

## Tuberculosis Burden by Country

### Jiaxiang Li

#### Original Dataset:

- *Tuberculosis Burden by Country* - The World Health Organization estimates the prevalence and mortality of Tuberculosis by country, Tableau public.  
[https://public.tableau.com/s/resources?qt-overview\\_resources=1](https://public.tableau.com/s/resources?qt-overview_resources=1)

#### Additional Data Sources:

- *Treatment coverage Data by country*, data from Global Health Observatory data repository (GHO). <https://apps.who.int/gho/data/node.main.1323?lang=en>
- TB incidence by Age Sex and Risk factor, data from WHO Global Tuberculosis Programme. <https://www.who.int/teams/global-tuberculosis-programme/data>

The original source dataset, which is provided by Tableau's sample data page, is credible for further data analysis and visualization, because it is collected and estimated by the World Health Organization (WHO).

The treatment coverage dataset is also collected and curated by WHO and is further exhibited in Global Health Observatory data repository (GHO) for any deepening in scientific research. GHO was founded solely for public to use authoritative health-related statistics collected by WHO<sup>[1]</sup>. All datasets in GHO are publicly released and free.

The third dataset is collected by the WHO Global Tuberculosis Programme, an authoritative team led by Dr. Tereza Kasaeva. It aims to achieve the Sustainable Development Goal 3, the End TB Strategy. The team provides global leadership programs, shape TB research agenda, gives technical support, and monitor the status of the TB epidemic.

The last two datasets are all created in less than two years ago, which ensures the timeliness of the information. They both have clear purpose, being supportive toward the original dataset of tuberculosis burden. Both investigate and collect relevant data of tuberculosis from all countries.

With the first dataset researching about TB burden, incidence, and mortality, the second one investigates specific treatment coverage to support data from the first set and reiterate the burden. The third dataset further included age, sex, and risk factor of different countries, adding more detailed descriptions of the status quo for global TB infection and treatment. All of them are set up by WHO and its affiliated groups, which ensures their validity.

The datasets consist of both numerical (year, number of cases, mortality, etc.) and categorical (age group, gender, risk factor, etc.). They are structured and primary, with types of integers (number of incidences, death, etc.), characters/strings (gender: m/f/a, country, income level, etc.), and void (when data in specific region is not available). In this scenario, the independent variables are countries/regions while the dependent variables are:

- Original dataset: Years. Numbers of total populations and estimated prevalence/mortality of TB in different circumstances and in low/high bounds. Methods to derive incidence estimates.

- Second dataset: Years. Percentage of treatment coverage. Numbers of TB cases and new/relapse cases.
- Third dataset: Years. Age group. Sex. Risk factor. Number of cases of low/high bound.

For the Tableau data I've chosen. The intended audience is most likely to be country/region government and their public health organizations of all size and form.

According to the data solely in the first dataset, we can make basic assumptions about population/TB case ratio, how HIV affects TB incidences, and how high case detection rates are in different regions.

For the ratio, the higher the ratio is in a region, the more prevalent and widespread TB is in that place. But if the percentage of HIV is relatively high in TB cases, the assumption made before about TB in one region should be re-considered.

The basic information derived from the dataset can show how comprehensive and well-thought-out the research about the topic is. In this case, the data WHO provided is comprehensive and of full scale, with date, region, HIV intervention considered. Although it is still primary source, it is well-organized enough to allow analyst to make some fundamental summaries and comprehension about it.

[\[1\]](#) Vardell E. (2020). Global Health Observatory Data Repository. *Medical Reference Services Quarterly*, Volume 39, 2020 - Issue 1.