Investigating Trends in Neglected Tropical Disease (NTD) Clinical Studies

Group ID:16

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

Project Host: 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

Document Type: Project proposal

Team Members:
Peizhe **J**iang
Lancaster University, p.jiang6@lancaster.ac.uk

Congyao **R**en
Lancaster University, c.ren1@lancaster.ac.uk
Zixu **W**ang
Lancaster University, z.wang120@lancaster.ac.uk

Alaghwani **B**alsam
Lancaster University, b.alaghwani@lancaster.ac.uk

1. Introduction

Neglected tropical diseases (NTDs) affect more than one billion people worldwide [1], causing a particularly heavy and disproportionate health burden on socio-economically disadvantaged groups, and further exacerbating the cycle of poverty and the lack of medical resources. Despite its significant impact on public health, compared with other major disease areas, NTDs-related research has been in a state of underfunding and low attention for a long time [2], resulting in the persistence of problems such as weak evidence base and uneven research layout.

This project is based on 315 research data registered by the World Health Organisation International Clinical Trial Registration Platform (WHO ICTRP) from 1999 to 2023, covering four main types of NTDs, including Chagas disease, schistosomiasis, endemic worm disease and visceral leishman disease. The research aims to sort out the changing trends and differences in clinical trials of these diseases in terms of time, geography and funding structure, and reveal the representative patterns and characteristics. Relevant findings are expected to provide strong data support and practical reference for evidence-based policy formulation, resource optimisation and future research direction formulation.

2. Aim And Research Questions

2.1 Aim

This project aims to conduct a comprehensive analysis of global clinical research related to neglected tropical diseases (NTDs) from 1999 to 2023. Through data processing and observation, it seeks to clarify the evolution of NTD clinical trial activities during this period, understand their changing trends and key characteristics, and provide data support and directional guidance for public health departments and related institutions in reducing disease prevalence, optimizing resource allocation, and improving the cost-effectiveness of prevention and intervention.

2.2 Research questions

2.2.1 *What factors are associated with whether results are published?*

2.2.2 *Can a network analysis reveal frequent partnerships and identify key hub countries?*

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

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

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

3. Proposed Approach

3.1 Research plan

3.1.1 Data collection and organisation:

The data set used in this project is provided by the Infectious Diseases Data Observatory (IDDO) of Oxford University.

The data comes from the International Clinical Trials Registry Platform (ICTRP) of the World Health Organisation (WHO), and has been partially standardised by IDDO, covering clinical research records related to neglected tropical diseases (NTD) from 1999 to 2023.

On this basis, the project team will clean and organise the data for the second time.

3.1.2 Data cleaning and preprocessing:

Delete duplicate and missing records; Unify classification variables such as date, disease name, type of funder, etc.; Classify sponsors according to keyword(Non-profit, Government, Industry)[3] rules. Classify countries(LOW-INCOME, LOWER-MIDDLE, UPPER-MIDDLE ,HIGH)through the World Bank[4]

3.1.3 Exploratory data analysis:

Analyse the annual trend of change, regional distribution and research stage characteristics; The proportion of public publication of statistical results and the structure of funding; Identify special groups (children, pregnant women) in the situation.

3.1.4 Modelling and Statistical Analysis

This project will use a logistic regression model to analyse the main factors (such as research stage, funding type, region and research population) that affect the publication of research results, and use network-centred analysis to identify cooperation models and core countries between countries. As emphasised by [5], excessive model complexity or unrealistic data demands can undermine decision relevance. Our analysis attempts to balance parsimony and realism to ensure interpretability and policy applicability.

3.1.5 Result visualisation and report writing:

Use Python to generate time series diagrams, pie charts and network relationship diagrams; Organise the main findings, write research reports and prepare for the final presentation.

3.2 Key Methods and Tools

3.2.1 Key Methods:

Logical regression analysis: used to identify which factors affect whether the research results are public, such as the research stage, the type of funder, the national income level and the inclusion of special groups.
Network analysis: By building a cooperation network between countries, it reveals the cooperation structure, common collaboration models and the central countries with key influence.
Classification method based on keywords: used to sort out non-standard fields, such as national income type, sponsor type, drug type, etc. The relevant fields are standardised through keyword rules, so that there is consistency between different records and convenient for statistical analysis.
Data cleaning, exploratory analysis and visualisation: complete data de-weighting, variable unification, trend analysis and chart presentation in the Python environment to ensure the accuracy and interpretability of the analysis results.

3.2.2 Main tools:

Programming Language: Python (pandas, NumPy, matplotlib, sklearn, networkx)
Programming Environment: VSCODE
Collaboration platform: GitHub, Google Docs

4. Proposed Timeline

Phase	Tasks	Duration
Week 1-2	Analyse 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, visualisation, and analysis	1 Week
Week 9	Project review, validation of findings and preparation of the final report	1 Week
Week 10	Final report submission and presentation preparation	1 Week

5. Reference

[1] World Health Organisation. 2025. Global Report on Neglected Tropical Diseases 2025. Retrieved from https://www.who.int/teams/control-of-neglected-tropical-diseases/global-report-on-neglected-tropical-diseases-2025

[2] Nick Feasey, Mark Wansbrough-Jones, David C. W. Mabey, and Anthony W. Solomon. 2010. Neglected tropical diseases. British Medical Bulletin, 93(1), 179–200. https://doi.org/10.1093/bmb/ldp046

[3] Florence T. Bourgeois, S. Murthy, and Kenneth D. Mandl. 2010. Outcome reporting among drug trials registered in ClinicalTrials.gov. Annals of Internal Medicine, 153(3), 158–166. https://www.acpjournals.org/doi/10.7326/0003-4819-153-3-201008030-00006

[4] World Bank. 2025. Country and Lending Groups – World Bank Data Help Desk. Retrieved from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519

[5] Bruce Y. Lee and Sarah M. Bartsch. 2017. How to determine if a model is right for neglected tropical disease decision making. PLoS Neglected Tropical Diseases, 11(4): e0005457. https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0005457

6. Team Setup

Group ID: 16

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

Name	StudentID	E-mail	Roles	Responsibilities
Peizhe Jiang	37012523	p.jiang6@lancaster.ac.uk	Data Scientist	Perform analysis, build models and extract insights
Congyao Ren	37002623	c.ren1@lancaster.ac.uk	Data Analyst	Explore data, visualisations and analytics reporting.
Zixu Wang	36989093	z.wang120@lancaster.ac.uk	Data Engineer	Responsible for data cleaning, sponsor classification, data visualisation and basic statistical analysis
Alaghwani Balsam	37060704	b.alaghwani@lancaster.ac.uk	Data Analyst	Generate visualisations, assist with exploratory analysis, and contribute figures for reporting