MTH208: Term Project Proposal
Group 24
Global Tuberculosis Analysis

Team Members: Naina Bhalla (240674) Puspak Kumar Laha (251080084)
Ananya Aggarwal (251080056) Ravi Kumar (251080085)

Problem Statement

Tuberculosis (TB) is caused by bacteria (Mycobacterium tuberculosis) and it most often affects the lungs. Despite multiple national programs and the WHO's *End TB Strategy*, India has the highest burden of TB with two deaths occurring every three minutes from it. Social factors such as poverty, undernutrition, literacy, and rapid urbanization or lack thereof significantly influence TB prevalence. This project aims to analyze TB incidence and mortality trends, explore socioeconomic correlations, and assess the extent of reporting disruptions. Over and above, it aims to develop an interactive, data-driven dashboard that integrates WHO and World Bank data to provide insights into global trends and determine factors affecting case-scenarios related to TB.

Objectives

- To track global and country-specific TB incidence, mortality, and notification trends.
- To explore correlations of TB burden with various socioeconomic indicators (GDP, poverty, health expenditure, urbanization).
- To analyze the scale and trend of multidrug-resistant TB (MDR-TB) and rifampicin-resistant TB (RR-TB).
- To quantify the impact of COVID-19 on TB notifications and outcomes.
- To provide a user-friendly, data-driven, dashboard-app (Shiny app) that can be used and extended for further research purposes.

Planned Data Sources

- **WHO TB Statistics:** Publicly available datasets from the Global TB Programme Data Portal.
- **World Bank World Development Indicators (2000-2023):** GDP per capita, poverty rates, health expenditure, urbanization, literacy, population.
- **Supplementary health indicators:** HIV prevalence (UNAIDS), diabetes prevalence (IDF), and nutrition (FAO, NFHS).

The datasets/websites are open access and ethically usable for academic research.

Proposed Research Questions

1. How has TB incidence and mortality changed in India over the past two decades? Did notifications decline significantly during COVID-19 (2020–2021)?
2. Which social factors (literacy, poverty, urbanization, nutrition) show the strongest correlation with TB prevalence across regions/countries?
3. Which regions bear the highest MDR-TB and RR-TB burden?
4. What are the trends in multi-drug resistant TB (MDR-TB) and how do treatment success rates relate to mortality reduction?

Proposed Methodology

- **Data Acquisition:** Download WHO TB CSV datasets; scrape/merge socio-economic indicators from World Bank and other sources using `tidyverse` and `dplyr`.
- **Cleaning & Integration:** Standardize country names using `countrycode` package, handle missing values
- **Exploratory Data Analysis (EDA):**
  - **Trend Plots:** Line charts of TB incidence, notifications, and mortality over time (2000–2023). Separate plots for new and retreatment cases to capture differences.
  - **MDR-TB Trends:**
    * Percentage of new TB cases that are MDR-TB.
    * Percentage of retreatment TB cases that are MDR-TB.
    * Total number of estimated MDR/RR-TB cases.
  - **Socioeconomic Correlation Plots:** Scatter plots of TB incidence against socioeconomic indicators with regression lines.
  - **COVID-19 Disruption Plots:** TB notifications and incidence before, during, and after the COVID-19 period (2020–2021), highlighting drops in reporting.
- **Shiny App:**
  - Interactive plots of TB trends across years and countries.
  - Filters for socioeconomic indicators (poverty, literacy, urbanization).
  - An "About" tab explaining data sources and methodology.
  - A "Home" tab answering the research questions along with global trends.
  - A "Glossary" tab containing definitions of some medical terms.

Tentative Team Roles

- **Data Acquisition & Cleaning:** Puspak Kumar Laha
- **EDA & Shiny App Development:** Naina Bhalla
- **Analysis:** Shared responsibility based on collective analysis
- **Report Writing :** Ananya Aggarwal
- **Presentation :** Ravi Kumar

Feasibility & Ethics

All data sources are open access, published by the WHO, World Bank, and UN agencies, and thus carry no privacy concerns. The scope is feasible within deadlines provided.

References

- WHO Global TB Programme Data
- World Bank World Development Indicators and WDI library
- Global Tuberculosis Report 2024
- The Social Determinants of Tuberculosis: From Evidence to Action
  Hargreaves JR, Boccia D, Evans CA, Adato M, Petticrew M, Porter JD. The social determinants of tuberculosis: from evidence to action. Am J Public Health. 2011 Apr;101(4):654-62. doi: 10.2105/AJPH.2010.199505. Epub 2011 Feb 17. PMID: 21330583; PMCID: PMC3052350.
- Sociodemographic factors affecting knowledge levels of tuberculosis patients in New Delhi
  Zaidi I, Sarma PS, Khayyam KU, Toufique Ahmad Q, Ramankutty V, Singh G. Sociodemographic factors affecting knowledge levels of tuberculosis patients in New Delhi. J Family Med Prim Care. 2024 Nov;13(11):5152-5158. doi: 10.4103/jfmpc.jfmpc_387_24. Epub 2024 Nov 18. PMID: 39722901; PMCID: PMC11668486.