TB Report

Shabnam Hajian,Dina Haji Zeynaly Biooky

2020-02-26

Data

Tuberculosis (TB) is caused by bacteria (Mycobacterium tuberculosis) that most often affect the lungs. People ill with TB can infect up to 10-15 other people through close contact over the course of a year. Without proper treatment up to two thirds of people ill with TB will die.
Since 2000, 53 million lives have been saved through effective diagnosis and treatment. Active, drug-sensitive TB disease is treated with a standard 6-month course of 4 antimicrobial drugs that are provided with information, supervision and support to the patient by a health worker or trained volunteer. The vast majority of TB cases can be cured when medicines are provided and taken properly.
This Dateset is provided with the information of about **216 countries** around the world and is categorized in **6 different variables**. we will talk about how **TB Incidence** and **TB Mortality** changed from 2000 to 2018. Then, based on the results of these change we will develop some hypothesis about the relations of **TB Mortality** and **Number Of visits** with each other and try to improve calculation opf this association with **TB Incidence**.

TB Incidence and **TB Mortality** are the estimated incidence and the estimated mortality of TB cases in per 100,000 population. Both of these two variables are skewed to the right because not all countries are equaly engaged with **TB**; this disease is more common in some areas of world such as Africa and Eastern Mediterranean, so the higher observed rates (observations in right tail of histograms) are related to these lands.
TB Incident's mean is 125.8 and most parts of the world (75% TB Incidents) have fewer than 200 people (out of each 100,000) with TB in each year; however in some countries we have reports up to 1280 people.**TB Mortality**'s numbers are even fewer than those of incidence: mean=27.05, Q3=24,and max=538. **TB Mortality** also has less variaty (SD=54.31 compared to TB Incidence's SD=183.25), so in most of our senarios we cosider it as better factor that shows us how TB changed all over the world.

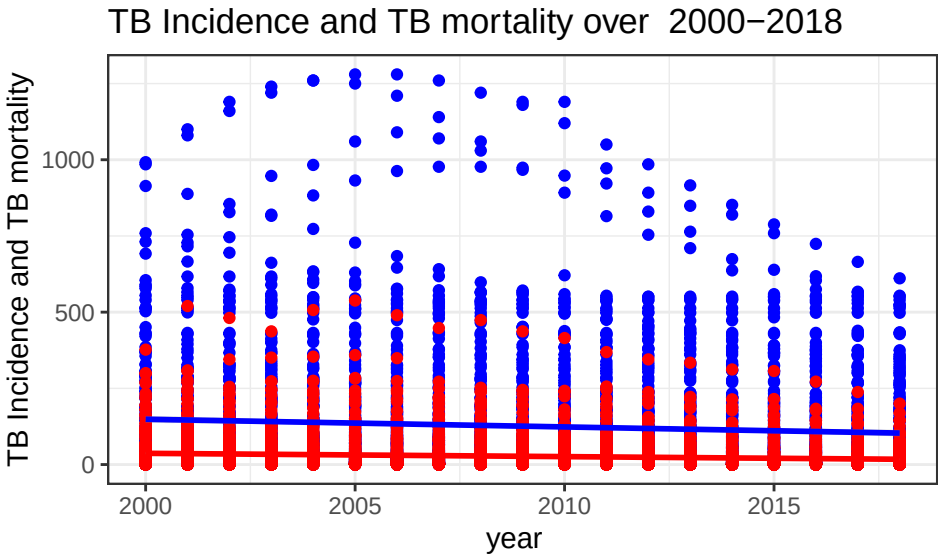
Number Of visits talk about typical number of visits that a TB patient makes to a health facility after diagnosis. We also consider this variable as a factor that shows how medical improvement can help TB patients to cure. The mean of **Number Of visits** is 67.09 when we considered all observations, but this number was modified to 66.23 after the elimination of an outlier. This outlier was a report of 2585 visits which is 8 times higher than that of next highest report and does not make sense; a person has to visit a health place 3 times in a week for 18 years or at leas every day for 7 years to reach this number. For **Number Of visits**, there is 93 missing variables: we beleive that this is because of a poor report system in health care of some countries and we can remove it casewise.

country and **Who Region** are both categorial variables which name and group different areas of the world. **year** in this dataset varies from 2000-2018.

Planning

Based on WHO's claim ,53 million of TB patients have cured since 2000. As a result we wanted to check whether the improvement in diagnosis and treatment methods caused the overal **TB** cases to decrease all

around world.



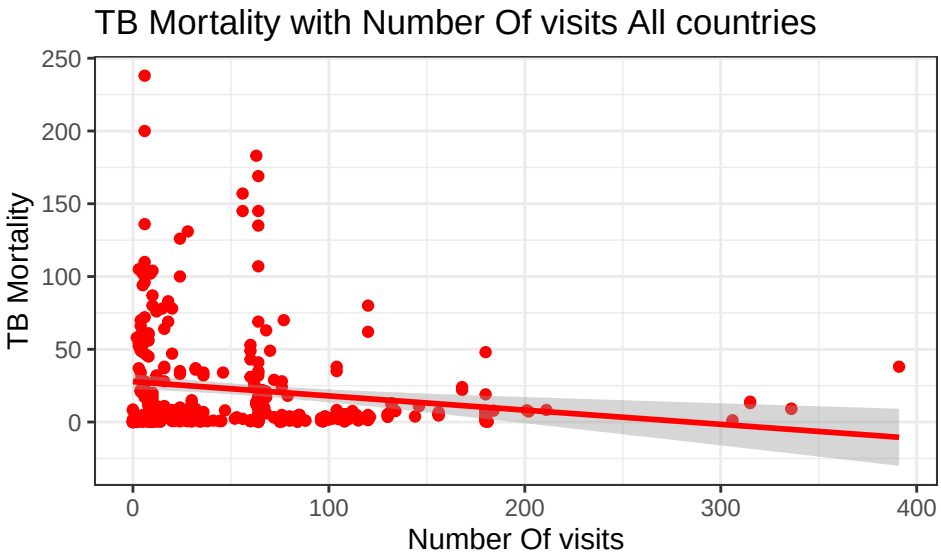
As we can see neither **TB Incidence**(Blue) nor **TB Mortality**(red) significantly decreased in these years which is in conflict with improvement in global health care for TB. We believe cosidering these relations over 18 years and over so many countries might cause many variables to has an effect on trend of TB factors.To get a better result we consider the changes over the 2 last years and tried to study the relation of **TB Mortality** ,which is a better factor for TB as we discussed, with **Number Of visits**, which we assume shows how medical treatment can fight TB.

Based on all these information and assumptions, we develop following hypothesis:

H0: **TB Mortality** does not have a correlation with **Number Of visits**. H1:**TB Mortality** have a correlation with **Number Of visits**.

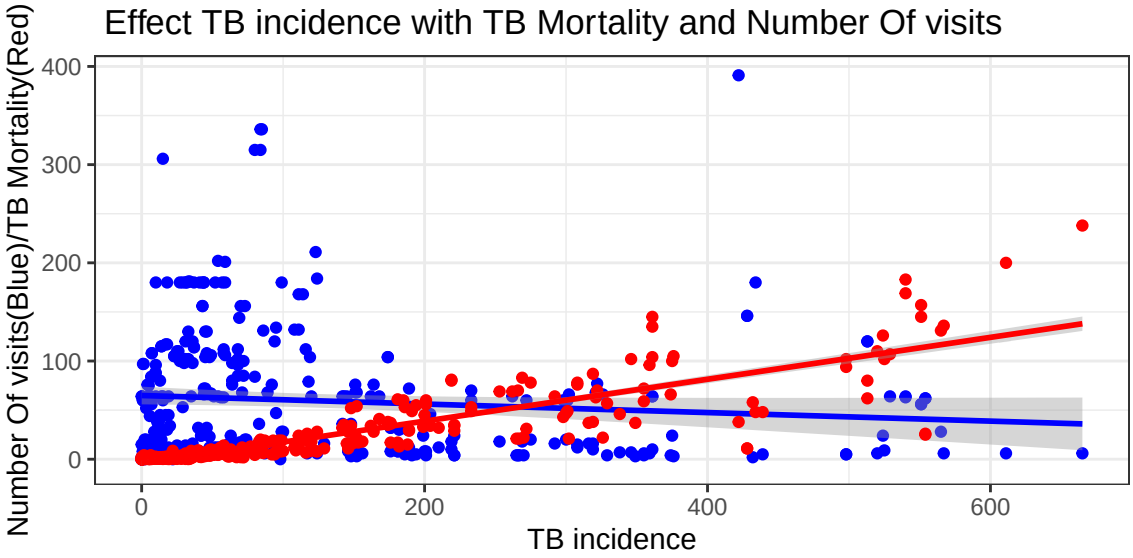
Based on our variables we use Pearson correlation, the only assumption ,interval data ,is also met. We then tried to consider the effect of **TB Incidence** in order to get a better result and we carry out partial correlation test for our hypothesis.

Analysis



Based on graph and correlatin testing result, we can say, by 95% confidence, that Our hypothesis is true and **Number Of visits** of patients have association with **TB Mortality** and r=-0.1775.

Although our hypothesis is true, we believe the correlation is not big enough (as we expected), so we think maybe other variables have and effect on the correlation of TB Mortality and Number Of Visits. As an example we check the relation of TB Incidence with these two variables.



TB Incidence has a strong positive effect on TB Mortality (cor=.87) which is obvious and has a negative small effect on Number of visits (cor=-.1). we beleive that if we could eliminate the effect of TB Incidence from the relation of TB Mortality and Number of Visits, our correlation will be more accurate and we could test hypothesis more strongly.

```
## [1] -0.18842
```

As can be seen from the table above, the partial correlation of TB Mortality and Number of visits is -0.18842, and this shows tha TB incidence does not change the correlation as we expected.

Conclusion

Based on WHO's claim and public evidence TB has decreased over past few years. when we study this decrease based on two factors of **TB Incidence** and **TB Mortality** we did not find this decline as strong as we expected. We then tried to analyze the relations of **TB Mortality** and **Number Of visits** without the effect of other variable such as **TB Incidence** to limit the effect of unwated variables, but the results showed us that this change have a small effect on this association. we assumed that this relation is influenced by many other variables and as an exmaple we study the effect of TB incidence on our main variables. We then realize that when we eliminate the effect of TB incidence our model not improves as we expected. We can conclude that in order to study how TB changed over years ,we have to consider the behaviour of many variables but we should not jump to conclusion because the the effect of these variables need more critical analysis. It might also be better to categorize countries based on the level of issue and try to find out some more variables to investigate the issue in these subgroups.