Predicting Tuberculosis Fatality Rate

By: Evan Zimmerman,Paige Crawford, Drew Sternberg, and Shiloh Gelnett

Introduction

- Tuberculosis is a deadly disease that can have extreme consequences if not identified and treated properly. This increasingly deadly disease has motivated us to consider what specific factors affect the fatality rate for TB cases. We investigate several factors such as HIV rate, GDP, and government spending on health care:
- Government spending on healthcare: According to an article by the National Institutes of Health, "A hospital-based TB diagnosis is a critical opportunity to identify those at high risk of early and overall mortality" (Osman, 2021, pg. 1)
- HIV: For example, "Recent data estimates show that 3-7 million HIV patients develop TB per year and up to 5 million people develop acute pulmonary TB" (Obeagu, 2023, pg. 128)
- GDP: GDP per capita "was highest for the type IV countries (high-income countries), which indicates that a lower TB incidence is accompanied by increasing affluence" (Lei, 2023, pg. 5)

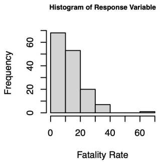
Research Questions

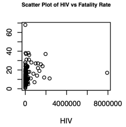
1. Does a higher HIV rate in a country make the tuberculosis death rate higher?
2. Does a country having a low GDP make the tuberculosis death rate higher?
3. Do the countries with the least amount of government spending on healthcare have the highest tuberculosis death rate?

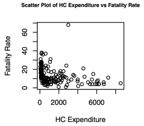
Data Summary

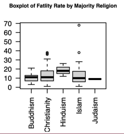
Variable	Description
Fatality rate	Percentage of cases of people with TB that die from it in each country; units: fatality percentage
GDP per capita (GDP)	The gross domestic product that measures a country's economic activity; units: Currency in US dollars
Healthcare expenditure (HC)	The amount spent on healthcare services by residents divided by a country's population; units: currency in US dollars
Number of people living with HIV (HIV)	The number of people living with diagnosed HIV (human immunodeficiency virus) in each country; units: number of people living with diagnosed HIV
Multidrug resistant Tuberculosis (MULTIDRUG)	A type of TB that is resistant to at least 2 different types of anti-TB drugs; units: number of people with diagnosed multidrug resistant TB
Rate of new cases (NEWTB)	The proportion of people recently diagnosed with TB per 100,000 people in each country; units: proportion of people with recently diagnosed TB
Political regime	A set of rules, protocols, and cultural norms that regulate how a government functions; units: N/A; Levels: Closed autocracy - 0, Electoral autocracy - 1, Electoral democracy - 2, and Liberal democracy - 3
Hemisphere	Hemisphere that a country belongs to; units: N/A; Levels: Northern, Southern
Religious Affiliation)	Dominant religion in each country; units: N/A; Levels: Christianity, Islam, Buddhism, Hinduism, Judaism, Other

EDA





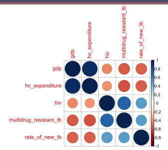




Model Building & Assumptions

Multicollinearity

- Correlation matrix indicates strong pairwise relationship between GDP and healthcare expenditure.
- Average VIF = 6.53, GDP VIF = 12.47, healthcare expenditure VIF = 13.37
- Built two models: one with gdp and one with healthcare expenditure



Stage One Quantitative Variables

Initial: FR=β0+β1log(GDP)+β2log(HIV)+β3log(MULTIDRUG)+β4(NEWTB)
Final: FR= β0 +β1log(GDP)+β2log(HIV)+β3(NEWTB)
*Quant x Quant interactions were not significant

Stage Two Qualitative Variables

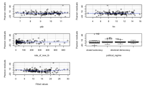
Initial: FR= β0 +β1log(GDP)+β2log(HIV)+β3(NEWTB)+β4(EA)+β5(ED)+β6(LD)+β7(South)+β8(Christian)+β9(Hindu)+β10(Islam)+β11(Judaism)
Final: FR = β0 +β1log(GDP)+β2log(HIV)+β3(NEWTB)
- EA = {1 if electoral autocracy, 0 otherwise}, ED = {1 if electoral democracy, 0 otherwise}, LD = {1 if liberal democracy, 0 otherwise}, South = {1 if southern hemisphere, 0 otherwise}, Christian = {1 if Christianity, 0 otherwise}, Hindu = {1 if Hinduism, otherwise}, Islam = {1 if Islam, 0 otherwise}, Judaism = {1 if Judaism, 0 otherwise}
*hemisphere x political regime interaction was not significant

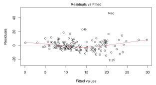
Stage Three Qual x Quant Interaction

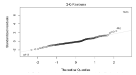
Initial: FR = β0+β1log(GDP)+β2log(HIV)+β3(NEWTB)+β4EA+β5ED+β6LD+β7log(GDP)*EA+β8log(GDP)*ED+β9log(GDP)*LD
Final: FR = β0+β1log(GDP)+β2log(HIV)+β3(NEWTB)+β4EA+β5ED+β6LD+β7log(GDP)*EA+β8log(GDP)*ED+β9log(GDP)*LD

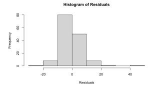
Assumptions

- No assumptions were violated except for a minor violation of constant variance seen in the rate of new TB residual plot
- We corrected the violation by log transforming rate of new TB. However, we decided that it made our model less statistically significant so we kept the minor violation.







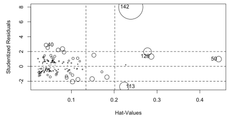


Removing Influential Observations

Outliers observed and removed:
40,50,113,128,142

	StudRes <dbl>	Hat <dbl>	CookD <dbl>
40	2.8756752	0.04098799	0.03358716
50	0.9746034	0.44386767	0.07583819
113	-2.7506685	0.22355040	0.20801421
128	1.3231661	0.28648184	0.06991678
142	7.9849503	0.23873372	1.37753478

5 rows



Variable Screening

Stepwise regression:
-Multi-drug removed
-Kept GDP, HIV, and rate of new TB
- p_ent and p_rem=0.15

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	25.138	8.306		3.026	0.003	8.721	41.555
gdp	-1.915	0.725	-0.234	-2.639	0.009	-3.348	-0.481
rate_of_new_tb	0.013	0.006	0.197	2.125	0.035	0.001	0.026
hiv	0.484	0.321	0.132	1.508	0.134	-0.150	1.118

External Model Validation

Weighted least squares:
-Weight was inverse of residual squared.
-Significantly overfitted the model.
- Adj R^2 = 0.9855

```
Residual standard error: 0.9463 on 134 degrees of freedom
Multiple R-squared:  0.9864,    Adjusted R-squared:  0.9855
F-statistic:  1084 on 9 and 134 DF,  p-value: < 0.00000000000000022
```

Conclusion

Final Model: FR = -1.11 + 0.76log(GDP) + 0.42log(HIV) + 0.01(NEWTB) + 35.71EA + 46.38ED + 67.25LD -3.86log(GDP)*EA -4.79log(GDP)*ED -6.46log(GDP)*LD
- RMSE: 6.04
- P-value: <0.001
- Adjusted R^2: 0.3625

Interpretation: Increased GDP, higher HIV prevalence, and a greater number of new tuberculosis cases are associated with higher fatality rates, if all other variables are held constant . However, the impact of GDP on increasing fatality rates is mitigated in the political regimes represented by EA, ED, and LD. While these groups exhibit higher fatality rates compared to the baseline, the positive effect of GDP on fatality rates is diminished within these groups as seen by the negative coefficients in the interaction terms (i.e. as GDP increases for a country with an electoral autocracy regime, the log(GDP)*EA interaction term has a greater negative value), This suggests a complex interplay between economic development and health outcomes under different political regimes.

Efficiency: The model is statistically significant as indicated by the very low P-value, suggesting that the predictors have a meaningful contribution to the model. However, the Adjusted R^2 value shows that 36.25% of variability in the fatality rate hasn't been captured by the model. This might suggest the need for additional variables, more complex modeling, or that inherent variability in the dataset is high.

Example: We wanted to test how accurate our equation is at predicting the fatality rate of TB in Austria based on the data about its GDP, HIV rate, rate of new TB cases, and political regime. Our model predicts a fatality rate of 8.05%, and the actual fatality rate is 8.00%, resulting in a residual of 0.05.

Improvements

- Including air quality and alcohol consumption for additional explanatory variables
- Using more recent, current data instead of 2019 data, however we only able to find 2019 data
- We would also add more explanatory variables in general

References

Healthcare expenditure vs. GDP per capita. Our World in Data. (n.d.). https://ourworldindata.org/grapher/healthcare-expenditure-vs-gdp?tab=table

Lei, Y., Wang, J., Wang, Y., & Xu, C. (2023). Geographical evolutionary pathway of global tuberculosis incidence trends. *BMC Public Health*, *23*(1), 755.

Northern Hemisphere countries 2024. (n.d.). https://worldpopulationreview.com/country-rankings/northern-hemisphere-countries

Number of people living with HIV. Our World in Data. (n.d.-b). https://ourworldindata.org/grapher/number-of-people-living-with-hiv?tab=table

Obeagu, E. I., & Onuoha, E. C. (2023). Tuberculosis among HIV Patients: A review of Prevalence and Associated Factors. *Int. J. Adv. Res. Biol. Sci*, *10*(9), 128-134.

Osman, M., Meehan, S. A., von Delft, A., Du Preez, K., Dunbar, R., Marx, F. M., Boulle, A., Welte, A., Naidoo, P., & Hesseling, A. C. (2021). Early mortality in tuberculosis patients initially lost to follow up following diagnosis in provincial hospitals and primary health care facilities in Western Cape, South Africa. *PloS one*, *16*(6), e0252084. https://doi.org/10.1371/journal.pone.0252084

Pai, M., Dewan, P. K., & Swaminathan, S. (2023). Transforming tuberculosis diagnosis. *Nature Microbiology*, *8*(5), 756-759.
Pew Research Center. (2022, December 21). *Religious composition by country, 2010-2050*. Pew Research Center's Religion & Public Life Project. https://www.pewresearch.org/religion/interactives/religious-composition-by-country-2010-2050/

Saloni Dattani, Fiona Spooner, Hannah Ritchie and Max Roser (2023) - "Tuberculosis" Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/tuberculosis' [Online Resource]

Southern Hemisphere countries 2024. (n.d.). https://worldpopulationreview.com/country-rankings/southern-hemisphere-countries
V-Dem (2023) – with major processing by Our World in Data. "Political regime – Regimes of the World" [dataset]. V-Dem, "Democracy and Human rights, OWID based on Varieties of Democracy (v13) and Regimes of the World v13" [original data]. Retrieved March 23, 2024 from https://ourworldindata.org/grapher/political-regime

WHO (2023); Population based on various sources (2023) – with minor processing by Our World in Data. "Rate of new tuberculosis cases" [dataset]. WHO, "Global Tuberculosis Report"; Various sources, "Population" [original data]. Retrieved March 23, 2024 from https://ourworldindata.org/grapher/incidence-of-tuberculosis-sdgs

World Bank (2023) – with minor processing by Our World in Data. "GDP per capita – World Bank" [dataset]. World Bank, "World Bank World Development Indicators" [original data]. Retrieved March 23, 2024 from https://ourworldindata.org/grapher/gdp-per-capita-worldbank