Global Inequalities in Tuberculosis Mortality: A Regional Comparison of Case Detection and Outcomes (2000–2021)

Abstract:

Tuberculosis (TB) remains a major global health threat, and outcomes vary widely across regions due to differences in healthcare access and diagnostic capacity. The goal of this study is to examine whether TB mortality rates differ among the six WHO regions and whether they are associated with case detection rates. Using WHO global Tuberculosis data from 2000-2021, we visualized regional patterns through two boxplots and conducted statistical analysis, including one-way ANOVA, pairwise t-tests, and linear regression inference. Results show significant regional variation in TB mortality, with Europe and the Americas having the lowest mortality burden while Africa and South-East Asia experiencing the highest ones. Regression analysis using log-transformed variables demonstrates an overall strong negative correlation between detection rate and mortality. In general, we conclude that TB mortality is not equally distributed across regions, and lower diagnostic coverage is closely linked to higher mortality, highlighting the need for strengthening detection systems in high-burden areas.

Background & Purpose

Tuberculosis (TB) remains one of the leading infectious causes of death worldwide. Although effective treatments exist, TB outcomes vary across regions due to various reasons. The World Health Organisation classifies countries into six regions – Africa, the Americas, Europe, Eastern Mediterranean, South-East Asia, and the Western Pacificeace – each facing different challenges in TB control.

A key indicator of health system performance is the TB case detection rate (CDR), which reflects the proportion of estimated cases that are successfully identified and treated. In other words, low detection rates indicate barriers such as limited access to diagnostic tools and inadequate healthcare infrastructure. As untreated TB has a fatality rate of nearly 50%, regions with lower detection rates often experience a rapid increase in mortality (WHO, 2019). These systemic factors contribute to regional disparities in TB mortality, making regional comparisons essential for identifying areas needing support.

Research Question & Methodology

This paper investigates TB outcomes from 2000 to 2021 with two primary aims:

1. to determine whether TB mortality rates differ significantly across WHO regions, and
2. to examine how mortality is associated with detection rates

To address these questions, we used the data from the World Health Organisation's Global Tuberculosis Dataset (2000-2021) and conducted related analyses. First, we performed a one-way ANOVA (Analysis of Variance) to test whether the mean mortality rate differs across six WHO regions. This allowed us to determine if regional inequalities exist overall. Then, we ran a pairwise t-test to identify exactly which regions differ from each other. Next, to explore potential causes for these differences, we conducted a linear regression inference between case detection rate and mortality, testing whether higher detection coverage tend to lead to overall lower mortality. If so, we will have evidence showing that diagnostic access plays a role in regional disparities.

Data Analysis

We started with exploratory data visualizations to better understand the regional pattern of tuberculosis outcomes from 2000 to 2021. Two boxplots are drawn to examine the distribution of mortality and case detection rates across the six WHO regions (Figures 1 & 2).

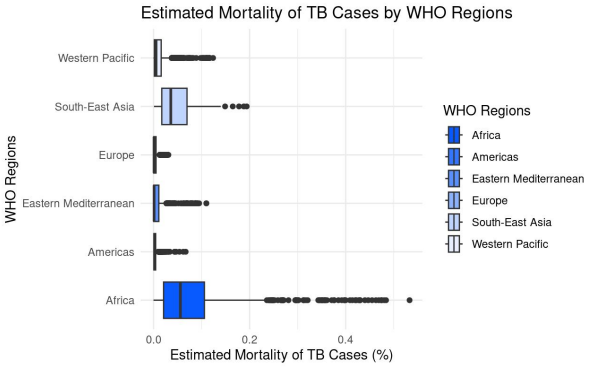


Figure 1 Estimated Mortality of TB Cases by WHO Regions Boxplot

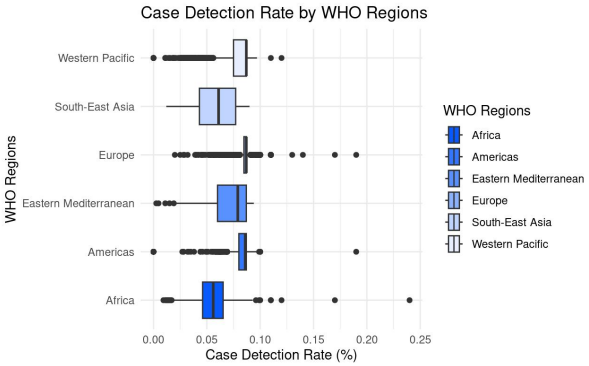


Figure 2 Case Detection Rate by WHO Regions Boxplot

Although both graphs contain multiple outliers, the mortality boxplot shows many more extreme values with higher mortality, indicating the right skewness of the data. Such observations suggest non-homogeneity of regional outcomes. Specifically, the mortality medians are the highest in Africa (.056) and South-East Asia (.036), whereas Europe (.0013) and the Americas (.0025) display very low medians, highlighting their relatively low burden (Figure 1 & Appendix 1). On the other hand, the case detection boxplot

shows an oppposite pattern. Africa (0.056) and South-East Asia (0.061) have the lowest median detection coverage, while Europe and the Americas have relatively higher medians of 0.087 and 0.086 (Figure 2 & Appendix 2).

These opposite trends imply an inverse association between treatment coverage and fatal outcomes and give us impetus to the subsequent regression analysis.

Difference in Mortality Means

To test whether mortality differs among WHO regions, we conducted a one-way ANOVA on untransformed mortality data.

$$H_0: \mu_{Afr.} = \mu_{Ame.} = \mu_{E.Med.} = \mu_{Eur.} = \mu_{S.E.Asia} = \mu_{W.Pac.}$$

$$H_A: \text{at least one } \mu_i \neq \mu_j$$

At a significance level of $\alpha = 0.05$, the model yields a p-value $< 2 \times 10^{-16}$, suggesting that at least two regional means differ significantly ($F = 518.4$) (Appendix 3).

Since ANOVA does not indicate which particular groups differ from one another significantly, we conducted two pairwise t-tests, once with Holm adjustment and once using Tukey's Honest Significant Difference test. According to Holm-adjusted tests, almost all regional comparisons are statistically significant, with only Europe-Americas and Western Pacific-Eastern Mediterranean being non-significant at family-wise $\alpha = 0.05$. On the other hand, all comparisons involving Africa and Southeast Asia yield p-value $< 2 \times 10^{-16}$, implying that these two regions differs the most in mortality patterns among all WHO regions (Appendix 4).

The Tukey test yields similar substantive results: Europe-Americas' mean difference is insignificant with a p-value > 0.05,whereas Africa and Southeast Asia are the only regions whose means differ significantly from all other regional means in mortality (p-value < 0.001), and together with the boxplot in Figure 1a, we can conclude that their mortality means are exceptionally higher than all the rest (Table 1).

	Comparison <chr>	p_value <dbl>	Classification <chr>
Americas-Africa	Americas-Africa	0.000000e+00	Significant
Eastern Mediterranean-Africa	Eastern Mediterranean-Africa	0.000000e+00	Significant
Europe-Africa	Europe-Africa	0.000000e+00	Significant
South-East Asia-Africa	South-East Asia-Africa	0.000000e+00	Significant
Western Pacific-Africa	Western Pacific-Africa	0.000000e+00	Significant
Eastern Mediterranean-Americas	Eastern Mediterranean-Americas	6.618249e-03	Not Significant
Europe-Americas	Europe-Americas	9.999972e-01	Not Significant
South-East Asia-Americas	South-East Asia-Americas	0.000000e+00	Significant
Western Pacific-Americas	Western Pacific-Americas	1.053519e-05	Not Significant
Europe-Eastern Mediterranean	Europe-Eastern Mediterranean	3.185587e-03	Not Significant
South-East Asia-Eastern Mediterranean	South-East Asia-Eastern Mediterranean	0.000000e+00	Significant
Western Pacific-Eastern Mediterranean	Western Pacific-Eastern Mediterranean	9.778816e-01	Not Significant
South-East Asia-Europe	South-East Asia-Europe	0.000000e+00	Significant
Western Pacific-Europe	Western Pacific-Europe	1.941207e-06	Not Significant
Western Pacific-South-East Asia	Western Pacific-South-East Asia	0.000000e+00	Significant

Table 1 Pairwise comparisons of regional TB mortality with Turkey test. Significance assigned when p-value <0.001

Relationship of Mortality and Detection Rate

Finally, we modeled the linear regression between two quantitative variables to assess whether mortality is associated with case detection rates. The fitted line in the initial model with raw values suggests a negative association with a slope of -1.47, and the overall regression p-value $< 2.2 \times 10^{-16}$ (Figure 3 & Appendix 5d).

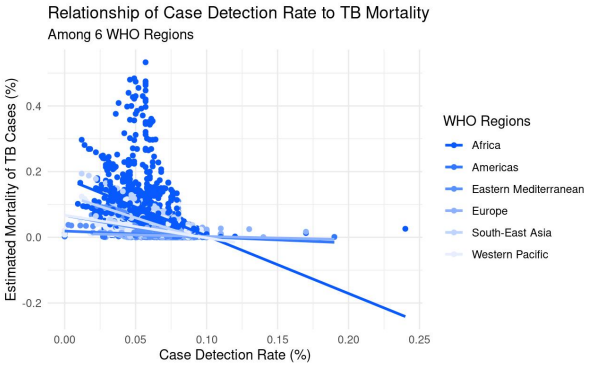


Figure 3 Case Detection Rate to Mortality Raw Regression Model

However, diagnostic checks indicate that conditions suitable for linear inference are violated: the residual plot shows a pattern of decreasing variability (heteroscedasticity), and the two variables are both heavily skewed, with mortality right skewed and detection rate left skewed (Appendix 5a, 5b, 5c). Such conditions undermine the validity of inference from the untransformed model, and thus we conducted a logarithmic transformation to justify the negative association.

After log-transforming both mortality and detection rate, the distributions become approximately normal and the residual plot no longer displays a funnel pattern (Appendix 6a, 6b, 6c). The transformed regression also indicates a negative association between detection rate (predictor) and mortality (response). The slope

estimate is substantial (p-value $< 2.2 \times 10^{-16}$), with much less variability around the fitted line. The regression equation and the measure of variance are shown below (Appendix 6d):

$$\log \widehat{Mortality} = -15.17 - 3.75 \times \log Detection\ Rate$$

$$R^2 = 0.41$$

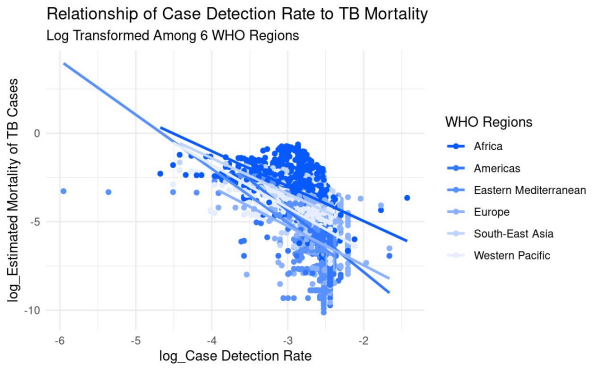


Figure 4 Case Detection Rate to Mortality Logged Regression Model

This means that countries with higher detection rates tend to have lower mortality, and this association holds after accounting for skewness and non-constant variance.

In summary, the data analysis supports two main findings: mortality differs across world regions, where Europe & the Americas being the best performers and Africa & Southeast Asia being the poorest performers; secondly, high TB detection coverage is strongly related to low mortality. These results support the global health agenda on increasing access to diagnosis and treatment delivery in high-burden regions where coverage and survival outcomes remain poor.

Results and Conclusions

Our analysis show clear regional inequalities in TB outcomes. Mortality rates differ significantly across WHO regions (ANOVA: F = 518.4, p-value $< 2 \times 10^{-16}$), with pairwise t-tests confirming that Europe and the Americas having similarly lower mortality while Africa and South-East Asia having significantly higher mortality (all p-values < 0.001). Logged regression results further show a strong negative relationship between detection and mortality, producing a slope of –3.75 justified by the p-value $< 2.2 \times 10^{-16}$ and $R^2 = 0.41$.

Together, the analysis provides evidence of large regional inequalities in TB outcomes and demonstrates that regions with lower diagnostic coverage, specifically Africa and South-East Asia, experience significantly higher mortality, and the opposite holds true as well for Europe and the Americas. Furthermore, the strong negative association between detection and mortality suggests that improving diagnostic access and treatment coverage may be helpful to reduce TB deaths in affected areas. These findings highlight the importance of strengthening healthcare systems and encourage WHO interventions to support identified high-burden areas to reduce preventable mortality worldwide.

Limitations

Certain limitations must be considered when interpreting the results. First, since the data is from an observational study, causality cannot be inferred. The relationships found, such as the relationship between detection and mortality, may be confounded by external factors such as HIV prevalence or socioeconomic conditions. Thus, our findings should be interpreted as correlations only.

Moreover, some observations are excluded in analysis due to missing values, which may introduce bias if some regions are more prone to missing data. In addition, each WHO region consists of countries with considerable internal variation. Grouping dozens of countries into six regional groups simplifies the analysis and hides differences at the country level. Hence, the results and regional averages may not accurately reflect individual countries' circumstances.

Finally, the mortality data used in this analysis are estimates rather than direct measurements. They rely heavily on assumptions and incomplete national reporting, especially in settings with weak surveillance systems. Such measurement uncertainty may affect the precision of the statistical results.

References

WHO. “2.2 TB Mortality.” *Who.int*, 2019, www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports/global-tuberculosis-report-2022/tb-disease-burden/2-2-tb-mortality.

Appendix 1. Medians for Mortality

```
              Africa               Americas Eastern Mediterranean                 Europe
              0.05600                 0.00245                0.00125
0.00115
      South-East Asia       Western Pacific
              0.03600                 0.00530
```

Appendix 2. Medians for Case Detection Rate

```
              Africa               Americas Eastern Mediterranean                 Europe
                0.056                   0.086                  0.079
0.087
      South-East Asia       Western Pacific
                0.061                   0.087
```

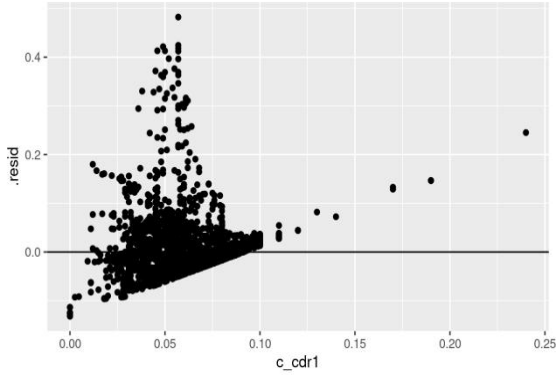
Appendix 3. ANOVA Results

```
             Df Sum Sq Mean Sq F value Pr(>F)
g_whoregion    5  4.876  0.9752   518.4 <2e-16 ***
Residuals   5087  9.570  0.0019
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
24 observations deleted due to missingness
```

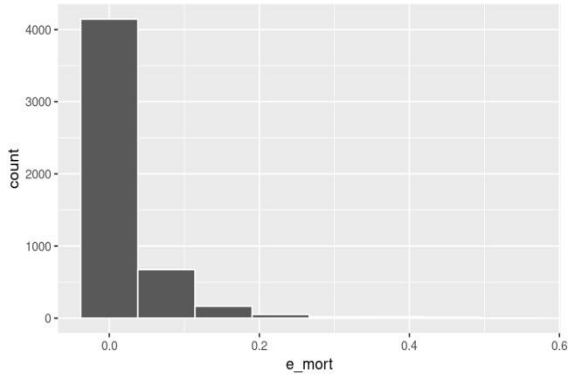
Appendix 4. Pairwise t-tests with pooled SD and Holm Adjustment

```
       Pairwise comparisons using t tests with pooled SD

data:  who_tb_data$e_mort and who_tb_data$g_whoregion

                    Africa  Americas Eastern Mediterranean Europe  South-East
Asia
Americas            < 2e-16 -        -                     -       -
Eastern Mediterranean < 2e-16 0.00150  -                     -       -
Europe              < 2e-16 0.92716  0.00093               -       -
South-East Asia     < 2e-16 < 2e-16  < 2e-16               < 2e-16 -
Western Pacific     < 2e-16 3.6e-06  0.92716               7.8e-07 < 2e-16

P-value adjustment method: Holm
```

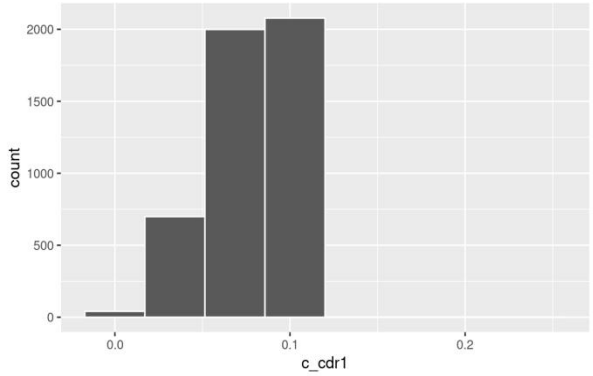
Appendix 5A. Residual Plot of Case Detection Rate v.s. Mortality with Raw Data



Appendix 5b. Histogram of Raw Mortality (Right-skewed)



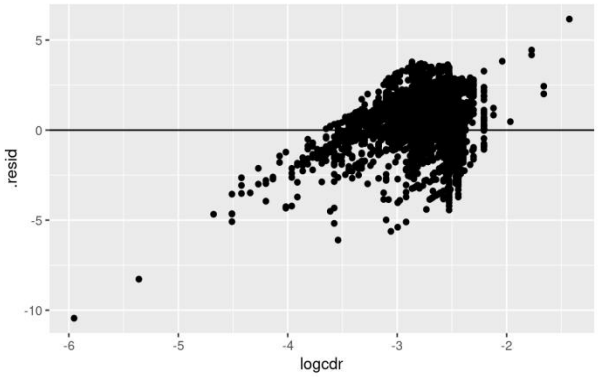
Appendix 5c. Histogram of Raw Case Detection Rate (Left-skewed)



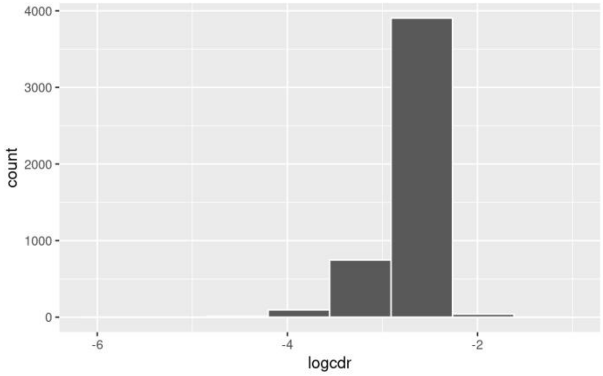
Appendix 5d. Numeric Values of Raw Regression

```
Call:
lm(formula = e_mort ~ c_cdr1, data = who_tb_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.13188 -0.01342 -0.00579 -0.00015  0.48235

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.134682   0.002596   51.88   <2e-16 ***
c_cdr1      -1.474260   0.034180  -43.13   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04598 on 4798 degrees of freedom
  (317 observations deleted due to missingness)
Multiple R-squared:  0.2794,      Adjusted R-squared:  0.2793
F-statistic:  1860 on 1 and 4798 DF,  p-value: < 2.2e-16
```

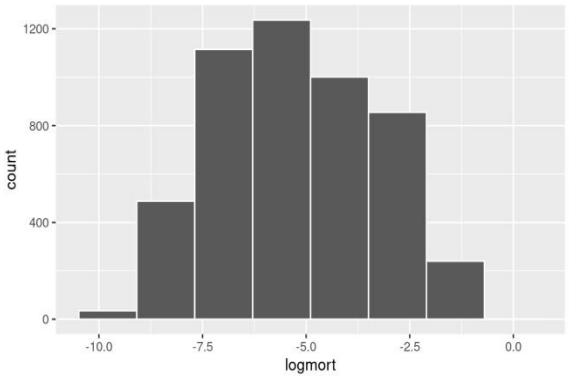
Appendix 6a. Residual Plot of Log-transformed Regression



Appendix 6b. Histogram of Logged Case Detection Rate



Appendix 6c. Histogram of Logged Mortality Rate



Appendix 6d. Numeric Values of Logged Regression

```
Call:
lm(formula = logmort ~ logcdr, data = who_tb_data)

Residuals:
     Min       1Q   Median       3Q      Max
-10.4453  -1.0856   0.0736   1.0037   6.1652

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -15.1731     0.1726  -87.89   <2e-16 ***
logcdr       -3.7546     0.0646  -58.12   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.472 on 4770 degrees of freedom
  (345 observations deleted due to missingness)
Multiple R-squared:  0.4146,      Adjusted R-squared:  0.4145
F-statistic:  3378 on 1 and 4770 DF,  p-value: < 2.2e-16
```