Investigating Trends in Neglected Tropical Disease (NTD) Clinical Studies

1. Project information

Project Name: Investigating Trends in Neglected Tropical Disease (NTD) Clinical Studies

Project owner/sponsor: Infectious Diseases Data Observatory, University of Oxford

Organisation name: Infectious Diseases Data Observatory, University of Oxford

Main contact name: Rhys Peploe

Main contact position: Statistician

Team Members: 3 Students

Peizhe Jiang

Congyao Ren

Zixu Wang 36989093

2. Executive Summary

Neglected Tropical Diseases (NTDs) afflict more than one billion individuals globally, exerting a disproportionate burden on socioeconomically disadvantaged populations. Nevertheless, research pertaining to NTDs remains significantly underfunded relative to other global health priorities. This project synthesizes data from 315 studies

registered in the World Health Organization International Clinical Trials Registry Platform (ICTRP), encompassing Chagas disease, schistosomiasis, soil-transmitted helminthiases, and visceral leishmaniasis, published between 1999 and 2023.

The objective of this study is to elucidate patterns, disparities, and salient insights within NTD clinical research. The findings are expected to inform evidence-based policy formulation, optimize resource allocation, and guide the prioritization of future research within the global health community.

3. Data Source

The data for this project originates from the World Health Organization's (WHO) International Clinical Trials Registry Platform (ICTRP), which collates trial registrations from multiple international and national registries. The dataset encompasses all available years of Neglected Tropical Disease (NTD) clinical trial registrations between 1999–2023, covering every record retrievable from the ICTRP search portal.

To improve consistency, the dataset has been curated and partially standardised by the Infectious Diseases Data Observatory (IDDO), while retaining some raw variables where full standardisation was not feasible. The classification of NTDs in this study follows the definitions provided by WHO

This comprehensive dataset provides a reliable global perspective on NTD research activities, enabling robust analysis of long-term trends across all recorded years.

4. Research Questions

4.1. What factors (e.g., study phase, funder type, geographic region, population inclusion) are associated with whether results are published?

4.2. Can a network analysis reveal frequent partnerships and identify key hub countries?

4.3. What proportion of trials are funded by pharmaceutical companies, and do these align with regions of high disease burden?

4.4. Are children and pregnant women included or excluded from studies, and how does this vary by disease or study phase?

4.5. For one selected disease, which drugs are being studied, and how have these trends evolved?

5. Methodology

5.1 Data Processing

Data Cleaning: Handle missing values, remove duplicates, standardize categorical variables.

Feature Selection: Extract year, disease category, country, sponsor type, and participant demographics, and use statistical correlation and domain knowledge to retain the most relevant variables.

Results Analysis: Logistic regression and tree-based models to identify predictors of result publication.

5.2 Tools and Technologies

Programming: Python (pandas, NumPy, Sklearn)

Visualization: matplotlib

Collaboration: GitHub, Google Docs

6. Team Structure and Responsibilities

Peizhe Jiang:

Congyao Ren:

Zixu Wang:

7. Project Timeline

Phase	Tasks	Duration
Week 1-2	Analyze group projects and select, research literature	2 Weeks
Week 3-4	Team coordination and data collection	2 Weeks
Week 5	Draft project plan and initial data cleaning.	1 Weeks

Week 6-7	Data preprocessing and creation of classification tables.	2 Weeks
Week 8	Clustering, visualization, and analysis	1 Week
Week 9	Project review , validation of findings and preparation final report	1 Week
Week 10	Final report submission and presentation preparation	1 Week