# Tuberculosis Burden Country Data Visualization

Christina Lu Jin | March 2021

This is analysis report includes three dashboard each with charts/graphs. The idea was to create more comprehensive visualizations and more in-depth analysis of the TB cases worldwide records during year 1990 to 2014. As well as exploring the relationships between prevalence, incidence, and mortality, also between TB and HIV.

Before diving into the dashboards, first, I would like to clarify the differences between the concept of incidence and prevalence to avoid confusions. Prevalence is the measure of disease in a population, it refers to *the proportion of individuals in a population that have a disease at a particular time*, and is also known as a measure of the disease burden. Different from prevalence, incidence refers to the number of new cases that develop within a given time period.

While researching online, I found the following image which perfectly described the difference between the two and thought it is worthy to look at. *When thought of like a bathtub, incidence is the water being added to the bathtub, prevalence is the contents of the bathtub, and cases that recover or die are the water leaving the bathtub via evaporation or down the drain, as represented in the diagram below* [1]. *Therefore, as new cases occur, disease prevalence values will accordingly rise. However, if individuals with the disease die or recover, the prevalence will accordingly drop.*[1]
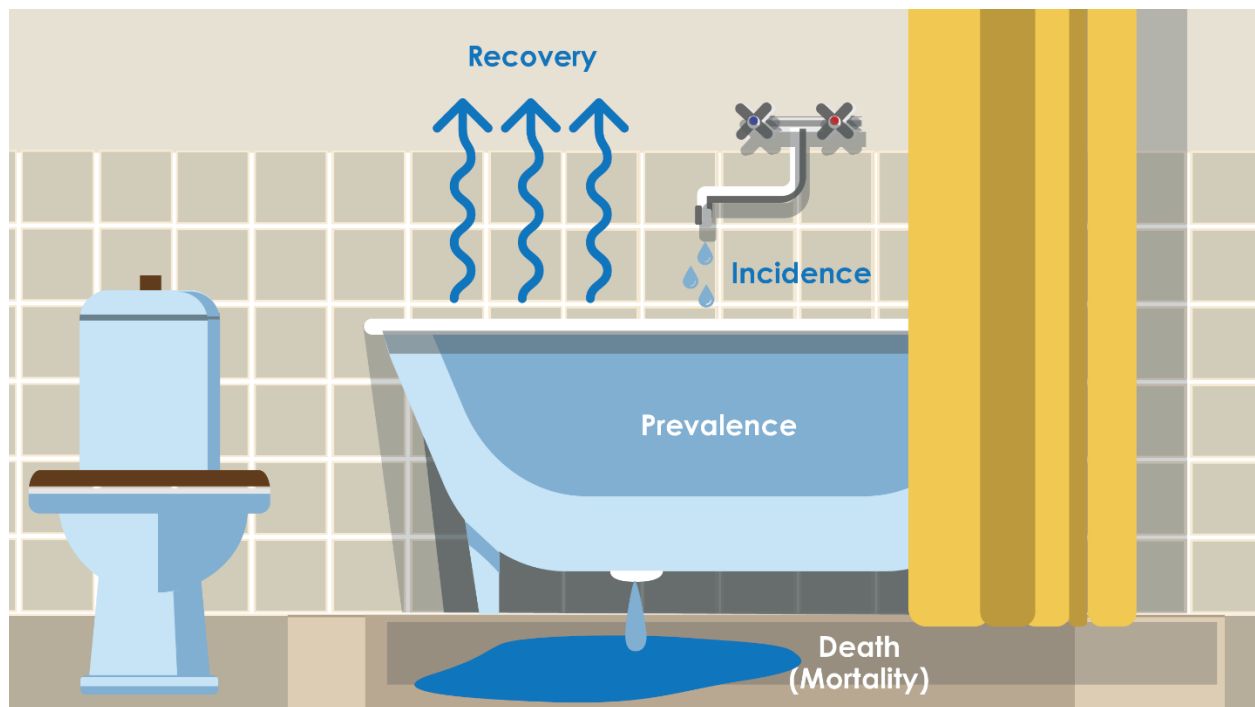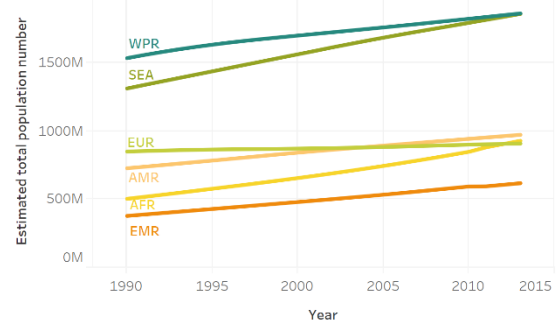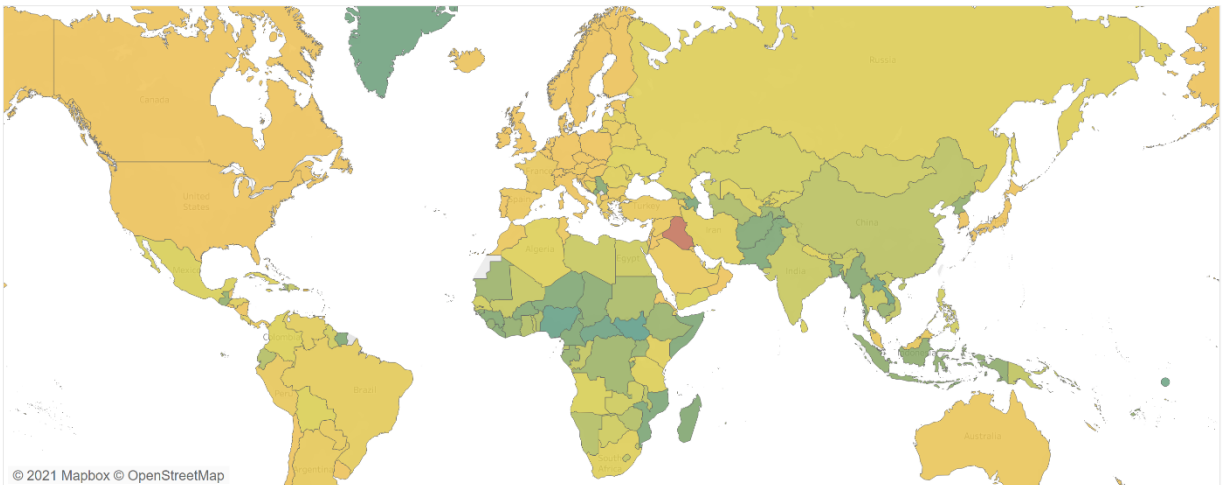
Dashboard 1

Total Population by Country (Top 10)
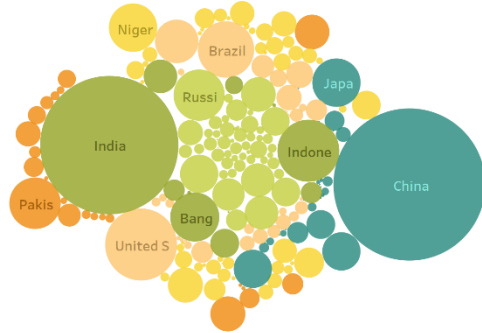
Total Population by Region

TB Case Detection Rate (all form) by Country

This first dashboard displays a general overview of some basic categories of the TB data. The bubble chart on the top left shows the total population situation around the world during 1990 and 2014, ranked by country and color coded by region. Sizes of circles are indicators of the size of the population, it indicates that China and India had tremendous amount of population during that period comparing to rest of the world.

The top right is a line graph shows the population growth within each region. All regions had some degree of growth, European Region (EUR) is the only one that didn't change much over the years. Western Pacific Region (WPR) and South-East Asian (SEA) are observed to have the most population, almost doubled the rest of the regions.

The geo-map, on the bottom, displays the detection rate of all form of TB case of each country. It is the sum of the case detection rate of the entire 24 years, so the added-up percentage values are pretty large in general. Based on the color scale, Iraq (the red area in the middle)
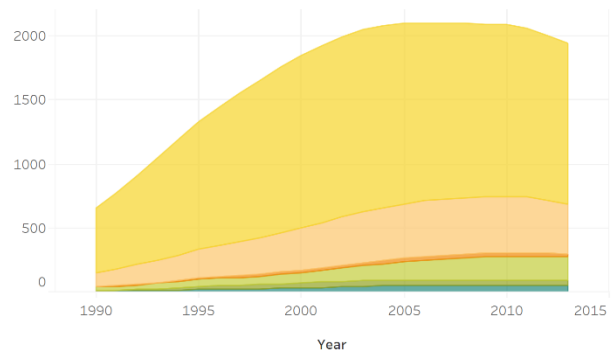
from EMR has the highest case detection rate. And South Sudan from AFR has the least case detection rate among all countries.
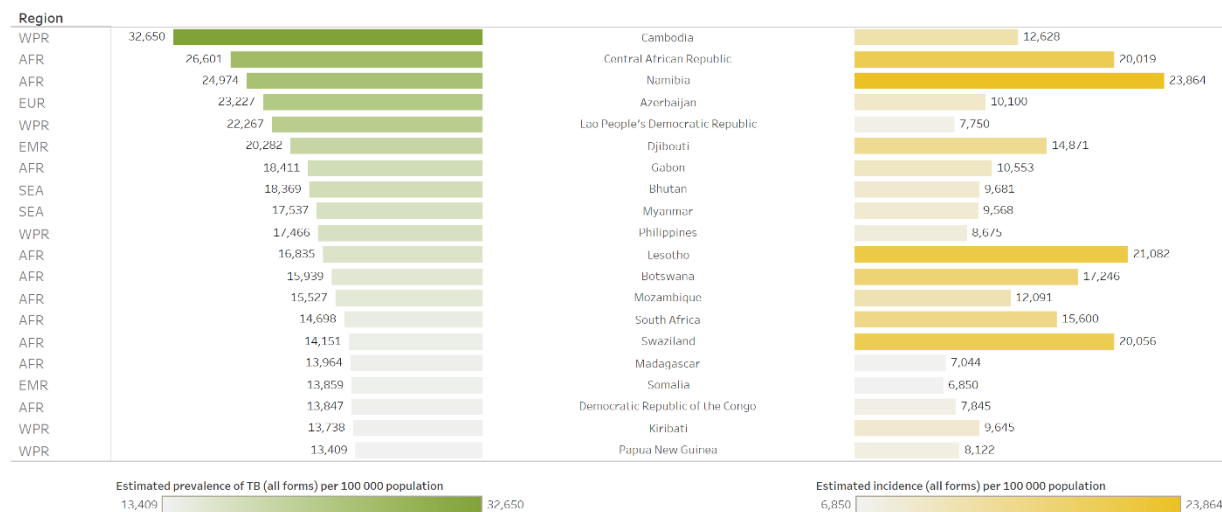
Dashboard 2

Prevalence, Incidence and Mortality Rates
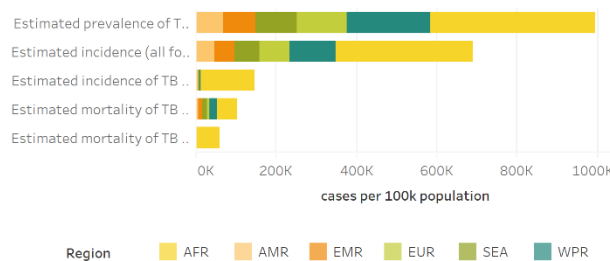
Estimated HIV in incident TB (1990 - 2014)

Estimated Prevalence vs. Incidence of TB per 100k Population (top 20 countries ranked by prevalence)

The bar chart on the top left of the second dashboard shows the rates of all five metrics, starting from the top, they are: prevalence rate, incidence (exclude HIV) rate, incidence (HIV-positive) rate, mortality (exclude HIV) rate, and mortality (HIV-positive) rate. From the bar chart we could tell that, although all regions had decent amount of incidence of TB cases, these new cases are mostly not due to HIV (except African). The mortality rate is quite low in most regions besides African, and barely any death cases are HIV patients. Since AFR has significant HIV infection issues, and the fact that it also has way more TB incidence than other regions suggests that HIV potentially has positive correlation with the chance of getting TB, HIV destroys the immune system and puts individuals at a slightly higher risk at TB infections.
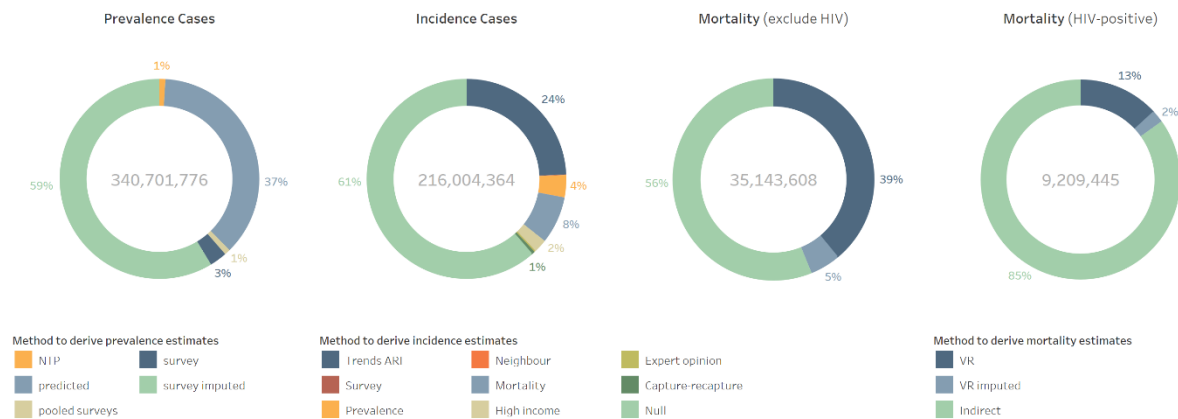
The graph on the top right is an area chart, depicts the changes of HIV in incident TB over that time period, each color represents a region. The HIV-positive cases in AFR region is drastically high, even more than the sum of all other regions together. And it also increased a huge

amount from 1990 to 2005, but it started to slowly decrease from 2005. Thus, we could conclude that HIV was a major issue in AFR regions during those years, by year 2005, the number of new cases (TB patients who are also infected with HIV) almost tripled the amount of 1990. This situation has been taken under-control from 2005, it starts with slower decrease during 2005 and 2010, and after that there are slightly more significant reduction each year. But there is still a huge amount of TB-HIV incidence rate by the end of 2014, therefore, HIV infections is still a negligible issue in AFR. Looking at the rest of the world, Region of Americas (AMR) has slightly higher rates than the other regions, and followed by European Region (EUR), but both regions had got it under control around 2010. HIV was not quite a huge concern in Eastern Mediterranean Region (EMR), South-east Asian (SEA), or Western Pacific Region (WPR).
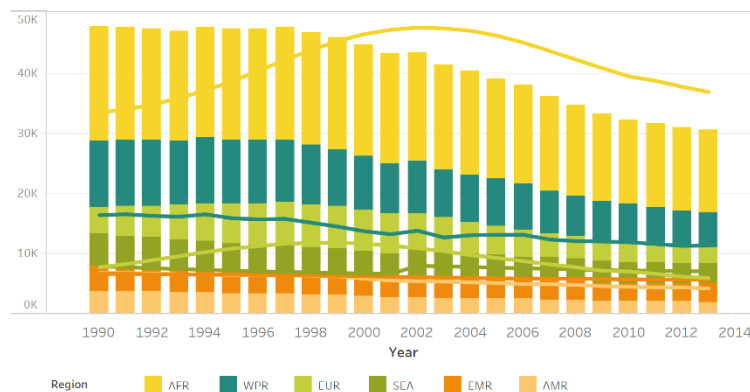
The last graph of this dashboard is a tornado chart, also known as butterfly chart. This tornado chart shows two major metrics side-by-side – prevalence and incidence of each country, it is showing the top 20 countries with highest prevalence. Both metrics are rates per 100k population, therefore, although China and India might have more exiting and new cases, but due to the enormous population in both countries, the rates are not significant. According to the chart, surprisingly there isn't any obvious relationship between prevalence rate and incidence rate. Since the left side of the tornado chart has prevalence rate ranked in a descending order, but the incidence rate on the right side is pretty rando and has no distinct patterns to correspond with the left. Therefore, although TB is a contagious disease, but it doesn't necessarily mean that the more existing cases (prevalence) there are, the more new cases (incidence) there would be. Looking at the region column to the very left, countries that had higher rates are mostly from AFR and WPR regions. Most of the countries that had higher incidence rates belongs to AFR, which corresponds with what the first two charts showed.
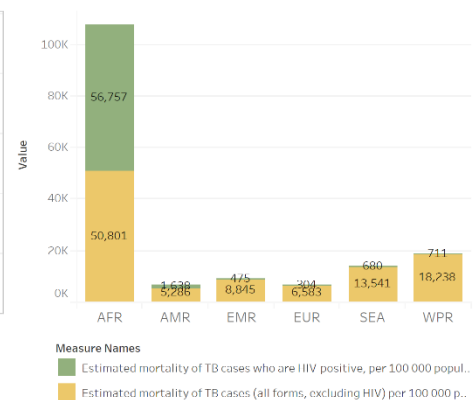
Dashboard 3

Methods for Derive Prevalence, Incidence and Mortality

Prevalence Cases
340,701,776

Incidence Cases
216,004,364

Mortality (exclude HIV)
35,143,608

Mortality (HIV-positive)
9,209,445

Method to derive prevalence estimates
- NIP
- survey
- predicted
- survey imputed
- pooled surveys

Method to derive incidence estimates
- Trends ARI
- Survey
- Prevalence
- Neighbour
- Mortality
- High income
- Expert opinion
- Capture-recapture
- Null

Method to derive mortality estimates
- VR
- VR imputed
- Indirect

Estimated Prevalence of TB Cases by Region
(per 100k population)

Region: AFR, WPR, EUR, SEA, EMR, AMR

Mortality of TB Cases by Region
(HIV-positive vs. HIV-negative)

Measure Names
- Estimated mortality of TB cases who are HIV positive, per 100 000 popul..
- Estimated mortality of TB cases (all forms, excluding HIV) per 100 000 p..

At the top of the third dashboard, there are a series of donut charts which shows the different method that were used to derive each of the metrics, and the number in the center of the rings are the total number of cases of each category. Among all prevalence cases, 59% were derived with survey impute, 37% were derived from predictions. Besides the 61% of null incidence cases, 24% were derived with trend ARI. There were some other derived methods such as from surveys, neighbors, and expert opinions, but they were less than 1%, thus not showing up very well on the chart. Lastly, for mortality, most cases were derived indirectly, including 56% of the cases that excludes HIV, and 85% of the ones that are HIV-positive. They second commonly used method was VR, which covered 39% of the non-HIV cases, and 13% of the HIV-positive cases.

The second chart of this dashboard is a combination chart on the bottom left, it combines a bar chart which displays the prevalence rate, and a line chart which displays the incidence rate, both rates are for per 100k population. From the trend of the bar chart, we could tell that the

prevalence rate for all 6 regions started to decrease from 1997. AMR has the least prevalence rate per 100k population, and EMR comes in second least. Not surprisingly, AFR and WPR are the top two regions with the most prevalence rate. Looking at the curve of the line chart, we can see there was a huge increase in incidence rate of AFR from 1990 to 2002, and it started to slowly changes into decrease in the rate. Similar trend also appears in EUR region, but at a much lower volume. Interestingly, SEA has a decreasing incidence rate from 1990 to 2000 and had a sudden increase in the rate over 2001, then it started to decrease again after 2002. All other regions, even including WPR, maintained a decreasing incidence rate over the years.

Lastly, the stacked bar chart on the bottom right exhibits the comparison between the two types of TB mortality rate - positive HIV and negative HIV. As we would suspect, AFR had the most mortality rate. Since HIV had been a major issue in AFR, there are even more TB deaths cases with positive HIV than negative, none like other regions where there were barely any positive HIV death cases.

References

1) Steward, K. Incidence vs Prevalence (Jan 08,2020). Retrieved from https://www.technologynetworks.com/immunology/articles/incidence-vs-prevalence-329073
2) Arts of Visualization. Tableau Butterfly Chart Tutorial. Retrieved from https://www.youtube.com/watch?v=M74p3N3lnbI
3) Smoak, A. Create Multiple KPI Donut Charts in Tableau. Retrieved from https://www.youtube.com/watch?v=5hMY95vgnlg
4) Smoak, A. Stacked Bar Chart Totals in Tableau. Retrieved from https://www.youtube.com/watch?v=CylyFY0UIj0&t=17s