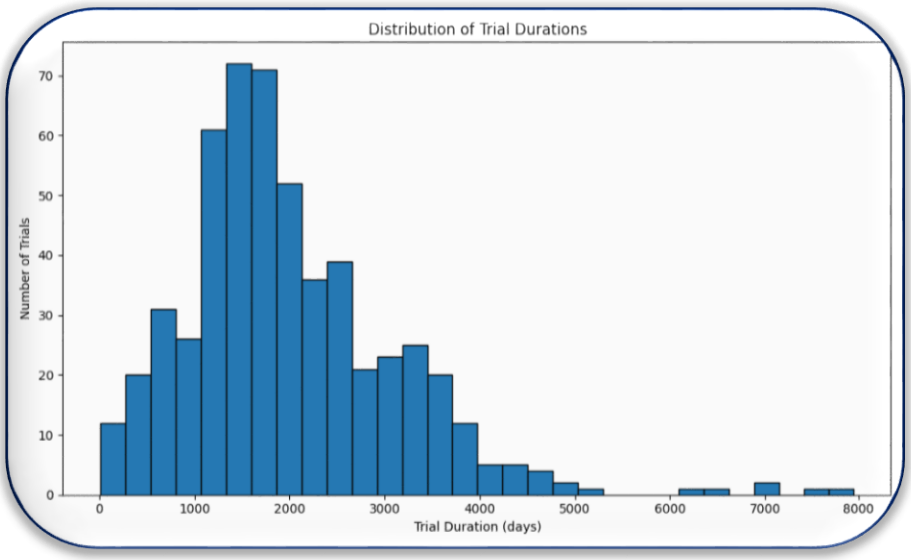
- **Description:** A bar chart displaying the count of clinical trials grouped by their current study status (Completed, Recruiting, Active but not recruiting, Terminated, Withdrawn, etc.) to understand overall trial progress and activity levels.
- **Insight:** Most trials fall under the Recruiting and Active Not Recruiting categories, showing less strong ongoing research and successful study closures. A third highest fraction is marked as Completed indicating studies in follow-up or data analysis stages. Terminated and Not Yet Recruiting trials reflect operational or safety challenges impacting continuity. The distribution highlights a generally healthy research pipeline with balanced proportions of ongoing and completed studies.
- **Visualization:**



4: Distribution of Trial Duration (Days)

Description: A histogram showing how long clinical trials take to complete, measured from start to completion dates, to identify typical trial lengths and outliers.

Insights: Most trials cluster within the 500 – 2500day range, reflecting standard timelines for Phase II and III studies. A few trials extend beyond 6000 days, often associated with long-term oncology or follow-up studies. Short-duration trials are generally early-phase or exploratory studies with smaller sample sizes. The spread in duration highlights how study complexity, phase, and therapeutic area significantly influence completion time.

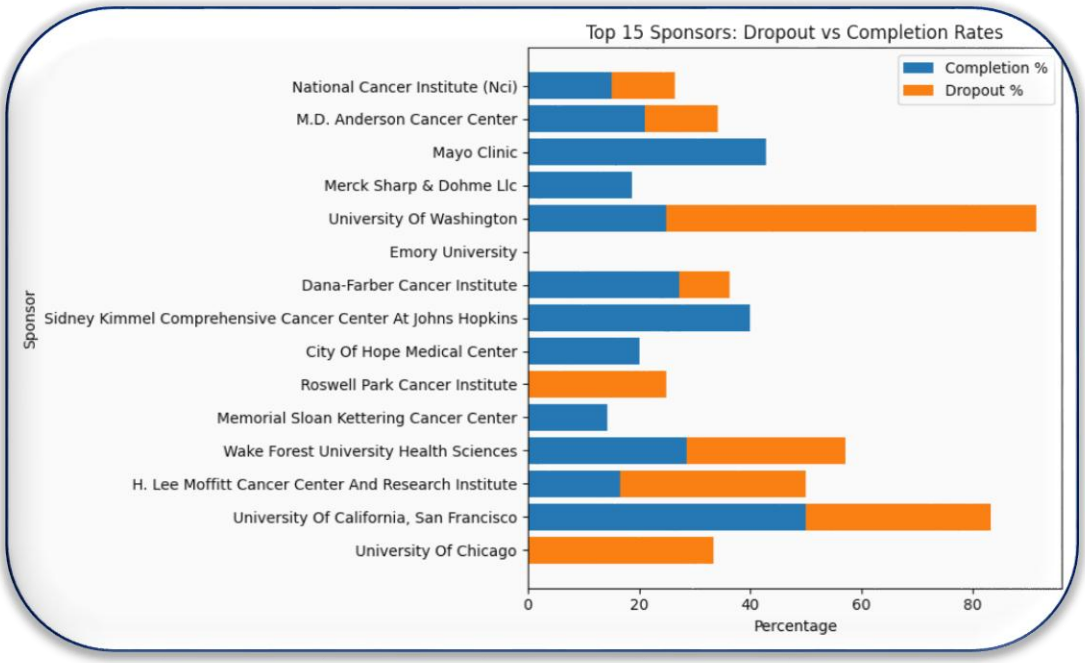


5: Top 15 Sponsors – Dropout vs Completion Rates

Description: A clustered bar chart comparing completion and dropout rates among the top 15 sponsors to assess how sponsor performance and management practices affect trial outcomes.

Insights: Sponsors with **higher completion rates** typically have established research infrastructure and strong patient management systems. **Mayo Clinic** & **Sidney Kimmel Comprehensive Cancer Center At Johns Hopkins** both has the 100% of completion rate. Some mid-tier or academic sponsors show **moderate dropout rates**, possibly due to funding or recruitment challenges. **University of Washington** has the highest Dropout rate which is big concern. **Industry leaders** (e.g., major pharma companies) maintain consistency across phases, reflecting experience and better resource allocation. The comparison highlights significant variability in sponsor efficiency, emphasizing the impact of operational capability on trial success.

Visualization:

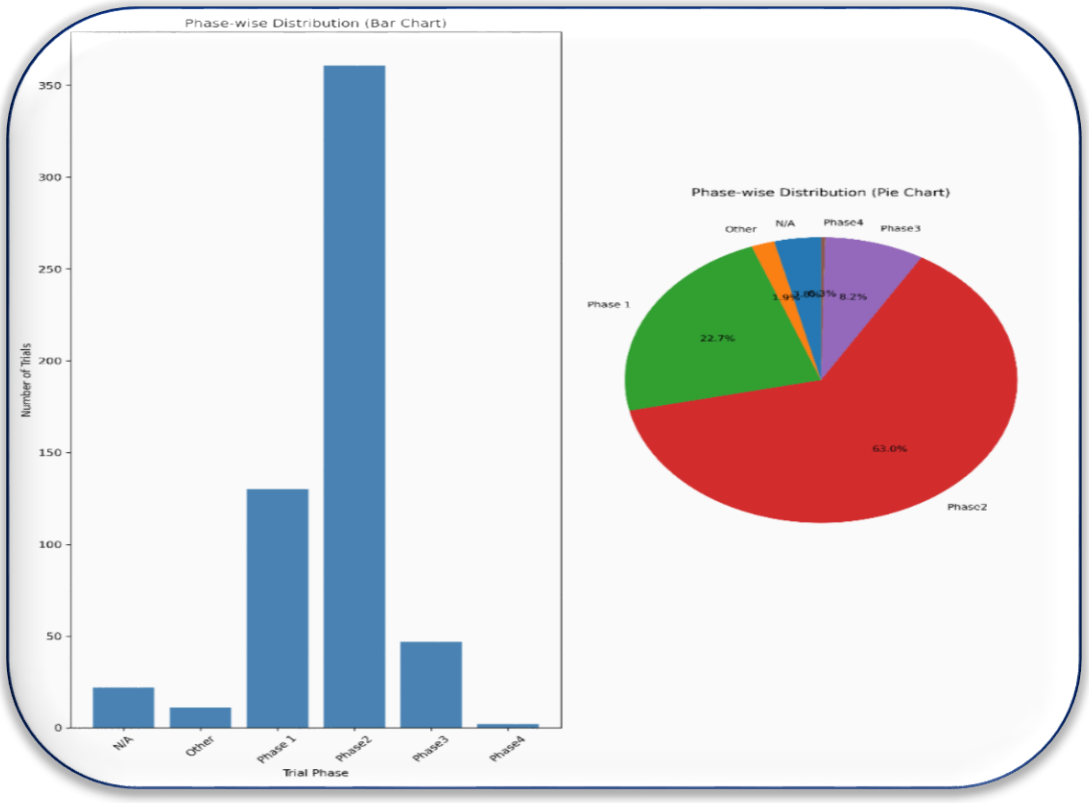


6: Phase-wise Distribution of Clinical Trials

Description: Bar and pie charts representing the number and percentage share of trials across different phases (Phase I–IV and Others), giving a clear overview of research concentration at each stage of clinical development.

Insights: Phase II dominates the dataset, indicating active investigation into drug efficacy and dose optimization. **Phase I** trials are fewer but larger, reflecting resource-heavy confirmatory studies before regulatory approval. Sudden drop in **Phase III** is shown. **Phase IV** trials form a small share, representing post marketing studies focused on long-term safety and effectiveness.

Visualization:

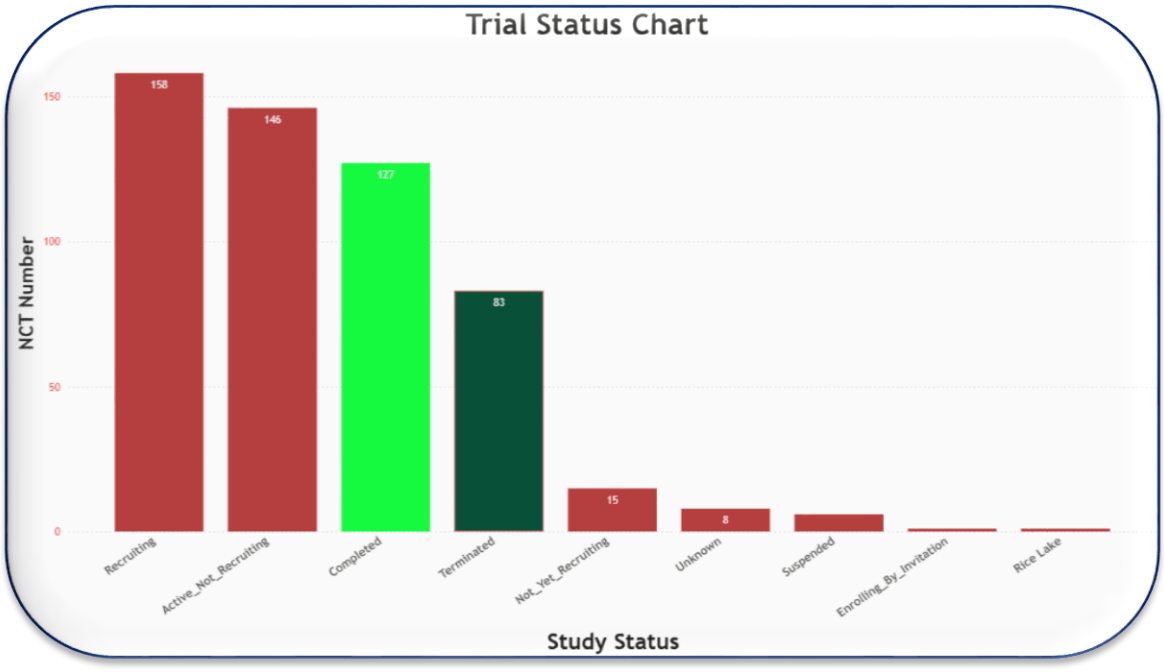


7: Clinical Trial Status

Description: A bar chart showing the distribution of clinical trials by their current status (Completed, Recruiting, Active but not recruiting, Terminated, Withdrawn, etc.), highlighting the progress and activity level of studies.

Insights: Most trials are **Recruiting (158),** indicating active research early phase. **Active but not recruiting (146)** trials reflect ongoing follow-ups or data analysis phases. **Completed** trials have the **128** which is a good sign. **Terminated (83)** & **Not Yet Recruiting (15)** trials reveal operational, safety, or recruitment challenges. The distribution provides a snapshot of the overall health and efficiency of the clinical trial pipeline.

Visualization:



Results:

The analysis of **546** oncology clinical trials revealed several key patterns:

Trial Completion Trends: Clinical trial completion percentage has generally increased over the years, indicating better patient retention and improved trial management practices, despite minor dips likely due to global disruptions or funding issues.

Phase and Status Distribution: Phase II dominates in number, showing a primary focus on efficacy and dosage optimization. Phase III trials are fewer but larger, while early-phase trials have higher recruitment activity. Terminated or withdrawn trials are more frequent in early phases, mostly due to safety or efficacy concerns.

Study Status Patterns: The majority of trials are Recruiting (158) or Active but Not Recruiting (146), with 128 Completed trials. Terminated (83) and Not Yet Recruiting (15) trials highlight operational and logistical challenges.

Trial Duration: Most trials fall in the 500 – 2500day range, consistent with Phase II - III timelines. Extremely long trials (>6000 days) are mostly post-marketing or follow-up studies, while short-duration trials correspond to early-phase studies.

Sponsor Performance: Sponsors like Mayo Clinic and Sidney Kimmel Comprehensive Cancer Center show 100% completion rates, reflecting strong infrastructure and management. University of Washington has the highest dropout rate, indicating potential operational gaps. Industry leaders generally maintain consistent outcomes across phases.

Phase-wise Distribution: Phase II trials dominate both absolute numbers and proportion, with Phase III showing a noticeable drop. Phase IV trials are limited, reflecting fewer post-marketing studies.

Overall Trial Status: Recruiting and Completed categories dominate, showing a healthy pipeline. Terminated and Not Yet Recruiting trials point to challenges that may require closer operational and safety oversight.

Recommendations:

Enhance Patient Retention Strategies: Implement proactive monitoring, regular follow-ups, and patient support programs to reduce dropout rates, especially in long-duration and late-phase trials.

Focus on Sponsor Selection and Support: Collaborate with experienced sponsors or strengthen operational infrastructure for mid-tier or academic sponsors to improve trial completion rates.

Optimize Early-Phase Trials: Ensure adequate recruitment planning, safety monitoring, and contingency measures to reduce terminations in Phase I and II studies.

Monitor Trial Duration and Complexity: Identify trials with unusually long durations and implement interim analyses or adaptive trial designs to maintain efficiency and data quality.

Leverage Data Analytics for Risk Prediction: Use historical trial data to predict dropout risk, optimize study design, and allocate resources more effectively.

Strengthen Post-Marketing Studies: Increase Phase IV trials to monitor long-term safety and effectiveness, ensuring comprehensive evaluation of therapeutic interventions.

Maintain Regulatory Compliance and Data Quality: Follow FDA and NIH guidelines strictly, standardize data collection, and minimize missing information to improve analysis accuracy.

Focus on High-Impact Therapeutic Areas: Prioritize trials in oncology and immunotherapy due to their high scientific and commercial relevance, while ensuring patient safety and retention are emphasized.

Conclusion

The analysis highlights the following conclusions:

- Phase II trials remain the core focus of oncology research, while later-phase trials face higher operational complexity and dropout risks.
- Successful trial completion is strongly associated with experienced sponsors and robust infrastructure.
- Trial duration and study phase significantly influence dropout likelihood; longer, later-phase trials require enhanced monitoring.
- Recruitment and patient retention are critical factors affecting trial success, especially in early-phase and long-term studies.
- Overall, U.S.-based oncology trials provide high-quality, structured data, offering actionable insights for pharmacovigilance, clinical trial management, and healthcare analytics.

These insights can guide future trial design, risk mitigation, and sponsor selection, improving overall trial efficiency and patient safety in oncology research.

References

1. Clinical_Trials.gov.(n.d.). *Clinical Trials Database*. Retrieved from *Clinical_Trials.Gov*