

TB burden- Partial correlation Report

Data

In the burden data set we have 3591 observations about mortality due to all forms, estimated cases having both TB and HIV disease and the TB treatment coverage provided across 217 countries. Since all these interval variables were significantly skewed, and contains some outliers ,we have performed log transformation for achieving normal distribution. We removed records having null values.[Appx-1,2,3]

Analysis of the TB treatment coverage data [Appx-4,5,6] showed that mortality(due to all forms) was negatively related to TB treatment coverage, but positively related to no. of cases having both HIV and TB, and (TB+HIV)cases itself was negatively related to TB treatment coverage.Now, we know that mortality is related to both TB+HIV cases and TB treatment coverage, then to analyze exact relationship between mortality and TB treatment coverage need to take account of the effect of no of cases having both TB and HIV.

R^2	mortality	treatment_coverage	cases_tbhiv
Mortality	100.00	42.90	83.08
treatment_coverage	42.90	100.00	33.82
cases_tbhiv	83.08	33.82	100.00

Using the values of R2 for these relationships [Appx-7], we know that TB treatment coverage accounts for 42.90% of the variance in mortality, that no of cases accounts for 83.08% of the variance in mortality, and that no of cases accounts for 33.82% of the variance in TB treatment coverage. If no of cases accounts for 33.82% the variance in TB treatment coverage, then it seems feasible that at least some of the 42.90% of variance in mortality that is accounted for by TB treatment coverage is the same variance that is accounted for by no of cases.

Planning (Assumptions):

- 1: You have one (outcome) variable mortality and (predictor) variable TB treatment coverage and these are both interval.
- 2: You have one control variable no of cases having both TB & HIV and is also interval data.
- 3: There is a linear relationship between all three variables (Visualized using scatterplot [Appx-8, 9, 10])
- 4: Since partial correlation is sensitive to outliers, we have already taken logarithmic transformation of data.
- 5: Variables should be approximately normally distributed. This is tested with Shapiro test for normality.From table we can see that all the variables are not normal ($p < 0.05$), however, due to large sample size i.e. 3591 we can assume normal distribution.

1	Data	Statistic	p value
2	mortality	0.99	4.3e-17
3	coverage	0.8	4.4e-55
4	cases	0.98	2.1e-23

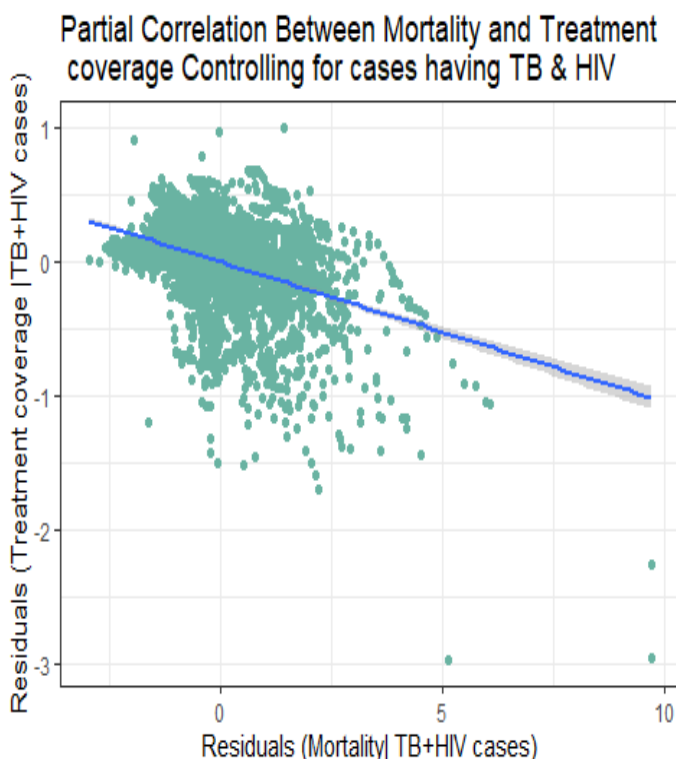
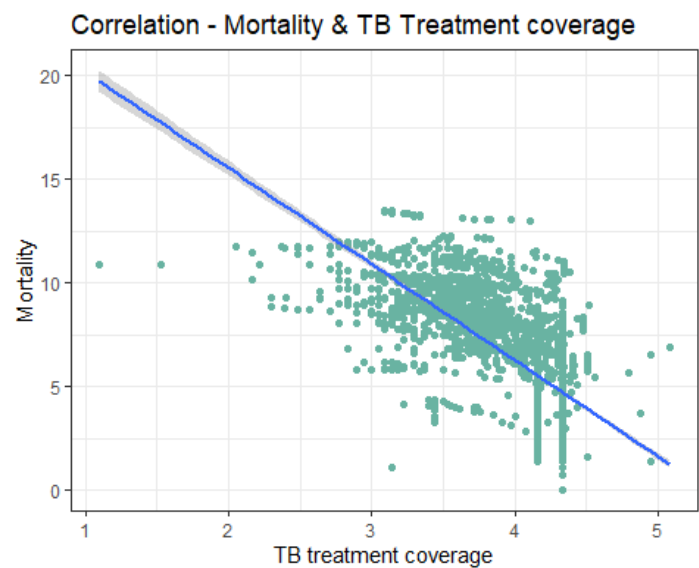
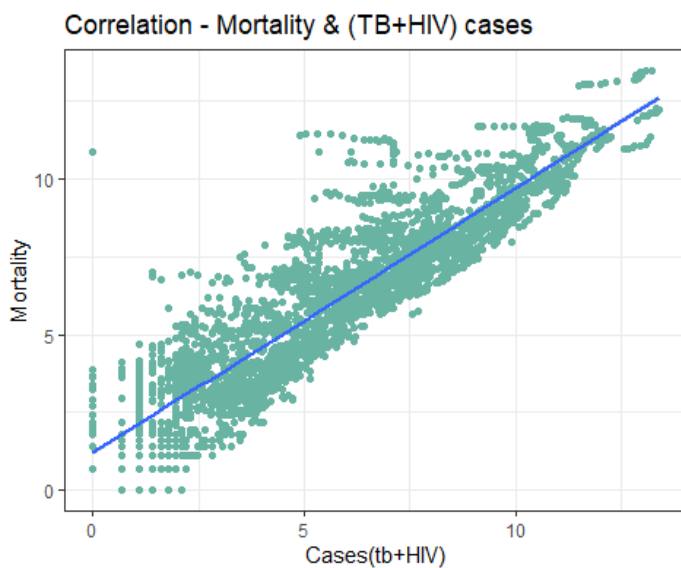
Analysis:

Table shows output for the partial correlation of TB treatment coverage and mortality controlling for number of cases having TB & HIV. [Appx-11, 12, 13]

pc	pc^2	t value	df	p value
-0.37	0.14	-24.09	3588	5.4e-119

1. The partial correlation between mortality and TB treatment coverage is -0.37 which is very less than the correlation when effect of no of cases having TB+HIV is not controlled for -0.65 (p-value < 2.2e-16).
2. The correlation coefficient is nearly half what it was before. While this correlation is still statistically significant (its p-value is 5.4e-119, which is very smaller than .05).
3. The value of R2 for the partial correlation is 0.14, which means that TB treatment coverage can now account for only 14% of the variance in mortality.
4. When the effects of number of cases were not controlled for, TB treatment coverage shared 42.90% of the variation in mortality and so the inclusion of number of cases has severely diminished the amount of variation in Mortality shared by TB treatment coverage.

As such, a more correct measure of the role of TB treatment coverage has been obtained. Running this analysis has shown us that TB treatment coverage alone does explain some of the variation in Mortality, but there is a complex relationship between TB treatment coverage, no of cases with TB+HIV and mortality that might otherwise have been ignored. Although causality is still not certain, because relevant variables are being included, the third variable problem is, at least, being addressed in some form.



Conclusion:

“A partial correlation coefficient was computed to assess the relationship between the mortality (all forms) and treatment coverage by controlling the no. of cases having both HIV & TB. The partial correlation between mortality and TB treatment coverage is -0.37 which is very less than the bi-variate correlation when the effect of no of cases having TB+HIV is not controlled for -0.65 (p-value < 2.2e-16). A lower partial r than bi-variate r means that much of the correlation was accounted for in the no of cases having TB & HIV. Once we REMOVED the correlation between mortality and no of cases, treatment coverage was no longer strongly correlated with mortality. That is because mortality and no of cases are so strongly correlated. A scatterplot summarizes the results Overall.”