

Drug resistance Tuberculosis

Uthman Olalekan Al-ameen

2025-06-05

Load required packages

Load Raw Datasets

```
library(readr)
TB_drug_resistance_Estimate <- read_csv("C:/Users/THIS PC/Downloads/TB drug resistance Estimate.csv")
```

```
## Rows: 9 Columns: 34
## -- Column specification -----
## Delimiter: ","
## chr  (11): IndicatorCode, Indicator, ValueType, ParentLocationCode, ParentLo...
## dbl  (4): Period, FactValueNumeric, FactValueNumericLow, FactValueNumericHigh
## lgl  (18): IsLatestYear, Dim1 type, Dim1, Dim1ValueCode, Dim2 type, Dim2, Di...
## dtm  (1): DateModified
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
library(readr)
TB_cases_started_on_MDR_RR_TB_treatment <- read_csv("C:/Users/THIS PC/Downloads/TB cases started on MDR,
```

```
## Rows: 16 Columns: 34
## -- Column specification -----
## Delimiter: ","
## chr  (10): IndicatorCode, Indicator, ValueType, ParentLocationCode, ParentLo...
## dbl  (3): Period, FactValueNumeric, Value
## lgl  (20): IsLatestYear, Dim1 type, Dim1, Dim1ValueCode, Dim2 type, Dim2, Di...
## dtm  (1): DateModified
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
library(readr)
TB_Drug_resistance_confimed_cases <- read_csv("C:/Users/THIS PC/Downloads/TB Drug resistance confimed c
```

```
## Rows: 17 Columns: 34
## -- Column specification -----
## Delimiter: ","
```

```
## chr (10): IndicatorCode, Indicator, ValueType, ParentLocationCode, ParentLo...
## dbl (3): Period, FactValueNumeric, Value
## lgl (20): IsLatestYear, Dim1 type, Dim1, Dim1ValueCode, Dim2 type, Dim2, Di...
## dtm (1): DateModified
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

confirmed cases of multi drug Resistance tuberculosis

```
# Filter and rename for confirmed cases of MDR-TB (2015-2023)
confirmed_cases_of_MDR_TB <- TB_Drug_resistance_confimed_cases %>%
  filter(Period >= 2015 & Period <= 2023) %>%
  select(Year = Period, Confirmed_Cases = FactValueNumeric)

# View the cleaned dataset
flextable(confirmed_cases_of_MDR_TB)
```

```
## Warning: fonts used in 'flextable' are ignored because the 'pdflatex' engine is
## used and not 'xelatex' or 'lualatex'. You can avoid this warning by using the
## 'set_flextable_defaults(fonts_ignore=TRUE)' command or use a compatible engine
## by defining 'latex_engine: xelatex' in the YAML header of the R Markdown
## document.
```

Year Confirmed_Cases	
2,023	3,642
2,022	3,932
2,021	2,975
2,020	2,061
2,019	2,384
2,018	2,275
2,017	2,286
2,016	1,686
2,015	1,241

Estimated number of Muti drug resistance tuberculosis

```
# Filter and select actual estimated values (FactValueNumeric) from 2015 to 2023
estimated_number_of_MDR_TB_cases <- TB_drug_resistance_Estimate %>%
  filter(
    Period >= 2015 & Period <= 2023,
    IndicatorCode == "TB_e_inc_rr_num", # To ensure it's the right indicator
    SpatialDimValueCode == "NGA" # Nigeria
  ) %>%
```

```

select(Year = Period, Estimated_Cases = FactValueNumeric)

# View the cleaned dataset
flextable (estimated_number_of_MDR_TB_cases)

```

```

## Warning: fonts used in 'flextable' are ignored because the 'pdflatex' engine is
## used and not 'xelatex' or 'lualatex'. You can avoid this warning by using the
## 'set_flextable_defaults(fonts_ignore=TRUE)' command or use a compatible engine
## by defining 'latex_engine: xelatex' in the YAML header of the R Markdown
## document.

```

Year	Estimated_Cases
2,023	9,400
2,022	10,000
2,021	12,000
2,020	13,000
2,019	14,000
2,018	16,000
2,017	18,000
2,016	20,000
2,015	22,000

Cases of Multidrug resistance tuberculosis started on treatment

```

# Step 2: Rename the dataset
cases_started_on_MDR_TB_treatment <- TB_cases_started_on_MDR_RR_TB_treatment

# Step 3: Filter for years 2015 to 2023 and select Year and Value columns
cases_started_on_MDR_TB_treatment <- cases_started_on_MDR_TB_treatment[cases_started_on_MDR_TB_treatment$Year %in% 2015:2023,]

# Step 4: View the result
flextable (cases_started_on_MDR_TB_treatment)

```

```

## Warning: fonts used in 'flextable' are ignored because the 'pdflatex' engine is
## used and not 'xelatex' or 'lualatex'. You can avoid this warning by using the
## 'set_flextable_defaults(fonts_ignore=TRUE)' command or use a compatible engine
## by defining 'latex_engine: xelatex' in the YAML header of the R Markdown
## document.

```

Period	Value
2,023	3,047
2,022	3,185

Period	Value
2,021	2,197
2,020	1,584
2,019	1,975
2,018	1,895
2,017	1,786
2,016	1,251
2,015	656

Combine All 3 Datasets

```
# Ensure column names are consistent before merging
cases_started_on_MDR_TB_treatment <- cases_started_on_MDR_TB_treatment %>%
  rename(Year = Period, Started_Treatment = Value)

# Merge all three datasets by 'Year'
combined_tb_data <- confirmed_cases_of_MDR_TB %>%
  left_join(estimated_number_of_MDR_TB_cases, by = "Year") %>%
  left_join(cases_started_on_MDR_TB_treatment, by = "Year")
```

Print combined_tb_data

```
flextable(combined_tb_data)
```

```
## Warning: fonts used in 'flextable' are ignored because the 'pdflatex' engine is
## used and not 'xelatex' or 'lualatex'. You can avoid this warning by using the
## 'set_flextable_defaults(fonts_ignore=TRUE)' command or use a compatible engine
## by defining 'latex_engine: xelatex' in the YAML header of the R Markdown
## document.
```

Year	Confirmed_Cases	Estimated_Cases	Started_Treatment
2,023	3,642	9,400	3,047
2,022	3,932	10,000	3,185
2,021	2,975	12,000	2,197
2,020	2,061	13,000	1,584
2,019	2,384	14,000	1,975
2,018	2,275	16,000	1,895
2,017	2,286	18,000	1,786
2,016	1,686	20,000	1,251

Year	Confirmed_Cases	Estimated_Cases	Started_Treatment
2,015	1,241	22,000	656

combined_tb_data output

```
# Create the flextable from your data
ft <- flextable(combined_tb_data)

# Save it to a Word document
save_as_docx("Combined MDR TB Data Table" = ft, path = "combined_mdr_tb_data_table.docx")
```

Summary of the combined_tb_data

```
summary(combined_tb_data)
```

```
##      Year      Confirmed_Cases Estimated_Cases Started_Treatment
## Min.   :2015      Min.   :1241      Min.   : 9400      Min.   : 656
## 1st Qu.:2017      1st Qu.:2061      1st Qu.:12000      1st Qu.:1584
## Median :2019      Median :2286      Median :14000      Median :1895
## Mean   :2019      Mean   :2498      Mean   :14933      Mean   :1953
## 3rd Qu.:2021      3rd Qu.:2975      3rd Qu.:18000      3rd Qu.:2197
## Max.   :2023      Max.   :3932      Max.   :22000      Max.   :3185
```

Convert to Long Format

```
# Convert wide to long
combined_tb_data_long <- combined_tb_data %>%
  pivot_longer(
    cols = c(Confirmed_Cases, Estimated_Cases, Started_Treatment),
    names_to = "Variable",
    values_to = "Value")
```

Percentage of treated and confirmed

```
# Calculate derived percentages
derived_percentages <- combined_tb_data %>%
  mutate(
    Percentage_Confirmed = round((Confirmed_Cases / Estimated_Cases) * 100, 1),
    Percentage_Treated_of_Confirmed = round((Started_Treatment / Confirmed_Cases) * 100, 1)
  ) %>%
  select(Year, Percentage_Confirmed, Percentage_Treated_of_Confirmed)
```

Describe percentages

```
summary (derived_percentages)
```

```
##      Year      Percentage_Confirmed Percentage_Treated_of_Confirmed
## Min.   :2015   Min.    : 5.60      Min.    :52.9
## 1st Qu.:2017   1st Qu.:12.70      1st Qu.:74.2
## Median :2019   Median :15.90      Median :78.1
## Mean   :2019   Mean    :19.62      Mean    :76.3
## 3rd Qu.:2021   3rd Qu.:24.80      3rd Qu.:82.8
## Max.    :2023   Max.    :39.30      Max.    :83.7
```

Flextable of the percentage table

```
flextable(derived_percentages)
```

```
## Warning: fonts used in 'flextable' are ignored because the 'pdflatex' engine is
## used and not 'xelatex' or 'lualatex'. You can avoid this warning by using the
## 'set_flextable_defaults(fonts_ignore=TRUE)' command or use a compatible engine
## by defining 'latex_engine: xelatex' in the YAML header of the R Markdown
## document.
```

Year	Percentage_Confirmed	Percentage_Treated_of_Confirmed
2,023	38.7	83.7
2,022	39.3	81.0
2,021	24.8	73.8
2,020	15.9	76.9
2,019	17.0	82.8
2,018	14.2	83.3
2,017	12.7	78.1
2,016	8.4	74.2
2,015	5.6	52.9

. Chart 1: Line chart for Estimated, Confirmed, and Treated cases (raw counts)

```
# Convert combined_tb_data to long format for easier plotting
data_long_counts <- combined_tb_data %>%
  pivot_longer(cols = c(Estimated_Cases, Confirmed_Cases, Started_Treatment),
               names_to = "Case_Type",
               values_to = "Count")

ggplot(data_long_counts, aes(x = Year, y = Count, color = Case_Type)) +
```

```

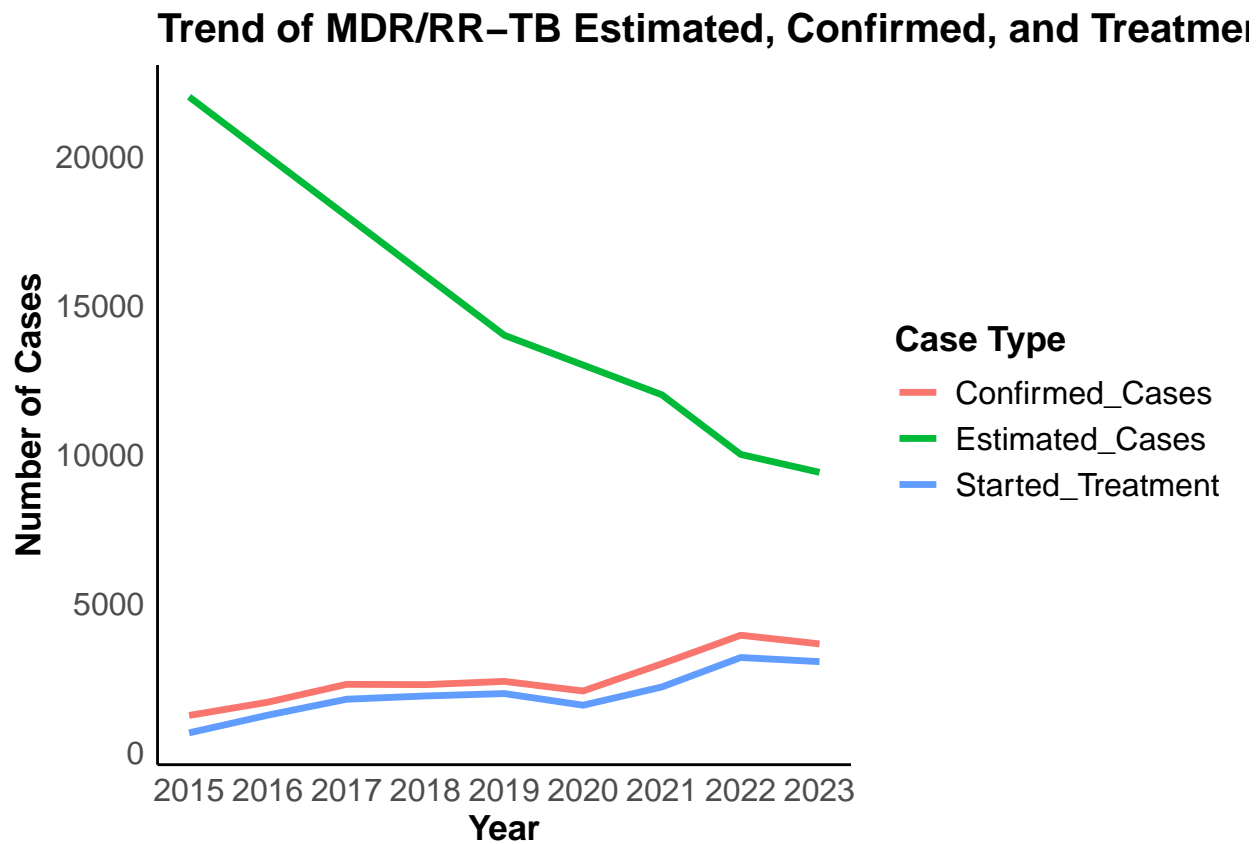
geom_line(size = 1.2) +
labs(
  title = "Trend of MDR/RR-TB Estimated, Confirmed, and Treatment-Started Cases in Nigeria (2015-2023)",
  y = "Number of Cases",
  color = "Case Type"
) +
scale_x_continuous(breaks = 2015:2023) +
theme_minimal() +
theme(
  panel.grid = element_blank(),
  axis.line = element_line(color = "black"),
  plot.title = element_text(face = "bold", size = 15),
  axis.title = element_text(face = "bold", size = 13),
  axis.text = element_text(size = 12),           # Axis tick labels
  legend.title = element_text(size = 13, face = "bold"),
  legend.text = element_text(size = 12)         # Legend labels
)

```

```

## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

```



```

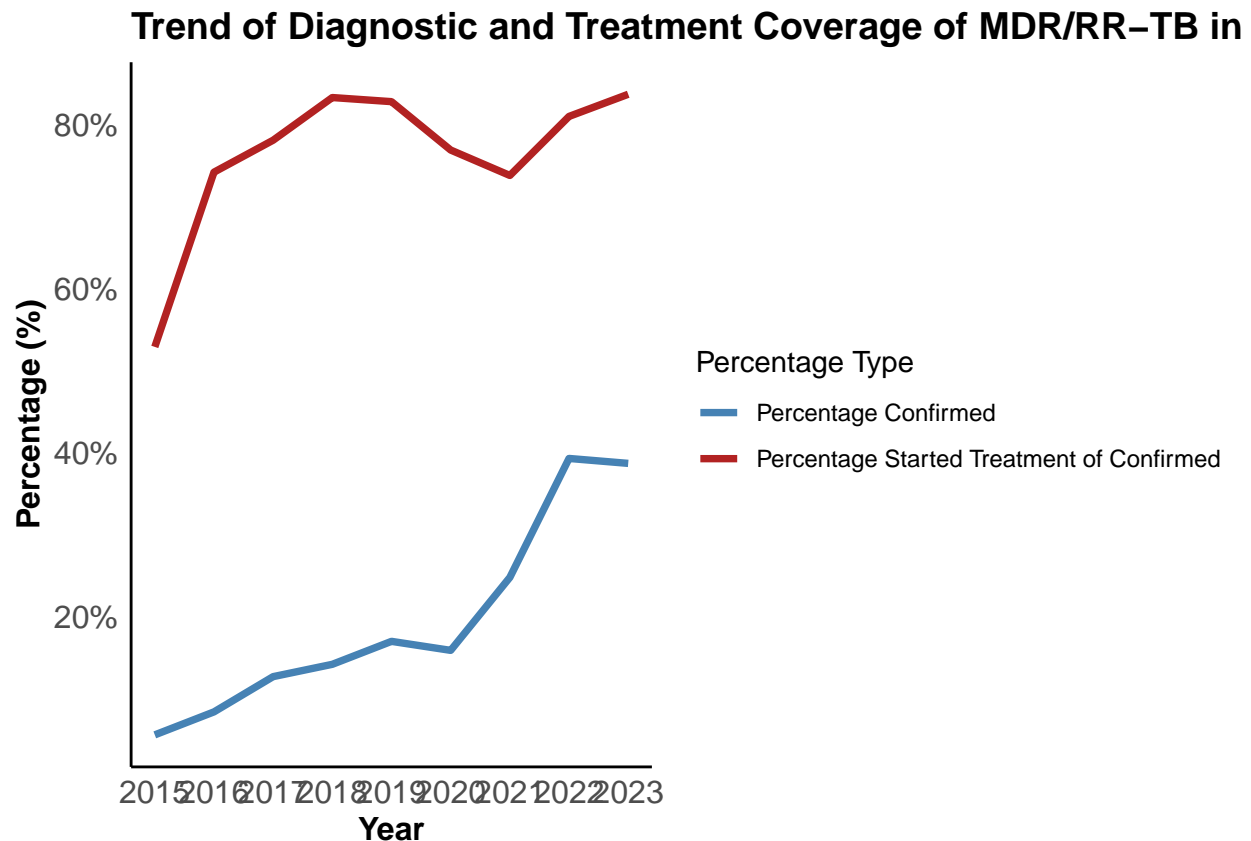
## Convert derived percentages to long format for plotting

```

```
# Convert derived percentages to long format for plotting
data_long_percent <- derived_percentages %>%
  pivot_longer(
    cols = c(Percentage_Confirmed, Percentage_Treated_of_Confirmed),
    names_to = "Percentage_Type",
    values_to = "Percentage"
  )
```

2. Chart 2: Line chart for % Confirmed and % Treated of Confirmed (percent-ages)

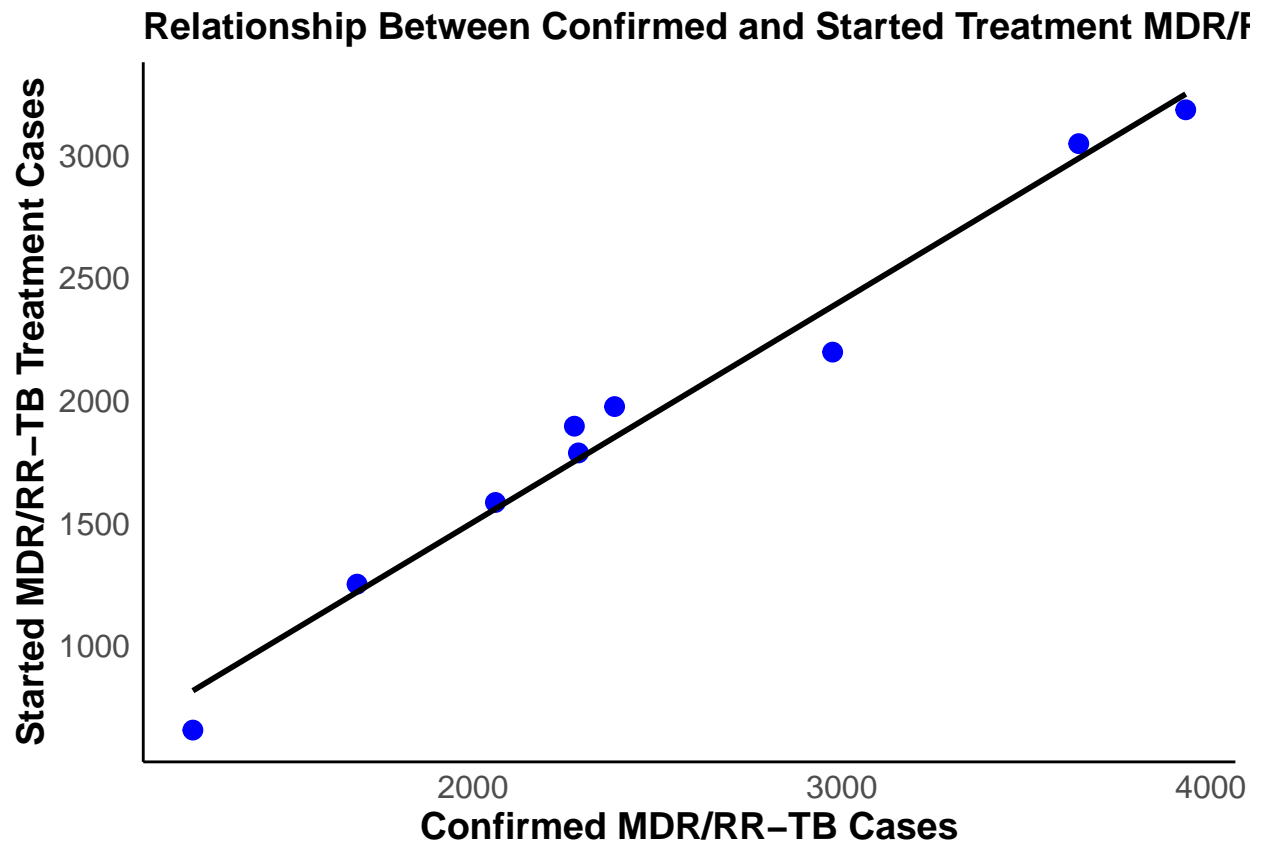
```
ggplot(data_long_percent, aes(x = Year, y = Percentage, color = Percentage_Type)) +
  geom_line(size = 1.3) +
  labs(
    title = "Trend of Diagnostic and Treatment Coverage of MDR/RR-TB in Nigeria (2015-2023)",
    y = "Percentage (%)",
    color = "Percentage Type"
  ) +
  theme_minimal() +
  scale_x_continuous(breaks = 2015:2023) +
  scale_y_continuous(labels = percent_format(scale = 1)) +
  scale_color_manual(
    values = c("Percentage_Confirmed" = "steelblue",
              "Percentage_Treated_of_Confirmed" = "firebrick"),
    labels = c("Percentage_Confirmed" = "Percentage Confirmed",
              "Percentage_Treated_of_Confirmed" = "Percentage Started Treatment of Confirmed")
  ) +
  theme(
    panel.grid = element_blank(),
    axis.line = element_line(color = "black"),
    plot.title = element_text(face = "bold", size = 14),
    axis.title = element_text(face = "bold", size = 12),
    axis.text = element_text(size = 12)
  )
```

Scatter Plot: Confirmed cases (X) vs. Treated cases (Y)

```
ggplot(combined_tb_data, aes(x = Confirmed_Cases, y = Started_Treatment)) +
  geom_point(color = "blue", size = 3) +
  geom_smooth(method = "lm", se = FALSE, color = "black", linetype = "solid") +
  labs(
    title = "Relationship Between Confirmed and Started Treatment MDR/RR-TB Cases in Nigeria",
    x = "Confirmed MDR/RR-TB Cases",
    y = "Started MDR/RR-TB Treatment Cases"
  ) +
  theme_minimal() +
  theme(
    panel.grid = element_blank(),
    axis.line = element_line(color = "black"),
    plot.title = element_text(face = "bold", size = 14),
    axis.title = element_text(face = "bold", size = 14), # Bold and increase axis titles
    axis.text = element_text(size = 12) # Increase axis scale (tick labels)
  )
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



```
# Ensure variables are numeric
combined_tb_data <- combined_tb_data %>%
  mutate(
    Confirmed_Cases = as.numeric(Confirmed_Cases),
    Started_Treatment = as.numeric(Started_Treatment)
  )
```

— 1. Pearson Correlation between Confirmed Cases and Treated Cases —

```
correlation_result <- cor.test(combined_tb_data$Confirmed_Cases, combined_tb_data$Started_Treatment)
cat("Pearson Correlation Test Results:\n")
```

```
## Pearson Correlation Test Results:
```

```
print(correlation_result)
```

```
##
## Pearson's product-moment correlation
##
## data: combined_tb_data$Confirmed_Cases and combined_tb_data$Started_Treatment
## t = 18.091, df = 7, p-value = 3.899e-07
```

```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9489153 0.9978667
## sample estimates:
##      cor
## 0.9894747
```

— 2. Simple Linear Regression: Predict Treated Cases from Confirmed Cases

```
lm_model <- lm(Started_Treatment ~ Confirmed_Cases, data = combined_tb_data)

cat("\nSimple Linear Regression Summary:\n")
```

```
##
## Simple Linear Regression Summary:
```

```
summary(lm_model)
```

```
##
## Call:
## lm(formula = Started_Treatment ~ Confirmed_Cases, data = combined_tb_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -187.09  -64.21   26.16   59.95  143.70
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -305.27809   131.45772  -2.322   0.0532 .
## Confirmed_Cases    0.90399    0.04997  18.091 3.9e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 123.7 on 7 degrees of freedom
## Multiple R-squared:  0.9791, Adjusted R-squared:  0.9761
## F-statistic: 327.3 on 1 and 7 DF, p-value: 3.899e-07
```