Unveiling Disparities in Clinical Trials: Success Rates, Funding, and Activity for Overlooked High-Impact Diseases

Anant Vishwakarma, Nafiz Imtiaz Khan, Zeerak Babar

Abstract

This study investigates disparities in clinical trial activity, success rates, and funding allocation for high-impact yet overlooked diseases, including pancreatic cancer, Chagas disease, endometriosis, drug-resistant tuberculosis, and Duchenne muscular dystrophy (DMD). Using data from ClinicalTrials.gov and PubMed, we compare these diseases with better-funded conditions such as Alzheimer's, breast cancer, influenza, hepatitis, and malaria to assess research prioritization and investment gaps. Our analysis reveals significant imbalances in clinical trial volume, progression rates across trial phases, and the distribution of public versus private funding. Diseases with high mortality but limited commercial appeal, such as Chagas disease and drug-resistant tuberculosis, exhibit significantly lower trial activity and success rates compared to diseases with strong pharmaceutical investment. Furthermore, industry-sponsored trials demonstrate higher progression rates, while publicly funded trials face greater delays and attrition. These findings highlight the critical need for equitable research investment and policy interventions to accelerate drug development for neglected diseases.

1 Introduction

Clinical trials are fundamental to advancing medical treatments [1]; however, research efforts are often unevenly distributed. It often favor diseases with substantial pharmaceutical investment and public awareness. This study focuses on five high-impact but overlooked diseases: **pancreatic cancer**, **Chagas disease**, **endometriosis**, **drug-resistant tuberculosis**, and **Duchenne muscular dystrophy (DMD)**. Despite their severe health implications, these diseases receive significantly less research funding and clinical trial activity compared to common diseases. In this project, we analyze these disparities by comparing the five overlooked high-impact disease with five most common diseases. The five most common diseases [2] in our study includes: Alzheimer's disease[3], hepatitis [4], influenza [5], malaria [6], and breast-cancer [7].

The detailed description of the overlooked high-impact diseases are as follows: **Pancreatic cancer** is one of the most lethal cancers, with a five-year survival rate of less than 12% [8]. The low percentage of survival rate is due to late-stage diagnosis and limited treatment options [9]. Thus, research into improving clinical trial success rates and funding allocation is essential for accelerating progress against this disease.

Similarly, **Chagas disease**, a neglected tropical disease affecting millions, particularly in Latin America, has far fewer clinical trials compared to malaria and HIV, despite its significant morbidity and mortality burden [10, 11]. This lack of research investment has hindered the development of effective new treatments for that disease.

Endometriosis, a chronic condition affecting one in ten women of reproductive age, remains understudied compared to male reproductive health conditions like prostate disease [12]. This imbalance in clinical trials has led to delayed diagnoses and inadequate treatment options for millions of affected individuals [13].

Drug-resistant tuberculosis (TB) is another urgent global health crisis [14]. While the rapid development of COVID-19 vaccines demonstrated the potential for accelerated clinical trials, TB treatment

advancements remain slow, prolonging the global burden of the disease [15]. Understanding why TB trials take longer and receive less funding is crucial for combating this infectious threat.

Lastly, **Duchenne muscular dystrophy (DMD)**, a severe genetic disorder primarily affecting children, exemplifies the disparities in pediatric versus adult disease research funding [16]. Rare pediatric diseases consistently receive less pharmaceutical investment, delaying the development of effective treatments [17]. Investigating why industry sponsorship is lower for rare diseases can help identify ways to incentivize investment and accelerate drug development.

2 Research Questions and Objectives

The research questions explored in this study address critical gaps in clinical trial activity, success rates, and funding disparities. Understanding how many clinical trials are currently active for each disease and how this compares to better-funded conditions can help assess whether research prioritization aligns with disease burden. If severe but neglected diseases have significantly fewer trials than those with lower mortality rates, this signals a systemic research gap that must be addressed. Thus, our first research question is:

> **RQ**$_1$: How many clinical trials are currently active for each disease, and how does this compare to better-funded diseases?

Evaluating clinical trial success rates across different phases can provide insight into disease-specific barriers to drug development. Some diseases might struggle to progress beyond Phase 1 or 2 due to high costs, patient recruitment challenges, or scientific limitations. Analyzing the transition rates from Phase 1 to Phase 3 and FDA approval highlights whether certain diseases face disproportionately high failure rates. Consequently, we construct our second research question:

> **RQ**$_2$: What is the success rate of clinical trials across different phases, and which diseases struggle to progress?

A key driver of research activity is funding allocation. Investigating how clinical trial funding compares between rare, high-mortality diseases and more common conditions helps determine if research fundings are being allocated equitably. Many neglected diseases rely on government funding rather than industry sponsorship, which can lead to slower drug development. Understanding the role of public versus private funding provides insights into whether diseases with lower commercial profitability receive adequate research investment. Thus, our third research question is:

> **RQ**$_3$: How does clinical trial funding compare between rare/high-mortality diseases and more common diseases?

Furthermore, getting government funds might be easy for certain diseases, compared to getting funds from industry. Whether a study is being funded by government or by Industry might also have omapt on clinical trial success rates. Industry investment often accelerates drug approval timelines due to greater resources and streamlined processes, while government-funded trials may focus on scientific advancement rather than commercialization. By comparing industry versus NIH-funded trial success rates, the impact of funding source disparities on treatment development can be identified. Thus, our final research question is:

> **RQ**$_4$: Are certain diseases more likely to have industry-sponsored vs. government- sponsored trials, and does this impact success rates?

By answering the above mentioned research questions, this project contributes to the broader discourse on healthcare equity and research funding disparities. In addition, it provides data-driven insight to inform policy changes, funding decisions, and future clinical trial investment strategies.

Table 1: Data Collection Summary for Rare and Common Diseases.

Disease Type	Disease Name	Associated Clinical Trials	Associated PubMed Articles
Rare	Pancreatic Cancer	2580	9831
	Endometriosis	589	2839
	Chagas Disease	55	680
	Drug Resistant Tuberculosis	60	1274
	Duchenne Muscular Dystrophy	364	1423
	Total	**3648**	**16047**
Common	Alzheimer's	2169	60
	Influenza	2303	8905
	Breast Cancer	5000	171
	Hepatitis	3517	7087
	Malaria	1129	6855
	Total	**14118**	**25078**

3 Methodology

3.1 Data Collection

In our project, we collected data from two sources. The first source was the ClinicalTrials.gov API [18], which provides detailed information about clinical trials. We used this API to gather data on clinical trials related to five rare diseases: pancreatic cancer, Chagas disease, endometriosis, drug-resistant tuberculosis, and Duchenne muscular dystrophy. Also, to answer RQ3, we gathered data on five most common diseases: alzheimer's, influenza, breast cancer, hepatitis, and malaria. Each disease was queried using its name as the search parameter. During data collection, we encountered pagination, which limits the number of results per page. To handle this, we utilized the pageToken parameter to retrieve subsequent pages. We set the pageSize parameter to 100 and implemented a loop to fetch data until no nextPageToken remained. The raw data was stored in JSON format for further analysis.

The second source was the PubMed API [19], which provides access to scientific articles. We used the PyMed library to query PubMed for articles related to the same five diseases. To interact with the API, we specified an example Gmail account as the email parameter and set tool = "PubMedSearcher". Articles were retrieved using the disease name combined with the keyword "Clinical Trials" to refine results. We set the maximum number of results to 10,000 to ensure comprehensive coverage. From the retrieved articles, we extracted key metadata, including PubMed ID, title, keywords, journal, abstract, methods, results, conclusions, copyrights, DOI, publication date, and authors. We stored the data from in CSV format for further analysis.

3.2 Data Processing

For the Clinical Trials data, the raw dataset contained detailed descriptions of various trial phases. However, for our study, we needed specific fields, thus, we extracted only the relevant keys from the raw JSON data. A list of the selected fields along with their descriptions is provided in Table 2. The Phases field indicates the different phases the study has gone through. If a study includes multiple phases, we consider only the most recent phase. For instance, if a trial has recorded phases 1, 2, and 3, we classify it as phase 3. Similarly, if a study is categorized as early-phase-1, we treat it as phase-1.

Table 2: Detailed Descriptions of Extracted Clinical Trial Fields

Field Name	**Description**
NCT ID	A unique identifier assigned to each clinical trial by ClinicalTrials.gov, used to track and reference specific studies.
Acronym	A short name or abbreviation by which the trial is commonly known, often used for ease of reference.
Overall Status	The current state of the clinical trial, indicating whether it is ongoing, completed, withdrawn, or in another status such as recruiting or suspended.
Conditions	The diseases, disorders, or medical conditions that the trial aims to study, prevent, diagnose, or treat. These conditions define the primary focus of the research.
Interventions	The treatments, medications, procedures, behavioral therapies, or other actions being investigated in the trial to assess their effectiveness and safety.
Locations	The geographic sites where the clinical trial is being conducted, including hospitals, research centers, and other facilities where participants are enrolled.
Primary Completion Date	The date when data collection for the primary outcome measure is completed for all participants, marking a critical milestone in the study timeline.
Study First Post Date	The date when the clinical trial record was first published on ClinicalTrials.gov, making it publicly accessible for researchers, clinicians, and patients.
Last Update Post Date	The most recent date when the trial information was updated on ClinicalTrials.gov, reflecting any modifications to the study's status, design, or results.
Study Type	The classification of the trial based on its design and purpose, such as interventional (testing a treatment) or observational (studying outcomes without intervention).
Phases	The phase of the clinical trial, indicating its stage of research (e.g., Phase 1: safety testing, Phase 2: efficacy and safety, Phase 3: large-scale efficacy, Phase 4: post-market surveillance).
Lead Sponsor Name	The name of the organization, institution, or company responsible for initiating, managing, and funding the clinical trial.
Lead Sponsor Type	The category of the trial sponsor, such as industry (pharmaceutical companies), academic institutions (universities, research centers), or government agencies (NIH, FDA).

Table 3: Distribution of Clinical Trials with and without Associated Publications for Common and Rare Diseases

Disease-Type	Trials-with-Associated-Articles (%)	Trials-without-Associated-Articles (%)	Total
Common	1,137 (8.05%)	12,981 (91.95%)	**14,118**
Rare	422 (11.57%)	3,226 (88.43%)	**3,648**

For preprocessing the PubMed data, we extracted five key fields from the raw JSON data: PubMed ID, which serves as a unique identifier for each article; Title, representing the study's name; Keywords, which capture relevant terms to aid in information retrieval; and Journal, which records the name of the journal where the article was published.

3.3 Linking Clinical Trials with PubMed Articles

In this step, we identified which clinical trials have associated articles in PubMed. Typically, when researchers publish an article related to a specific clinical trial, they include the trial's unique identifier (NCT-ID) in the abstract. Leveraging this convention, we examined the collected PubMed data to determine which articles mention an NCT-ID and whether the mentioned ID matches any of the collected NCT-IDs from the clinical trial dataset. The results of this linkage analysis are presented in Table 3.

From Table 3, we observe that a significant majority of clinical trials do not have associated articles in PubMed. Specifically, only 8.05% of common disease trials and 11.57% of rare disease trials have linked publications, indicating that rare disease trials are slightly more likely to be published in the literature. However, in both cases, the vast majority of trials remain unpublished, with 91.95% of common disease trials and 88.43% of rare disease trials lacking associated articles. This highlights a potential gap in the dissemination of clinical trial findings, particularly for common diseases, where a lower percentage of trials have linked publications despite their higher overall volume.

3.4 Analyzing the Data

In this step, we conducted exploratory data analysis (EDA) to address our research questions. We began by examining active clinical trials for both common and rare diseases. Next, we analyzed the distribution of trials across different phases, followed by an assessment of the funding status for these diseases. Finally, we investigated the impact of government versus industry funding on trial success rates.

For data processing and analysis, we utilized Python's `pandas` and `numpy` libraries, while visualizations were generated using `matplotlib` and `plotly`.

4 Findings and Analysis

4.1 RQ_1: How many clinical trials are currently active for each disease, and how does this compare to better-funded diseases?

The analysis of active clinical trials reveals significant disparities in research prioritization among high-impact but overlooked diseases. Pancreatic cancer has the highest number of active trials, with 649 ongoing studies (486 recruiting and 163 active but not recruiting). In contrast, endometriosis has 135 active trials, while Duchenne muscular dystrophy (DMD) has 81. Diseases such as drug-resistant tuberculosis (9 active trials) and Chagas disease (5 active trials) are critically underrepresented in clinical research.

When compared to better-funded diseases, the differences become even more pronounced. Breast cancer has 797 actively recruiting trials, Alzheimer's has 356, and hepatitis has 199. These numbers significantly surpass those of neglected diseases. Malaria, though a high-burden infectious disease, still

has more active trials (51) than Chagas disease and drug-resistant tuberculosis. Breast cancer alone has over 4,500 total clinical trials, highlighting the massive investment gap.

These numbers highlight a stark funding and research gap. The limited number of active trials for Chagas disease and drug-resistant tuberculosis suggests a lack of investment in diseases that primarily affect lower-income populations, despite their significant global burden. The relatively higher trial activity for pancreatic cancer indicates increased research attention but remains significantly lower than better-funded cancers. This disparity in clinical trial activity suggests that diseases with strong pharmaceutical investment receive far more research attention, while those with high mortality but low commercial interest struggle to attract funding and resources. Future efforts should focus on increasing clinical trial investment in neglected diseases, particularly those with a high unmet medical need.

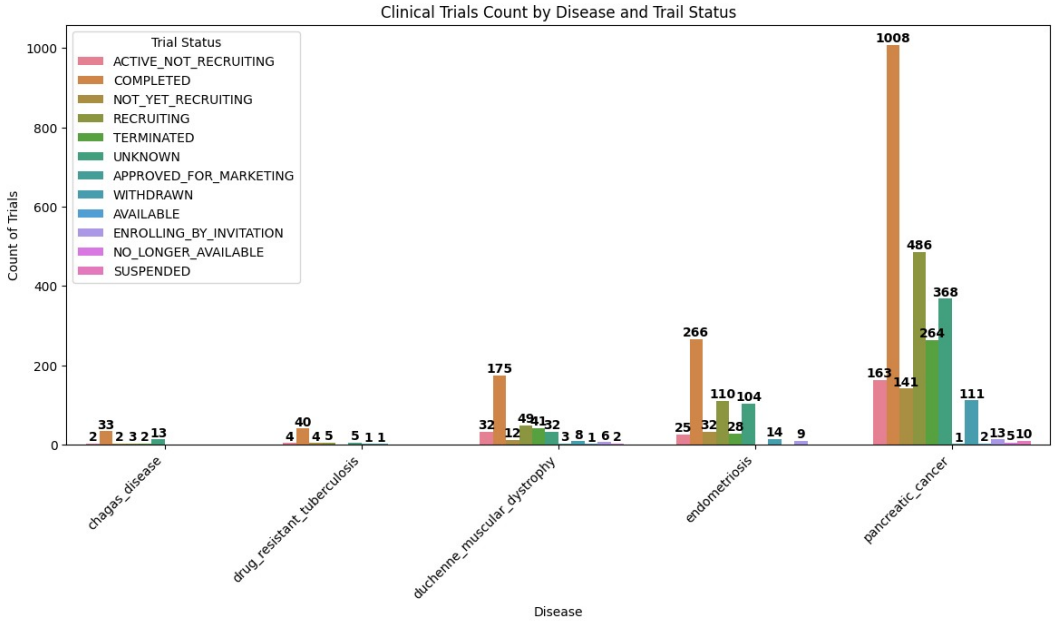


Figure 1: Number of Clinical Trials for Overlooked Diseases

The PubMed dataset further contextualizes these findings by showing that pancreatic cancer research has the highest publication volume (9,831 articles), whereas Chagas disease has the least (680 articles). The research exposure of a disease might correlate with its clinical trial activity, suggesting that diseases with fewer active trials also receive less overall research attention. Furthermore, diseases with strong pharmaceutical backing, such as pancreatic cancer, have research published in high-impact journals like Journal of Clinical Oncology, while neglected diseases like Chagas disease remain confined to niche tropical medicine journals.

RQ_1 Findings:

There are significant disparities in clinical trial activity between overlooked and common diseases. Pancreatic cancer has the highest number of active trials (649), while diseases like drug-resistant tuberculosis (9 trials) and Chagas disease (5 trials) remain critically underrepresented. In contrast, well-funded diseases such as breast cancer (797 recruiting trials) and Alzheimer's (356 recruiting trials) have far more active studies, highlighting a substantial research gap.

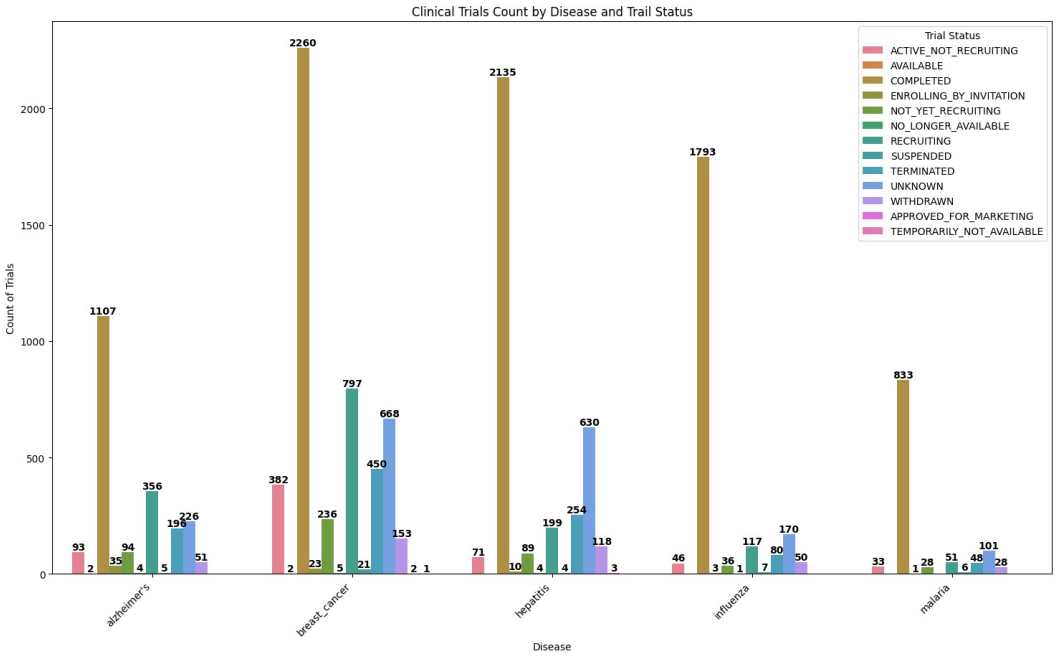
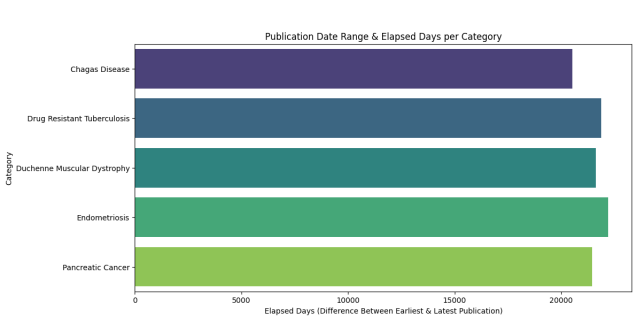
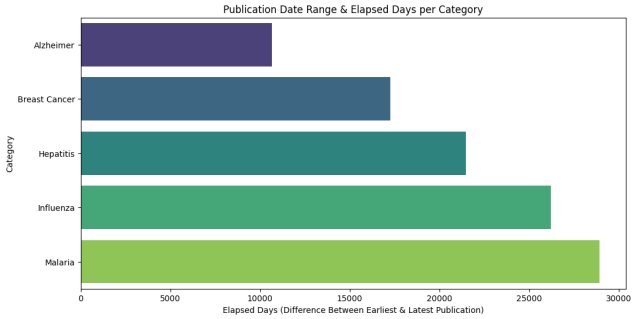


Figure 2: Number of Clinical Trials for Common Diseases



(a) Time span between the first and last publication for rare diseases



(b) Time span between the first and last publication for common diseases

Figure 3: Comparison of publication time spans for rare and common diseases.

4.2 RQ_2: What is the success rate of clinical trials across different phases, and which diseases struggle to progress?

The analysis of clinical trial success rates across different phases reveals significant disparities in how diseases progress through the clinical trial pipeline. Pancreatic cancer exhibits a severe drop-off rate after Phase 2, with 1,020 trials in Phase 2 but only 191 reaching Phase 3. This suggests that a significant portion of pancreatic cancer treatments fail to progress, potentially due to treatment inefficacy, lack of funding for late-stage trials, or high patient dropout rates. Similarly, Chagas disease and drug-resistant tuberculosis show minimal Phase 3 and Phase 4 trials, highlighting that many potential treatments never advance to late-stage clinical testing. Chagas disease, for instance, has only 9 Phase 3 trials, indicating a high attrition rate in the early phases of drug development.

In contrast, Duchenne muscular dystrophy (DMD) and endometriosis exhibit a relatively better transition from Phase 2 to Phase 3, likely due to higher levels of industry funding and better patient recruitment strategies. These findings suggest that late-stage trial bottlenecks exist for diseases with fewer commercial incentives, which ultimately hinders drug approvals and slows medical advancements.

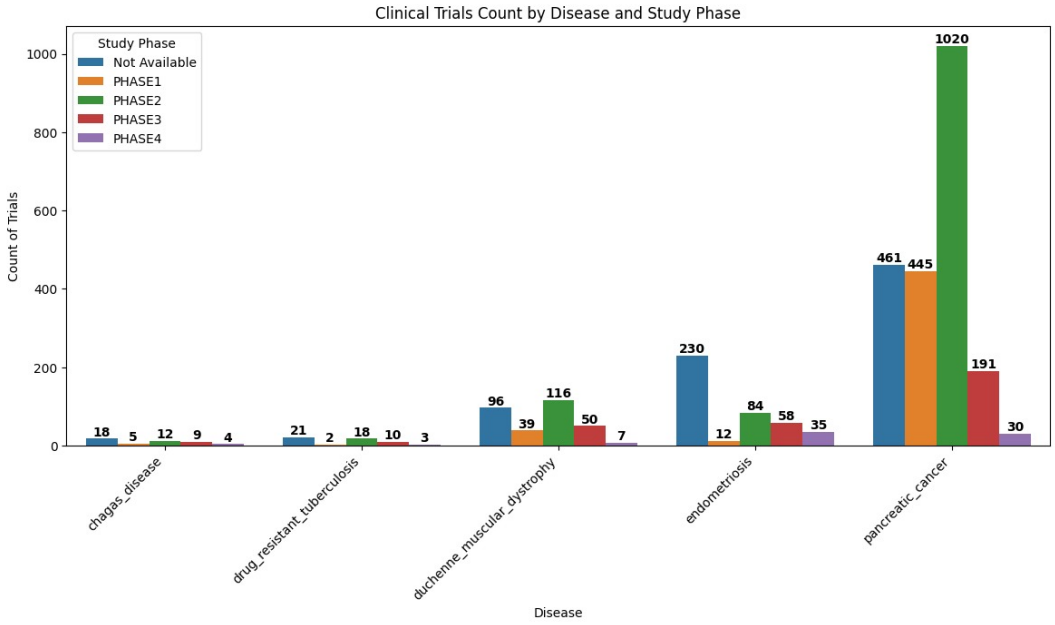


Figure 4: Clinical Trials by Phases for Overlooked Diseases

When compared to well-funded diseases, a clear pattern emerges. Breast cancer has 1,517 trials in Phase 2 and 589 in Phase 3, indicating a much higher transition rate than pancreatic cancer. Alzheimer's, another well-funded disease, has 516 Phase 2 trials but 273 in Phase 3, showing a better success rate than Chagas disease or drug-resistant tuberculosis. Hepatitis has the strongest pipeline, with 597 trials in Phase 3 and 610 in Phase 4, demonstrating an efficient clinical trial transition.

The PubMed dataset further reveals that endometriosis research spans over 22,000 days (since 1964), yet the disease still has only 58 Phase 3 trials, indicating that long-standing research efforts do not necessarily translate into clinical advancements without sufficient funding and trial prioritization.

Similarly, the fact that malaria has 180 Phase 3 trials while Chagas disease has only 9, despite both being neglected tropical diseases, reflects an imbalance in global research prioritization.

RQ_2 Findings: There are significant disparities in clinical trial success rates across different phases. Pancreatic cancer experiences a major drop-off after Phase 2, with only 19% of trials advancing to Phase 3. Chagas disease and drug-resistant tuberculosis struggle even more, with minimal late-stage trials (only 9 Phase 3 trials for Chagas disease). In contrast, common diseases like breast cancer and Alzheimer's show higher transition rates, with 589 and 273 Phase 3 trials, respectively. These findings suggest that diseases with strong pharmaceutical investment progress more efficiently, while those with fewer commercial incentives face significant bottlenecks in late-stage clinical trials.

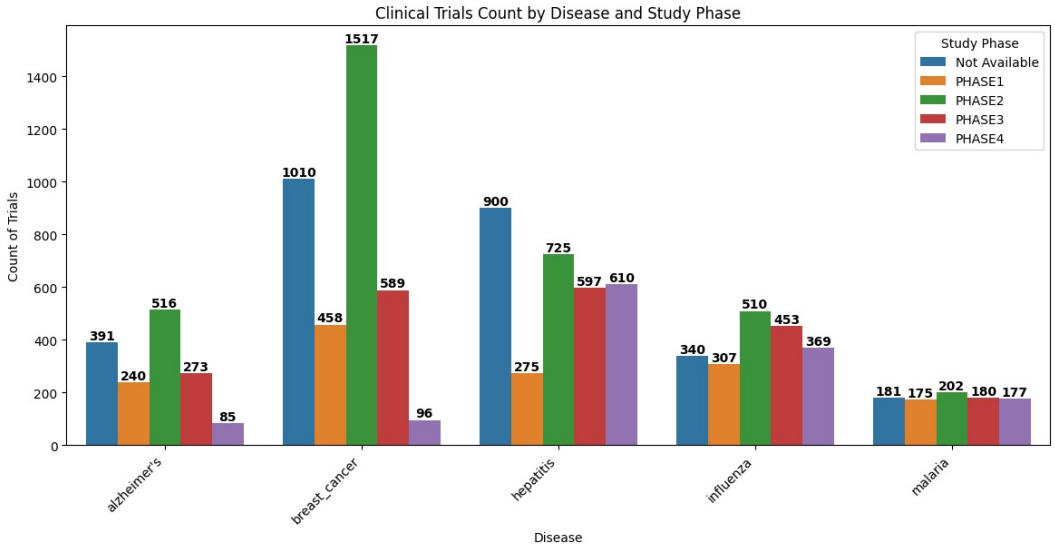


Figure 5: Clinical Trials by Phases for Common Diseases

4.3 RQ_3: How does clinical trial funding compare between rare/high-mortality diseases and more common diseases?

A closer examination of trial sponsorship reveals clear funding disparities between rare and high-mortality diseases versus well-funded conditions. Pancreatic cancer is heavily reliant on industry sponsorship, with 480 industry-funded trials compared to only 85 NIH-funded trials, indicating that private investment plays a dominant role in this disease's research landscape. Similarly, DMD receives substantial industry backing, with 163 industry-funded trials versus just 6 NIH-funded trials, reflecting the high commercial interest in genetic and pediatric diseases with established drug markets.

In contrast, Chagas disease and drug-resistant tuberculosis rely predominantly on public and university funding, with extremely low industry participation. Chagas disease, for instance, has only 13 industry-sponsored trials, while drug-resistant tuberculosis has 15, despite being a major global health burden. This suggests that diseases with high mortality but lower commercial appeal suffer from slower research progress due to limited private investment. Endometriosis presents a unique funding pattern, where 427 trials are classified as "OTHER" funded, likely indicating a reliance on university or non-profit funding rather than direct government or industry support. This fragmented funding structure could be a contributing factor to the slower pace of clinical advancements for endometriosis compared to diseases with stronger pharmaceutical backing.

In contrast, well-funded diseases exhibit a much stronger industry presence. Breast cancer has 947 industry-sponsored trials, significantly higher than pancreatic cancer, while Alzheimer's has 848 industry-funded trials. Hepatitis, a very common disease, leads with 1,494 industry-sponsored trials, nearly triple that of pancreatic cancer.

Publication trends from PubMed further highlight these disparities. Pancreatic cancer and DMD research are frequently published in high-impact medical and oncology journals, whereas Chagas disease research is mostly confined to specialized tropical disease publications. This reflects the funding gap between commercially viable diseases and neglected conditions that rely primarily on government and non-profit research funding.

The duration of published research provides an additional perspective on funding and research prioritization. The elapsed days between the first and latest publication serve as a proxy for how long a disease has been actively researched.

Rare diseases such as Chagas disease, drug-resistant tuberculosis, and Duchenne muscular dystrophy have a relatively shorter research span compared to more common, well-funded diseases. The first

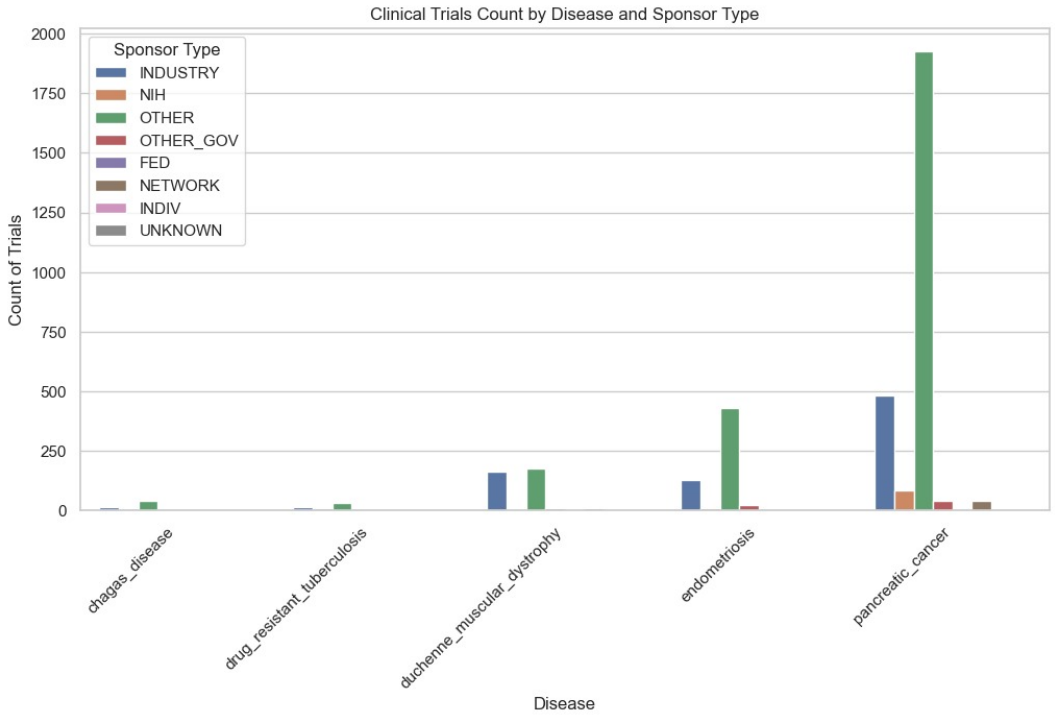


Figure 6: Type of Trial Sponsors per Overlooked Diseases

published studies on Chagas disease appeared around 1969, while pancreatic cancer research spans over 21,439 days. However, these durations are still shorter than those observed for better-funded diseases.

By comparison, common diseases such as influenza, hepatitis, and malaria have the longest research histories, spanning up to 28,929 days (Malaria) and 26,195 days (Influenza). These diseases have established global research networks, consistent funding sources, and industry involvement that sustain clinical trials over long periods.

The difference in elapsed research days may suggest that diseases with a longer publication history tend to attract more funding and have a more developed research infrastructure. This supports earlier observations that breast cancer, Alzheimer's, and hepatitis have significantly more industry-sponsored trials and successful clinical trial transitions compared to rare diseases.

These findings emphasize the importance of sustained research funding and publication exposure in ensuring successful clinical trial progression. Rare diseases, despite their severity, suffer from shorter research histories, lower industry participation, and fewer late-stage trials, reinforcing the need for increased investment and policy intervention.

RQ_3 Findings: Industry funding dominates well-funded diseases like breast cancer (947 trials) and Alzheimer's (848 trials), while neglected diseases like Chagas (13 trials) and drug-resistant tuberculosis (15 trials) rely on public and university funding. This disparity highlights how commercial appeal drives research investment, which leaves high-mortality but less profitable diseases underfunded and under-researched.

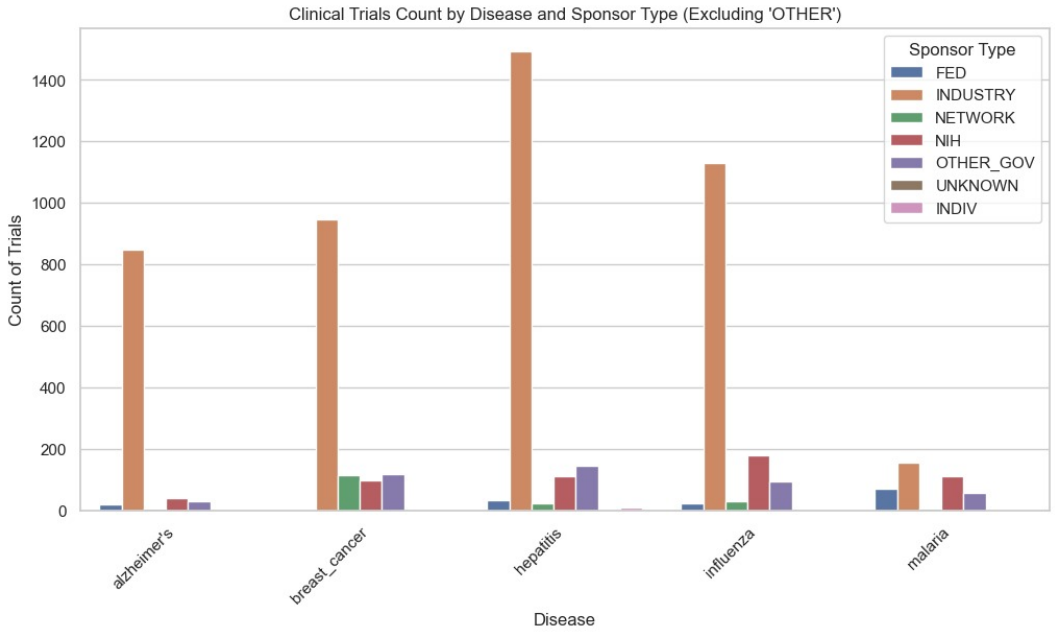


Figure 7: Type of Trial Sponsors per Common Diseases

4.4 RQ$_4$: Are certain diseases more likely to have industry-sponsored vs. government-sponsored trials, and does this impact success rates?

The distribution of trial sponsorship across diseases indicates that industry-backed diseases tend to progress through clinical trials more efficiently than those dependent on public funding. Pancreatic cancer and DMD are largely industry-driven, with hundreds of trials backed by private companies, suggesting a strong commercial pipeline that accelerates drug development. The high level of industry involvement in these diseases may explain why DMD exhibits relatively higher Phase 2 to Phase 3 success rates compared to neglected diseases. In contrast, Chagas disease and drug-resistant tuberculosis rely primarily on NIH and government funding, with minimal industry engagement. The lack of private investment in these diseases may contribute to their high failure rates and slow progression through clinical trials, as public funding alone may not be sufficient to carry treatments through the expensive late-stage trial phases. Endometriosis follows a mixed funding model, with a significant number of university and non-profit-backed trials but fewer direct industry or NIH investments. This may lead to delays in securing funding for large-scale Phase 3 trials, which could be a limiting factor in its clinical research advancement. These findings reinforce the importance of industry participation in clinical trials and suggest that diseases dependent solely on public funding may struggle to reach late-stage drug approvals.

The analysis confirms that diseases with strong industry investment are more likely to progress successfully through clinical trials. Breast cancer and hepatitis, which have the highest industry sponsorship, also show higher transition rates from Phase 2 to Phase 3. Among the target diseases Pancreatic cancer and DMD, which have the highest industry sponsorship, also show stronger success rates in transitioning from Phase 2 to Phase 3. In contrast, Chagas disease and drug-resistant tuberculosis, which rely primarily on NIH and government grants, struggle to reach late-stage trials.

Alzheimer's and breast cancer both have strong industry engagement and a balanced mix of public and private funding, which likely contributes to their clinical trial efficiency. Meanwhile, malaria, which receives more NIH support (110 NIH-funded trials), performs better than Chagas disease, highlighting the impact of sustained public investment.

A major implication of this finding is that publicly funded diseases may face longer development timelines due to a lack of commercial incentives. Diseases with strong industry participation benefit

from streamlined drug approval pathways and higher research visibility. The PubMed dataset supports this observation, showing that diseases with strong pharmaceutical engagement receive more high-impact journal coverage, which in turn attracts further investment and clinical development.

RQ$_4$ Findings: Industry-backed diseases progress more efficiently, while publicly funded ones, like Chagas disease and drug-resistant tuberculosis, face delays. Private investment accelerates trial success, whereas public funding struggles with higher attrition.

Figure 8: Trials across the world

4.5 Integration of PubMed and ClinicalTrials.gov Data

The integration of publication data with clinical trial statistics highlights a clear trend: diseases with higher industry funding and greater research exposure tend to have better clinical trial success rates. The number of published research articles correlates with the number of active clinical trials and progression through trial phases, suggesting that well-funded diseases benefit from both research visibility and industry sponsorship.

Key findings include:

- **Breast cancer and Alzheimer's disease** have significantly more active clinical trials and stronger Phase 2 to Phase 3 progression rates compared to pancreatic cancer and Chagas disease. Breast cancer has 797 actively recruiting trials, while pancreatic cancer has 486. Similarly, Alzheimer's has 356 recruiting trials, significantly more than Chagas disease (only 3 recruiting trials). These diseases also have a stronger presence in high-impact medical journals, further increasing research visibility.

- **Chagas disease, with only 5 active trials and 680 research articles, remains severely underfunded and under-researched**, despite its global burden. By comparison, malaria has 51 active trials and 6,855 research articles, indicating a significantly higher research priority. Both diseases affect millions globally, yet malaria receives stronger funding and publication attention, contributing to its relatively higher number of late-stage trials (180 in Phase 3 vs. 9 for Chagas disease).

- Endometriosis, despite having 2,839 research articles and spanning over 22,000 days of study, still faces challenges in late-stage trial progression. While breast cancer has 1,517 trials in Phase 2 and 589 in Phase 3, endometriosis only has 84 trials in Phase 2 and 58 in Phase 3. This reflects a gender-based research disparity where female-specific conditions receive lower funding despite their widespread prevalence.

- Hepatitis and breast cancer have the highest industry-sponsored trial counts, reinforcing the role of private funding in driving clinical success. Hepatitis has 1,494 industry-funded trials, while breast cancer has 947, both significantly higher than pancreatic cancer (480 industry-funded trials) and Chagas disease (only 13). This pattern is reflected in the research output: Hepatitis has 7,087 published articles, while pancreatic cancer, despite its high mortality, has 9,831 articles but lags in industry-sponsored trials.

- Diseases with stronger publication volume in high-impact journals also show better clinical trial outcomes. - Influenza research spans over 26,195 days and has 8,905 articles, indicating sustained long-term research investment. - Breast cancer research has 17,245 elapsed study days and appears frequently in high-impact journals like *Journal of Clinical Oncology* and *The Lancet Oncology*. - Chagas disease research, despite its severe burden, appears mostly in niche journals like *PLoS Neglected Tropical Diseases* rather than widely read general medical journals. - Malaria, which has 6,855 published articles, frequently appears in journals like *The Lancet* and *Clinical Infectious Diseases*, demonstrating stronger research visibility than Chagas disease.

These findings reinforce the hypothesis that higher research visibility, strong industry sponsorship, and extensive publication in high-impact journals correlate with better clinical trial success rates and faster drug development.

Diseases like breast cancer, Alzheimer's, and hepatitis benefit from strong industry participation and broad research dissemination, leading to better clinical outcomes. Meanwhile, Chagas disease and drug-resistant tuberculosis struggle with limited industry sponsorship and lower research visibility, which significantly slows drug development progress.

5 Discussion & Conclusion

This study highlights significant disparities in clinical trial activity, phase progression, and funding allocation for high-burden yet underfunded diseases. Compared to well-funded conditions like breast cancer and Alzheimer's disease, diseases such as Chagas disease, drug-resistant tuberculosis, and endometriosis have fewer active trials and lower phase transition rates, limiting treatment advancements.

The analysis reveals that industry-backed diseases progress more efficiently through clinical phases, while those reliant on public funding face higher attrition rates. Geographic disparities further restrict research efforts, particularly in low-income regions. Additionally, publication trends correlate with trial activity, as diseases with greater research visibility attract more funding and clinical studies.

These findings underscore the influence of commercial incentives on research prioritization, often at the expense of diseases with high morbidity but low market potential. Addressing these imbalances is crucial for ensuring equitable research investment and accelerating drug development for neglected diseases.

Code & Data Availability Statement

The project's codebase is available on GitHub at this link: https://github.com/Nafiz43/STA-220-Project-Clinical-Trial-Analysis. While the raw data of this study is available in the following drive link: G-Drive

References

[1] S. Piantadosi, *Clinical trials: a methodologic perspective*. John Wiley & Sons, 2024.

[2] J. Fry, *Common diseases: their nature incidence and care*. Springer Science & Business Media, 2012.

[3] C. A. Lane, J. Hardy, and J. M. Schott, "Alzheimer's disease," *European journal of neurology*, vol. 25, no. 1, pp. 59–70, 2018.

[4] J. De Groote, V. Desmet, P. Gedigk, G. Korb, H. Popper, H. Poulsen, P. Scheuer, M. Schmid, H. Thaler, E. Uehlinger *et al.*, "A classification of chronic hepatitis," *The Lancet*, vol. 292, no. 7568, pp. 626–628, 1968.

[5] E. D. Kilbourne, *Influenza*. Springer Science & Business Media, 2012.

[6] R. Tuteja, "Malaria- an overview," *The FEBS journal*, vol. 274, no. 18, pp. 4670–4679, 2007.

[7] T. J. Key, P. K. Verkasalo, and E. Banks, "Epidemiology of breast cancer," *The lancet oncology*, vol. 2, no. 3, pp. 133–140, 2001.

[8] M. E. Fernandez-Zapico, J. A. Kaczynski, and R. Urrutia, "Pancreatic cancer research: challenges, opportunities, and recent developments," *Current opinion in gastroenterology*, vol. 18, no. 5, pp. 563–567, 2002.

[9] R. G. Postier, "The challenge of pancreatic cancer," *The American Journal of Surgery*, vol. 186, no. 6, pp. 579–582, 2003.

[10] A. S. de Sousa, D. Vermeij, A. N. Ramos, and A. O. Luquetti, "Chagas disease," *The Lancet*, vol. 403, no. 10422, pp. 203–218, 2024.

[11] M. C. P. Nunes, W. Dones, C. A. Morillo, J. J. Encina, A. L. Ribeiro, and C. on Chagas Disease of the Interamerican Society of Cardiology, "Chagas disease: an overview of clinical and epidemiological aspects," *Journal of the American College of Cardiology*, vol. 62, no. 9, pp. 767–776, 2013.

[12] D. L. Olive and E. A. Pritts, "Treatment of endometriosis," *New England Journal of Medicine*, vol. 345, no. 4, pp. 266–275, 2001.

[13] B. Eskenazi and M. L. Warner, "Epidemiology of endometriosis," *Obstetrics and gynecology clinics of North America*, vol. 24, no. 2, pp. 235–258, 1997.

[14] C.-Y. CHIANG, R. Centis, and G. B. Migliori, "Drug-resistant tuberculosis: Past, present, future," *Respirology*, vol. 15, no. 3, pp. 413–432, 2010.

[15] J. L. Khawbung, D. Nath, and S. Chakraborty, "Drug resistant tuberculosis: A review," *Comparative immunology, microbiology and infectious diseases*, vol. 74, p. 101574, 2021.

[16] D. Duan, N. Goemans, S. Takeda, E. Mercuri, and A. Aartsma-Rus, "Duchenne muscular dystrophy," *Nature reviews disease primers*, vol. 7, no. 1, p. 13, 2021.

[17] I. E. Verhaart and A. Aartsma-Rus, "Therapeutic developments for duchenne muscular dystrophy," *Nature Reviews Neurology*, vol. 15, no. 7, pp. 373–386, 2019.

[18] ClinicalTrials.gov, "Clinicaltrials.gov api v2 documentation," 2025, accessed: 2025-03-16. [Online]. Available: https://clinicaltrials.gov/api/v2/studies

[19] N. C. for Biotechnology Information (NCBI), "Ncbi datasets api documentation," 2025, accessed: 2025-03-16. [Online]. Available: https://www.ncbi.nlm.nih.gov/datasets/docs/v2/api/