HIGH SCHOOL EDITION
Journal of Student Research
Volume 11 Issue 4 (2022)

Effects of Rural and Urban Resource Access on the Tuberculosis Burden in Least Developed Countries

Michael Chang[1] and Alfred Renaud[#]

[1]Newton South High School, USA
[#]Advisor

ABSTRACT

In 2020 alone, an estimated 1,500,000 deaths occurred due to tuberculosis (TB). TB continues to affect over 10 million people yearly, primarily in underdeveloped countries. The present paper uses regression analysis to compare the correlations between increased access to improved water and sanitation resources in rural versus urban areas and the national TB burden (incidence and mortality rates). The results suggest that enhancing water and sanitation resources reduces the national TB burden. Moreover, urban improvements correspond with a greater national TB burden reduction than rural improvements.

Introduction

Among the leading causes of death, tuberculosis (TB) mortality has declined from 7th to 13th place, but it stays prevalent. Least Developed Countries (LDCs) are low-income countries facing structural impediments to sustainable development. TB is exacerbated in these nations by poor infrastructure and remains in the top 10 causes of death (World Health Organization [WHO], 2000, 2021a).

In September 2018, the United Nations Member States affirmed their commitments to the Sustainable Development Goals (SDGs) and the WHO's End TB Strategy. SDG Target 3.3 aspires to end communicable disease epidemics by 2030. Environmental variables like water and sanitation are key determinants of the probability of disease. Cardoso (2017) reported that access to basic sanitation plays a factor in the probability of TB infection. Furthermore, Marcos-Garcia (2021) found a strong relationship between high respiratory infection mortality rates and low access to water, sanitation, and hygiene (WASH) services. Marcos-Garcia's discovery of this relationship led them to conclude that integrating the improvement of WASH services would reduce the mortality of prevalent diseases.

Critically, access to improved water and sanitation resources is limited in LDCs and often overlooked by public officials. The present paper uses linear regression to analyze correlations of improving resource allocation in rural and urban areas of LDCs and their national TB burden (incidence and mortality rates).

Materials and Methods

All code for cleaning, visualizing, and modeling data was written in Python using Jupyter Notebook. The libraries used were *Matplotlib*, *NumPy*, *Pandas*, *scikit-learn*, *Seaborn*, *and statsmodels*.

Datasets

1. TB burden estimates: A dataset from the WHO's 2021 edition of the Global Tuberculosis Report containing TB estimates for cases and deaths in all WHO member countries from 2000 to 2020 (WHO, 2021b).

Volume 11 Issue 4 (2022)

2. WHO and United Nations Children's Fund Joint Monitoring Programme (JMP) for WASH: Four datasets covering global household water, sanitation, hygiene, and menstrual health resources in WHO member states from 2000 to 2020. The data consisted of percentages for variables that indicate population access to resources at a rural, urban, and national level (JMP, 2021).

Merging Datasets

The primary datasets of interest were TB burden estimates, water resources, and sanitation resources. The hygiene and menstrual health datasets were not used. The three datasets were then filtered by removing non-LDCs, and the information was compiled into a single dataset using the rows' country and year variables. The compiled dataset had information on 46 LDCs over 21 years.

Variables

A correlation heatmap (Figure 1) was created to visualize the correlation between different variables in the combined dataset.

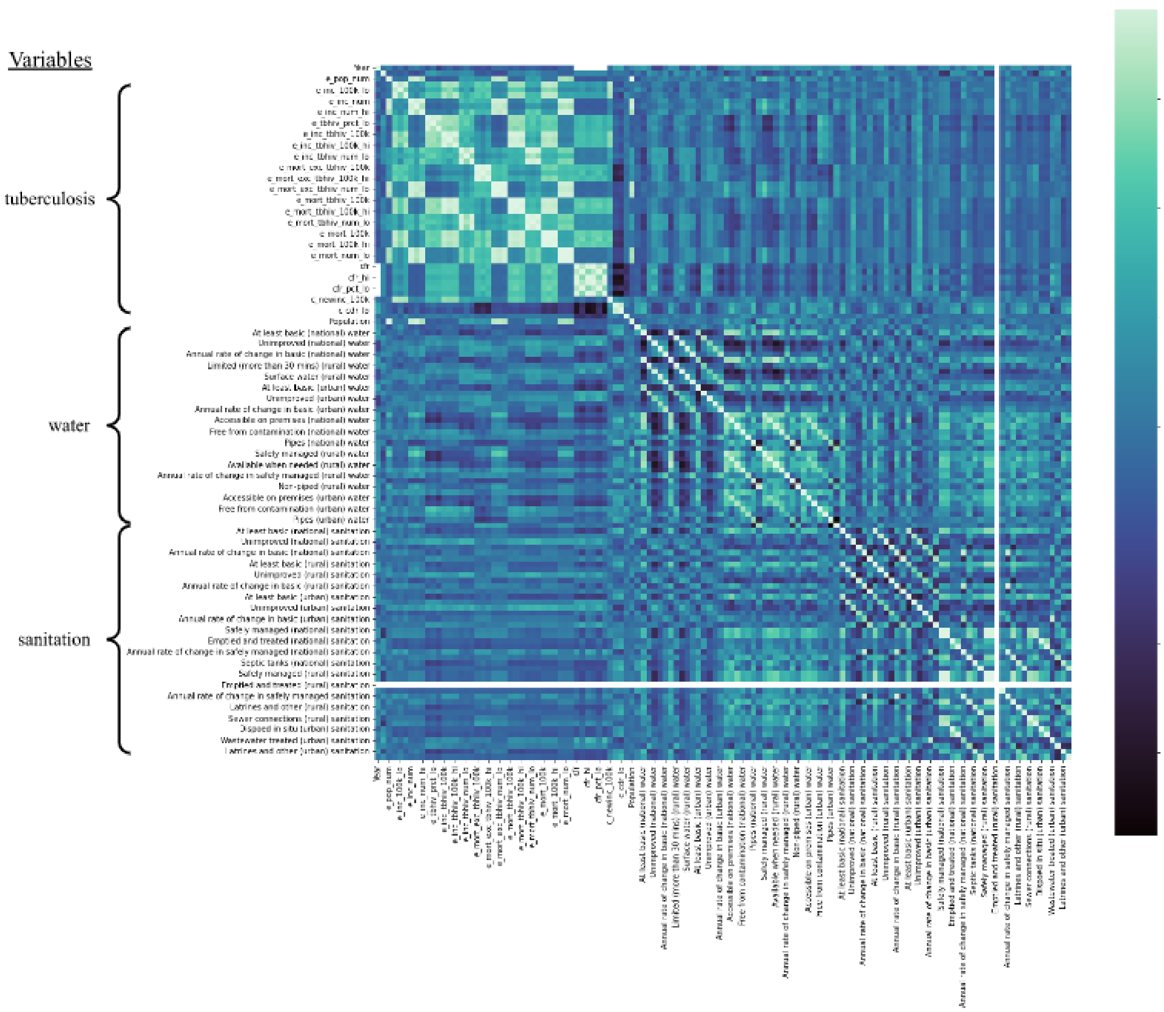

Figure 1. Initial correlation heatmap of combined dataset variables.

The original correlation heatmap had all the variables within the combined dataset that were numerical values and displayed different positive and negative correlations between individual variables. Desired variables were chosen from the initial list of 134 variables, so the correlation heatmap was narrowed down to two smaller (rural and urban) heatmaps. Figures 2 and 3 show the resulting heatmaps.

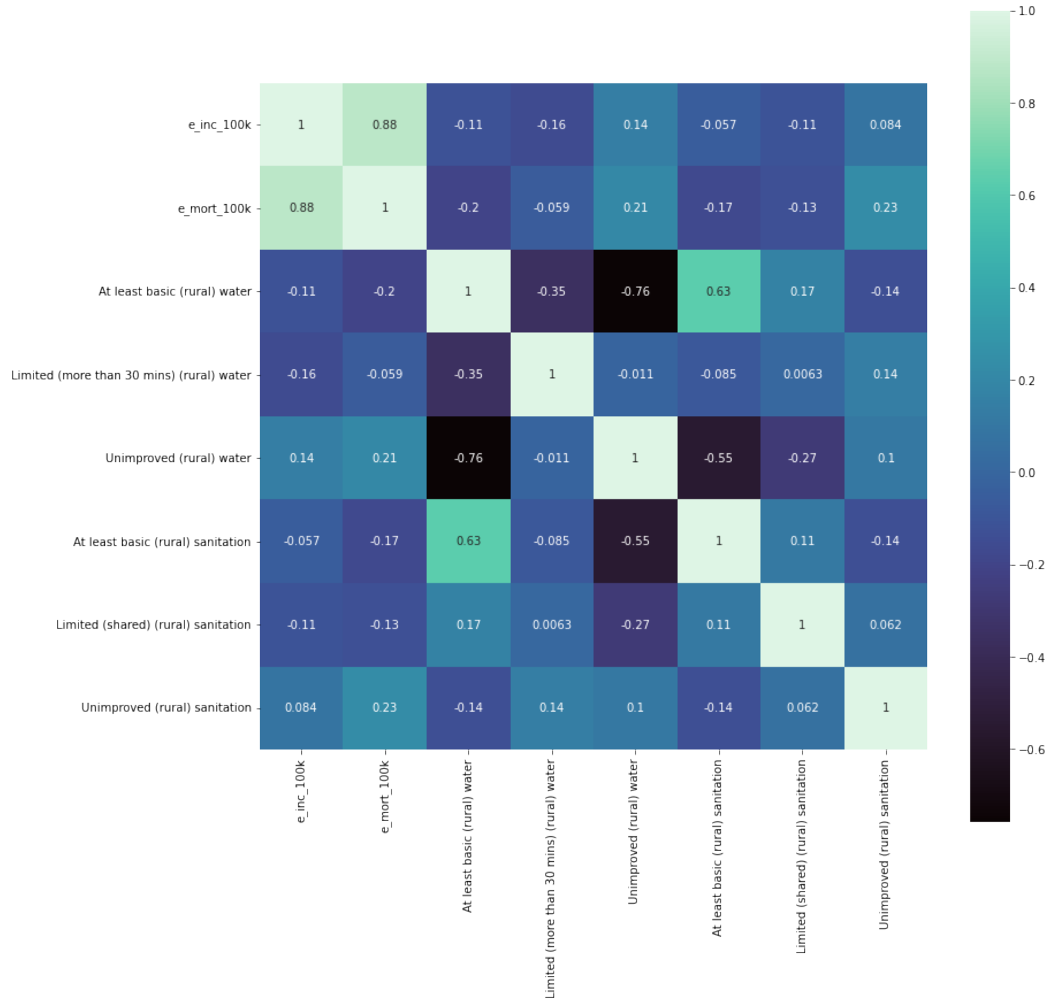

Figure 2. Rural independent variables and TB mortality and incidence variables.

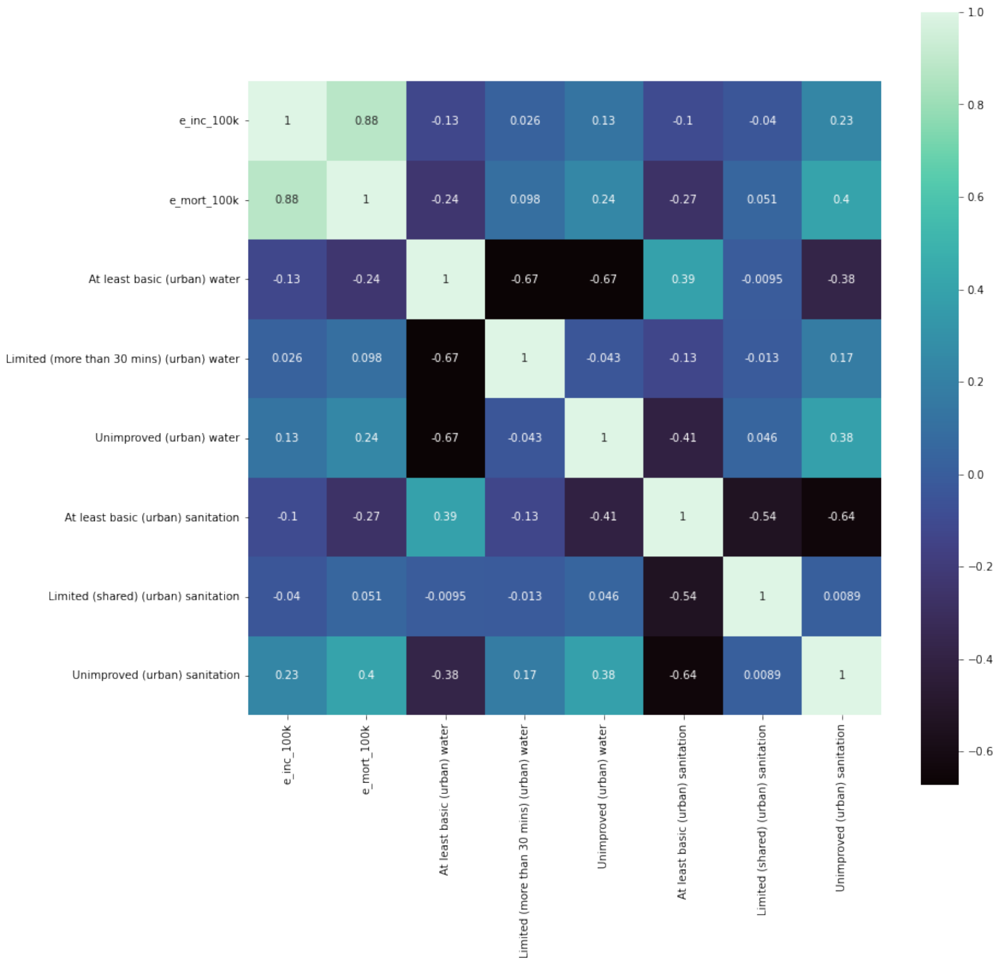

Figure 3. Urban independent variables and TB mortality and incidence variables.

Table 1. Definition of independent variables.

Water	Sanitation
Basic	
Drinking water from an improved source, provided collection time is not more than 30 minutes for a roundtrip, including queuing	Use of improved facilities that are not shared with other households
Limited	
Drinking water from an improved source for which collection time exceeds 30 minutes for a roundtrip, including queuing	Use of improved facilities shared between two or more households
Unimproved	
Drinking water from an unprotected dug well or unprotected spring	Use of pit latrines without a slab or platform, hanging latrines, or bucket latrines

Note. The definitions are from the JMP service ladder (UN, 2017). Improved drinking water sources include piped water, boreholes, tube wells, protected dug wells or springs, rainwater, and packaged or delivered water. Improved sanitation facilities include composting toilets, pit latrines with slabs, or flush/pour toilets connected to piped sewer systems or septic tanks.

HIGH SCHOOL EDITION
Journal of Student Research

As described in Table 1, the basic and limited conditions for water and sanitation quality represent improved resources. Thus, the two variables were summed, forming the new columns "improved water" and "improved sanitation." Ultimately, the rural and urban rural independent variables were narrowed down to improved sanitation, unimproved sanitation, improved water, and unimproved water.

The dependent variables were from the WHO's TB burden estimates dataset. The present paper analyzed the TB incidence rate per 100,000 people (e_inc_100k) and the TB mortality rate per 100,000 people (e_mort_100k).

Data Cleaning

The dataset was further cleaned before any modeling was done. Two LDCs — Timor-Leste and South Sudan — formed in 2002 and 2011, were missing data and were removed. Next, each independent variable was plotted against the dependent variables to catch outliers, as shown in Figure 4.

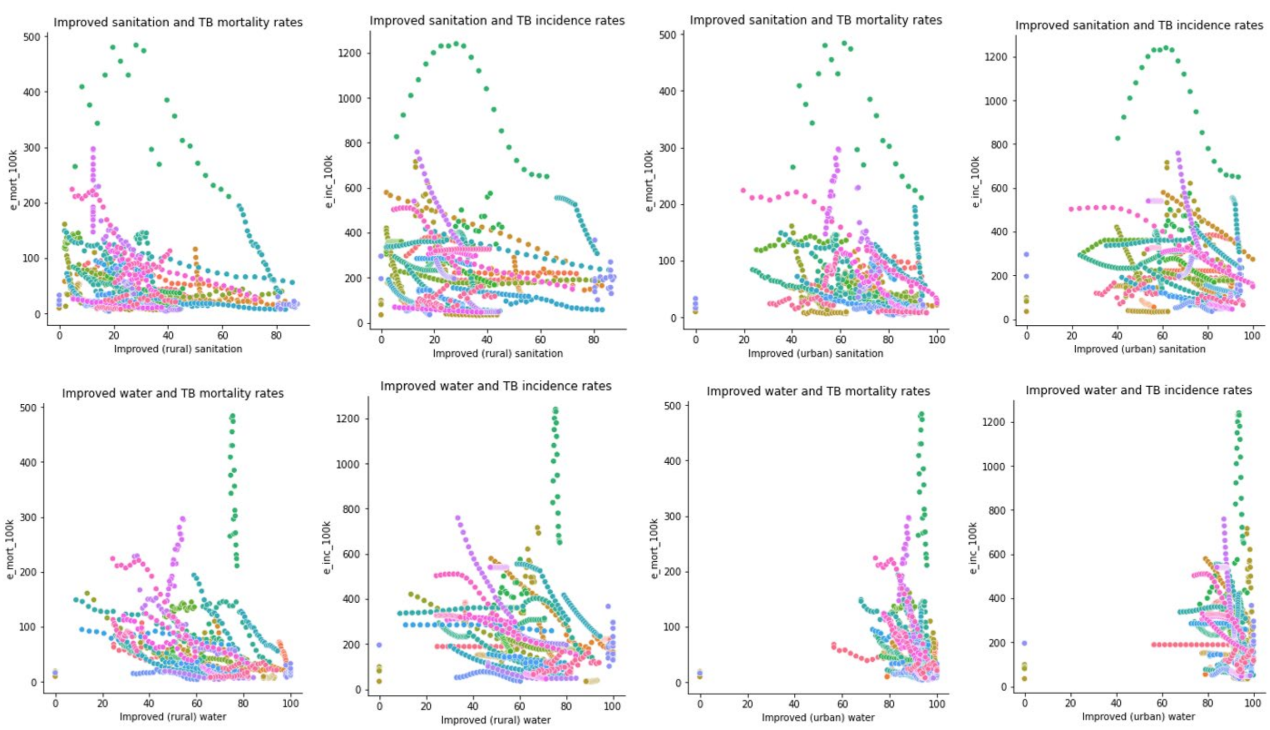

Figure 4. Plots of desired independent variables and dependent variables. The colors stand for different countries.

The topmost streaks of light green and blue were clearly separated from the general trend of data. Those streaks represented Lesotho and Myanmar, which were dropped from the dataset. Finally, the few data points at the bottom lefthand corners with less than five percent improved urban water and sanitation were removed. In the end, the final dataset had 826 entries.

Data Modeling

The cleaned dataset was split into 80% training and 20% testing with scikit-learn's random_state = 42. The present paper implemented linear regression to model the trends of TB incidence and mortality. Ordinary Least Squares was the method of choice used to estimate the best coefficients for a linear model by minimizing the sum of squared residuals.

Results

Tables 2 and 3 show the results from running multiple iterations of linear regression with one independent variable and one dependent variable.

Table 2. Results of linear regression with separate iterations for each independent variable and TB incidence.

Independent Variable	Coefficient	R-squared train	R-squared test	MSE	RMSE	p-value
Improved (urban) water	-5.925	0.076	0.114	18498.512	136.009	0.000
Improved (urban) sanitation	-1.929	0.052	0.045	18973.980	137.746	0.000
Improved (rural) water	-2.744	0.142	0.163	17169.568	131.033	0.000
Improved (rural) sanitation	-1.522	0.044	0.041	19123.170	138.287	0.000

Table 3. Results of linear regression with separate iterations for each independent variable and TB mortality.

Independent Variable	Coefficient	R-squared train	R-squared test	MSE	RMSE	p-value
Improved (urban) water	-3.158	0.168	0.184	2125.072	46.099	0.0000
Improved (urban) sanitation	-1.209	0.159	0.134	2147.822	46.345	0.0000
Improved (rural) water	-1.166	0.201	0.210	2041.943	45.188	0.0000
Improved (rural) sanitation	-0.873	0.114	0.140	2262.908	47.570	0.0000

The R-squared values of the linear regression models were surprisingly low and can be explained by high variability in the data, with points far from the line of best fit. The range for mortality was between 3.6 to 297 deaths per 100,000 people, while the range for incidence was between 33 to 759 cases per 100,000 people. The Root Mean Square Error (RMSE), which indicates variability, was smaller for TB mortality than the RMSE for TB incidence. As such, the R-squared values for TB mortality were marginally better than those for TB incidence.

Despite the low R-squared values among all iterations, the p-values show that the independent variables are highly statistically significant, with all values satisfying the condition $p < 0.001$.

The coefficients for improved urban resources were larger compared to their corresponding rural labels for both TB incidence and mortality. For TB incidence, the urban coefficients were -5.925 and -1.929 for water and sanitation, compared with rural coefficients of -2.744 and -1.522. For TB mortality, the urban coefficients were -3.158 and -1.166 for water and sanitation, compared with rural coefficients of -1.209 and -0.873.

Discussion

The present paper used linear regression to analyze correlations between improved water and sanitation and the national TB burden. Because the units for WASH variables in the dataset are percentages, an increase in access to improved resources corresponds with a reduction in the usage of worse quality resources. All the coefficients of the improved resources are negative, which means that better quality resources negatively correlate with the national TB burden. This observation is consistent with prior literature.

The present paper also revealed that improving resources in urban areas enhances the health of a greater proportion of the population. The specific values of regression coefficients were analyzed and interpreted. These coefficients denote mean changes. For example, the coefficient for improved urban water implies that if there is a one percent increase in improved urban water, there will be an average decrease in the national TB mortality rate by 3.158 deaths per 100,000 people. The larger urban coefficients showed that increasing access to improved water and sanitation in urban areas rather than rural areas significantly reduced TB incidence and mortality rates. A plausible explanation could be that in LDCs, most of the population lives in urban areas.

Conclusion

TB continues to ravage low-income countries. The present paper analyzed correlations between water and sanitation variables and the national TB burden. The results indicated that improvements in basic resources such as water and sanitation negatively correlate with TB incidence and mortality rates.

In addition, the present paper identified a difference between increasing access (in low-income countries) to improved resources in rural versus urban areas. Development in urban water and sanitation resources led to a greater correlation with the reduction of national TB incidence and mortality rates. This finding could influence resource management decisions in low-income countries.

Although there are numerous variables to consider in the battle against the TB epidemic, increasing access to fundamental resources is a crucial step in advancing population health and reducing preventable diseases.

Limitations

There is a large gap in reported TB cases. Of the 10 million people estimated to contract TB yearly, only 71% were identified and reported in 2019. Furthermore, national statistics are less likely to include rural patients who have reduced access to healthcare (Sikalengo, 2018). Underreported rural cases can mask the actual rate of decline of the national TB burden and the significance of rural resource improvements. In addition, the WHO estimates the TB burden using data from national prevalence surveys. While estimation methods are continuously being improved, relatively wide variation remains within country-level estimates for lower and upper bounds of TB incidence rates (Chakaya, 2021). Future research on the rural-urban comparison can consider more variables beyond the WASH dataset as well as the percentage of the population that is categorized as rural or urban.

Acknowledgments

I want to thank Dr. Ramezani for guiding me with research and Alfred Renaud for helping me throughout this project. I would also like to thank my fellow students Evan Assmus and Ryan Ayoub for working with me at the beginning of this project.

References

World Health Organization. (2020). The Top 10 Causes of Death. World Health Organization; World Health Organization. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

World Health Organization. (2021a). Tuberculosis. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/tuberculosis

Cardoso, B. A., Fonseca, F. O., Moraes, A., Neto, Martins, A., Oliveira, N., Lima, L., Dias, G., & Saad, M. (2017). Environmental aspects related to tuberculosis and intestinal parasites in a low-income community of the Brazilian Amazon. *Revista do Instituto de Medicina Tropical de Sao Paulo*, *59*, e57. https://doi.org/10.1590/S1678-9946201759057

Marcos-Garcia, P., Carmona-Moreno, C., López-Puga, J., & Ruiz-Ruano García, A. M. (2021). COVID-19 pandemic in Africa: Is it time for water, sanitation and hygiene to climb up the ladder of global priorities?. *The Science of the total environment*, *791*, 148252. https://doi.org/10.1016/j.scitotenv.2021.148252

World Health Organization. (2021b). WHO TB Burden Estimates. [Dataset]. Retrieved from https://extranet.who.int/tme/generateCSV.asp?ds=estimates

JMP. (2021). Estimates for drinking water, sanitation and hygiene services by country (2000-2020). [Dataset]. Retrieved from https://washdata.org/data/country/WLD/household/download

United Nations. (2017). WHO/UNICEF Joint Monitoring Program for Water Supply, Sanitation and Hygiene (JMP) - 2017 Update and SDG Baselines | UN-Water. UN-Water. https://www.unwater.org/publications/whounicef-joint-monitoring-program-water-supply-sanitation-hygiene-jmp-2017-update-sdg-baselines/

Sikalengo, G., Hella, J., Mhimbira, F., Rutaihwa, L. K., Bani, F., Ndege, R., Sasamalo, M., Kamwela, L., Said, K., Mhalu, G., Mlacha, Y., Hatz, C., Knopp, S., Gagneux, S., Reither, K., Utzinger, J., Tanner, M., Letang, E., Weisser, M., & Fenner, L. (2018). Distinct clinical characteristics and helminth co-infections in adult tuberculosis patients from urban compared to rural Tanzania. *Infectious diseases of poverty*, *7*(1), 24. https://doi.org/10.1186/s40249-018-0404-9

Chakaya, J., Khan, M., Ntoumi, F., Aklillu, E., Fatima, R., Mwaba, P., Kapata, N., Mfinanga, S., Hasnain, S. E., Katoto, P., Bulabula, A., Sam-Agudu, N. A., Nachega, J. B., Tiberi, S., McHugh, T. D., Abubakar, I., & Zumla, A. (2021). Global Tuberculosis Report 2020 - Reflections on the Global TB burden, treatment and prevention efforts. *International journal of infectious diseases: IJID: official publication of the International Society for Infectious Diseases*, *113 Suppl 1*(Suppl 1), S7–S12. https://doi.org/10.1016/j.ijid.2021.02.107