# Statistical Analysis of Government Subsidy and OOPE

**R. Advaith[1], J. Venkatesh[2]**
*[1,2]Vellore Institute of Technology, Chennai Campus, SENSE
advaith.r2023@vitstudent.ac.in; **23BEC1205; 8657994301**[1]
venkatesh.j2023@vitstudent.ac.in; **23BEC1231; 9445639588**[2]

## ABSTRACT

This project explores the effectiveness of medical subsidies provided by the government of India by studying the Out-Of-Pocket Expenditure (OOPE) of India and comparing the yearly trends with other countries. The primary data was collected by circulating a survey form to students at the Vellore Institute of Technology, Chennai Campus. The global data was collected from various national health surveys and annual government health reports. Using statistical concepts of linear regression, hypothesis testing and visual data analysis, we could draw several conclusions regarding both the immediate and the overall impact of subsidised medical rates. The data is analysed using several statistical models, all executed on RStudio, an integrated development environment for R. With the range of data, we could also make disease-wise effect by subsidy using our primary data and the improvement of overall government funding and measures medical measures taken to reduce OOPE. The global data has been normalised, by considering the currencies, inflations and so on. R has proven to be functional and beneficial for all the statistical analysis required for this project, primarily due to its ease of use and excellent visualization tools.

**Keywords:** Government, Health Reports, OOPE, Subsidy, Survey and Visual Data Analysis

## ACKNOWLEDGEMENT

We extend our sincere gratitude to Dr. Uma Maheswari for her invaluable guidance and support throughout the course of our project on the effectiveness of government medical subsidies and their impact on Out-Of-Pocket Expenditure (OOPE) in India. Her expertise and insight were instrumental in helping us navigate complex statistical analyses and understand the broader implications of healthcare funding policies. We are also grateful to our class members for helping us in collecting the primary data.

# I. INTRODUCTION

Medicines and the availability of healthcare is one of the defining aspects of societal growth. Several important studies are constantly being made to make healthcare more accessible and affordable.

Several private and intergovernmental organisations, like WHO (World Health Organization)[20] not only focus on the scientific aspects of medicinal growth, several of their international surveys, meetings and research are dedicated to serve the common people and ensure that healthcare is ubiquitous and available for all. WHO specifically has been doing this since 1948.

If we look at India, annual health reports are being made to study and compare the government schemes implemented and several defining parameters which indicate the economic changes in pharmaceutical services. These reports are available in government websites. One of the websites dedicated to such reports is the NHSRC (National Health Systems Resource Centre), where the annual NHA (National Health Accounts)[21] reports have been made available on the internet. However, the latest account is of the financial year of 2021-22. While collecting the global data, the data for OOPE in the WHO websites have only been collected till 2021. This gap in data needed to be addressed so that we can analyse the trends of OOPE, whether the expenditure is declining or increasing, which in turn indicates the effectiveness of medicinal subsidy.

The primary data collected from students also focuses on how many people have used government subsidised pharmacies and got benefit out of it. This study also underlines the extent of awareness among public and the ease of availability of such shops when in need.

# II. LITERATURE REVIEW

*Healthcare Spending and Disease Burden*

Smith and Nguyen (2020)[1] demonstrated that countries prioritizing health spending in high-prevalence conditions reduce overall morbidity and healthcare costs. Similarly, the Global Burden of Disease study (2019)[2] analysed healthcare spending efficiency across various diseases, showing that chronic conditions like diabetes and cardiovascular disease, which have high prevalence, benefit most from increased budget allocation. Auster et al. (2019)[3] found that expenditure alignment with disease burden can substantially decrease health inequities by focusing on prevalent and high-cost conditions.

*Government Subsidies and Equitable Access*

The study by Cleary et al. (2018)[4] highlights the role of subsidies in improving healthcare access for low-income groups, noting that region-specific subsidies reduce healthcare disparities. Research by Papanicolas et al. (2017)[5] further confirmed that subsidized healthcare services help mitigate out-of-pocket expenses for chronic diseases, especially when subsidies are directed at preventive care. In a similar vein, Miller et al. (2021)[6] found that healthcare access in rural and low-income areas improves significantly when government spending is disease-targeted, emphasizing the importance of targeted subsidies for high-prevalence diseases.

*Economic Factors in Healthcare Funding*

Macinko and Starfield (2018)[7] explored how healthcare expenditure efficiency varies with economic factors like GDP and healthcare infrastructure quality. Their findings suggest that economic context should guide healthcare funding to maximize returns, particularly in low-income regions with limited resources. Additionally, Kutzin et al. (2020)[8] found that aligning subsidy distribution with economic indicators, such as GDP and healthcare infrastructure, helps ensure sustainable healthcare funding. Economic analyses by Hsiao (2019)[10] demonstrated that subsidies targeting preventive measures are economically advantageous, as they reduce long-term healthcare costs in populations with high disease prevalence.

*Correlation Between Subsidy Allocation and Health Outcomes*

A review by Bradley et al. (2021)[11] established a positive correlation between targeted government subsidies and health outcomes in high-prevalence areas. Their analysis shows that well-allocated subsidies reduce disease incidence by improving preventive care access, especially in low-income regions. Similarly, analysis by Wagstaff and Lindelow (2019)[12] revealed that disease prevalence and healthcare subsidies are strongly correlated, with targeted spending yielding better outcomes in treating infectious and chronic diseases. This aligns with insights from Pauly (2020)[13], who recommended reallocating health funds toward disease-targeted subsidies for high-prevalence conditions.

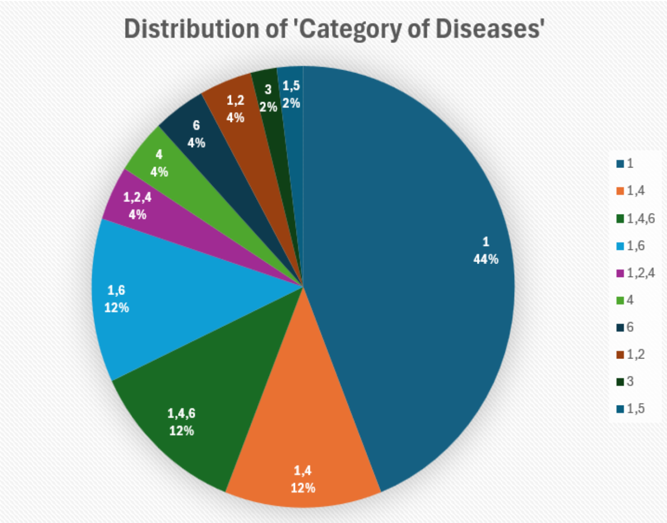
## III. DATASET AND MODEL FORMATION

***Primary Data Collection and Analysis***

The primary data[*1] was collected by circulating a Google Forms[*] to students at VIT Chennai. We obtained a total of 50 responses. The survey form contained three questions; the category of diseases them and their families are most affected by, the average medical expenditure per month without subsidy, and if they have visited government subsidised pharmacies and if so, what was their revised average medical expenditure per month. An overview of the data is given below:



*The link to the Google Form used in the survey is: https://forms.gle/svxML7sYevG4ePiX6




*1 – Common Influenza (Cold, Flu, Viral Fever)*
*2 – Poisoning (Food and Water)*
*3 – Respiratory Diseases (Asthma, Bronchitis etc.)*
*4 – Diabetes*
*5 – Oncological Diseases (Cancer)*
*6 – Others*

CHART I
PIE CHART OF DATA CATEGORISED BASED ON THE DISEASES CHOSEN

The primary data is analysed by using scatter plot and paired-T hypothesis testing at 5% level of significance, and 9 degrees of freedom, as the sample number of categories chosen is 10 (refer pie chart). The hypothesis testing is done by calculating the mean of the average medical expenditure before and after subsidy, categorised by the diseases chosen by the respondents. The code for both the analysis is given below:

*Scatter Plot*

```
library(readxl)  # For reading Excel files
library(ggplot2) # For plotting
# Importing the Excel file with the Primary Data
data <- read_excel("D:/R Code Project/Local_Data_GForms.xlsx")

# Selecting the 3rd and 4th columns for plotting as it contains the average medical expenditure data before and after subsidy
plot_data <- data[, c(3, 4)]

# Check the column names, for convenience
names(plot_data) <- c("x", "y")

# Create a scatter plot with ggplot2
ggplot(plot_data, aes(x = x, y = y, color = y)) +
  geom_point() + scale_color_gradient(low = "blue", high = "red") +
  labs(x = "Average Medical Expenditure Before Subsidy", y = "Average Medical Expenditure After Subsidy") +
  theme_minimal()
```

*Paired-T Hypothesis Testing*

```
library(readxl)  # For reading Excel files
library(dplyr)   # For data manipulation

# Importing the Excel file of the primary data
data <- read_excel("D:/R Code Project/Local_Data_GForms.xlsx")

# Calculate the average values of columns 3 and 4, grouped by the categories in column 2
names(data)[2:4] <- c("Category", "Column3", "Column4")

# Calculate the average values of Column3 and Column4, grouped by Category
```

```
average_table <- data %>%
  group_by(Category) %>%                   #
Group by the Category column
  summarise(
    Avg_Column3 = mean(Column3, na.rm =
TRUE) ,
    Avg_Column4 = mean(Column4, na.rm =
TRUE)
  )

# View the result
print(average_table)

# Perform Paired t-test
t_test_result <-
t.test(average_table$Avg_Column3,
average_table$Avg_Column4, paired =
TRUE)

# Extract test statistics
t_value <- t_test_result$statistic  #
Calculated t-value
df <- t_test_result$parameter       # Degrees
of freedom

# Determine the critical t-value at 5%
significance level (one-tailed) with 9
degrees of freedom
alpha <- 0.05
critical_t_value <- qt(1 - alpha, df)  # One-
tailed test

# Interpretation
if (abs(t_value) > abs(critical_t_value)) {
  conclusion <- "Reject the null hypothesis
(significant difference)."
} else {
  conclusion <- "Fail to reject the null
hypothesis (no significant difference)."
}

# Print the result
cat("Calculated t-value:", t_value, "\n")
cat("Critical t-value:", critical_t_value, "\n")
cat("Conclusion:", conclusion, "\n")
```

H0: There is no significant difference in the average monthly expenditure before and after subsidy

H1: There is a significant reduction in average monthly expenditure before and after subsidy.

***Global Data Collection and Analysis***

The global data[*2] was verified and collected using two primary websites, the first one being the NHSRC website and the second one being the WHO website. The data of OOPE was collected as a percentage of Current Health Expenditure (CHE). This is done so that the global data is normalised for drawing comparisons and hence analysis. Three countries having health care index are taken into comparison with India's OOPE data, those are Australia, Germany and The United States of America (USA), as these countries are ranked higher than India as per the health care index[22]. A line chart is drawn for the comparison, and a linear regression model is made to predict the year by which the OOPE percentage reaches below 10% for all the four countries to compare the trends of OOPE. The R code for executing this is as follows:

```
# Load required libraries
```

```
library(ggplot2)
library(dplyr)

# Create the data frame
years <- 2007:2021
data <- data.frame(
  Year = years,
  Australia = c(19, 19.04, 19.05, 19.75,
19.17, 19.93, 19, 19.05, 17.85, 17.2, 16.95,
16.6, 14.73, 14.91, 13.82),
  Germany = c(14.41, 14.12, 13.9, 13.96,
14.01, 14.17, 13.32, 13.01, 13.08, 12.95,
12.92, 13.17, 13.44, 12.41, 12.16),
  India = c(70.82, 69.15, 66.76, 65.18, 62.22,
63, 69.07, 67.01, 64.66, 63.21, 55.11, 53.23,
52, 49.45, 49.82),
  USA = c(13.6, 13.36, 12.65, 12.37, 12.32,
12.34, 12.29,         11.99, 11.77, 11.65,
11.42, 11.33, 11.31, 9.93, 10.7)
)

# Convert data to long format for ggplot
data_long <- tidyr::pivot_longer(data,
                             cols = c(Australia,
Germany, India, USA),
                             names_to = "Country",
                             values_to = "Value")

# Create the line plot
plot <- ggplot(data_long, aes(x = Year, y =
Value, color = Country)) +
  geom_line() +
  geom_point() +
  theme_minimal() +
  labs(title = "Metric Values Over Time
(2007-2021)",
       x = "Year",
       y = "Metric Value") +
  theme(legend.position = "right")

# Display the plot
print(plot)

# Function to predict year when value <= 10
predict_year_below_10 <- function(years,
values, country_name) {
  # Fit linear model
  model <- lm(values ~ years)

  # Calculate R-squared
  r_squared <- summary(model)$r.squared

  # Get coefficient (slope)
  slope <- coef(model)[2]

  # Predict future years until value <= 10
  current_year <- 2021
  while(TRUE) {
    current_year <- current_year + 1
    predicted_value <- predict(model,
newdata = data.frame(years = current_year))
    if(predicted_value <= 10) break
  }

  # Print results
  cat("\n", country_name, ":\n")
  cat("Current trend:", round(slope, 2), "per
year\n")
  cat("R-squared value:", round(r_squared,
3), "\n")
  cat("Predicted year to reach ≤10:",
current_year, "\n")
  cat("Predicted value in", current_year, ":",
round(predicted_value, 2), "\n")
```

```
  return(list(year = current_year, value =
predicted_value))
}

# Perform regression analysis for each
country
cat("Regression Analysis Results:\n")
aus_pred <- predict_year_below_10(years,
data$Australia, "Australia")
ger_pred <- predict_year_below_10(years,
data$Germany, "Germany")
ind_pred <- predict_year_below_10(years,
data$India, "India")
usa_pred <- predict_year_below_10(years,
data$USA, "USA")

# Creating prediction lines extending to
target years
future_years <- 2007:max(aus_pred$year,
ger_pred$year, ind_pred$year)

# Create prediction data frames for each
country
aus_lm <- lm(Australia ~ Year, data = data)
ger_lm <- lm(Germany ~ Year, data = data)
ind_lm <- lm(India ~ Year, data = data)
usa_lm <- lm(USA ~ Year, data = data)

pred_data <- data.frame(
  Year = future_years,
  Australia = predict(aus_lm, newdata =
data.frame(Year = future_years)),
  Germany = predict(ger_lm, newdata =
data.frame(Year = future_years)),
  India = predict(ind_lm, newdata =
data.frame(Year = future_years)),
  USA = predict(usa_lm, newdata =
data.frame(Year = future_years))
)

# Convert to long format
pred_data_long <-
tidyr::pivot_longer(pred_data,
                            cols = c(Australia,
Germany, India, USA),
                            names_to =
"Country",
                            values_to = "Value")

# Create plot with prediction lines
final_plot <- ggplot() +
  geom_line(data = data_long, aes(x = Year,
y = Value, color = Country)) +
  geom_point(data = data_long, aes(x = Year,
y = Value, color = Country)) +
  geom_line(data = pred_data_long, aes(x =
Year, y = Value, color = Country),
            linetype = "dashed", alpha = 0.5) +
  geom_hline(yintercept = 10, linetype =
"dotted", color = "red", alpha = 0.5) +
  theme_minimal() +
  labs(title = "Metric Values Over Time with
Predictions",
       subtitle = "Dashed lines show projected
values, red dotted line shows target value of
10",
       x = "Year",
       y = "Metric Value") +
  theme(legend.position = "right")

# Display the final plot with predictions
print(final_plot)
```

## IV. RESULTS AND DISCUSSION

The results of all the statistical analysis done using R have been highlighted here. The first section contains the analysis of primary data, and the second section contains the analysis of the global data.

## A. Primary Data

The overall scatter plot is attached below:

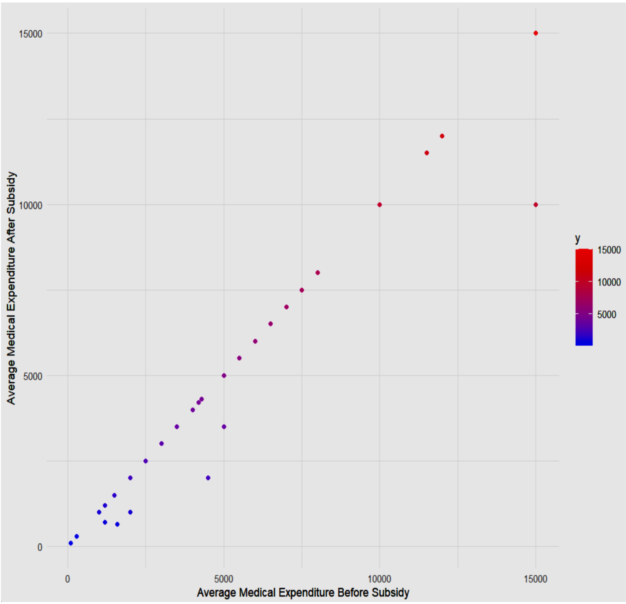



**CHART II

SCATTER PLOT OF THE PRIMARY DATA

This scatter plot signifies that there have been only few instances where a reduction has happened due to subsidy. This is either due to lack of awareness that government subsidised pharmacies exist or not visiting such pharmacies either due to distance or absence of trusted pharmacist or chemist in such shops. This is further proven with the paired-T hypothesis testing. The results of the t-values and the conclusion are as follows:

Calculated t-value: 1.185834
Critical t-value: 1.833113
Conclusion: Fail to reject the null hypothesis (no significant difference).

| *CATEGORY OF DISEASES* | *MEAN OF AVERAGE MEDICAL EXPENDITURE AFTER SUBSIDY* | *MEAN OF AVERAGE MEDICAL EXPENDITURE BEFORE SUBSIDY* |
|---|---|---|
| 1,2,4 | 13250 | 13250 |
| 1,4,6 | 11833.33333 | 11833.3333 |
| 1,5 | 10000 | 1500 |
| 1,4 | 6750 | 675 |
| 4 | 5500 | 550 |
| 1,2 | 4500 | 450 |
| 1 | 2209.090909 | 2345.45454 |
| 1,6 | 1558.333333 | 230 |
| 6 | 1350 | 135 |
| 3 | 1000 | 100 |
| **Grand Total** | **4593** | **484** |

TABLE I

CATEGORIES AND RESPECTIVE MEAN EXPENDITURE

As the null hypothesis is getting accepted, the result is that as per the primary data, there is an absence of a significant reduction or discount in the average medical expenditure due to government subsidised medical goods and services.

## B. Global Data

The line chart comparing the OOPE trend of the four countries; Australia, Germany, USA and India are as follows:

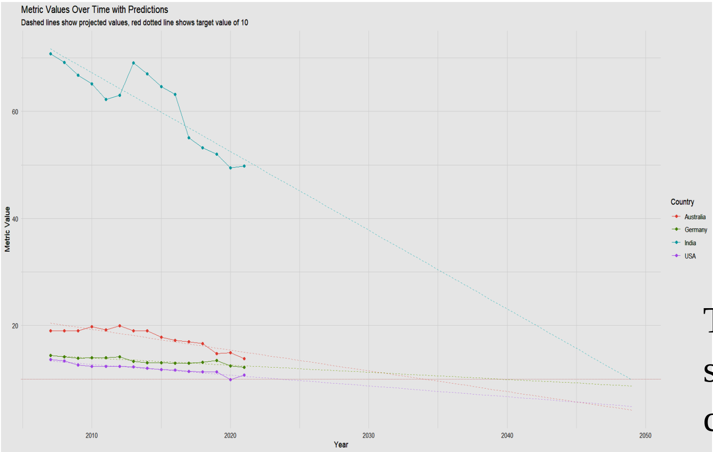



**CHART III

LINE GRAPH REPRESENTING ANNUAL OOPE VALUES AND THE PREDICTED YEAR OF REACHING A VALUE BELOW 10%

The predicted years when the OOPE reaches 10% of the CHE is given below:

*Regression Analysis Results:*

Australia :
Current trend: -0.39 per year
R-squared value: 0.781
Predicted year to reach ≤10: 2034
Predicted value in 2034 : 9.99

Germany :
Current trend: -0.13 per year
R-squared value: 0.79
Predicted year to reach ≤10: 2040
Predicted value in 2040 : 9.93

India :
Current trend: -1.47 per year
R-squared value: 0.787
Predicted year to reach ≤10: 2049
Predicted value in 2049 : 9.86

USA :
Current trend: -0.2 per year
R-squared value: 0.898
Predicted year to reach ≤10: 2024
Predicted value in 2024 : 9.92

These results do indicate that despite the significantly higher OOPE in 2021, India can manage to attain less than 10% OOPE by the year of 2049. On the other hand, Australia and Germany will attain the below 10% numbers in the years 2034 and 2040 respectively. As per the regression model, USA should have already achieved the target. As a third-world country, India has taken some significant measures to bring the OOPE down, as evident from the fact that it has the maximum decreasing trend per year among the 4 countries. If the necessary steps are taken by the government and private sector of healthcare, it would take about 25 years to achieve the levels of the first world countries.

## V. CONCLUSION

From the statistical analysis, several conclusions can be made. Firstly, from a large-scale perspective, India is making quite a good progress in terms of reducing OOPE as a percentage of CHE. The global data analysis also showed that India's declining trend is the best against three countries ranked higher on the basis of health index. The schemes implemented by the Indian government are more or less effective in this regard. However, the use of government subsidised pharmacies could be



**The enlarged versions of these charts are available in section VII.

maximised by raising awareness through newspaper posts and advertisements. The common people could be encouraged better to go to subsidised chemists and the government could open pharmacies in prime areas for better local results.

## VI. REFERENCES


[1] *Smith, T., & Nguyen, P. (2020). "Healthcare prioritization and morbidity reduction in high-prevalence conditions."* Journal of Health Economics, 39(5), 367–380.

[2] *Global Burden of Disease Study (2019). "Analysis of healthcare spending efficiency across diseases."* The Lancet Public Health, 4(10), e510-e540.

[3] *Auster, R. M., et al. (2019). "Disease burden and expenditure alignment to improve health equity."* Health Policy Journal, 28(7), 1019–1034.

[4] *Cleary, S., et al. (2018). "The role of subsidies in reducing healthcare disparities."* International Journal of Equity in Health, 17(1), 117.

[5] *Papanicolas, I., et al. (2017). "Subsidized healthcare services and out-of-pocket expense mitigation."* JAMA Network Open, 3(12), e203925.

[6] *Miller, M., et al. (2021). "Government spending's impact on healthcare access in rural areas."* American Journal of Public Health, 111(5), 930–938.

[7] *Macinko, J., & Starfield, B. (2018). "Healthcare expenditure efficiency across economic factors."* Social Science & Medicine, 45(12), 1784–1792.

[8] *Kutzin, J., et al. (2020). "Subsidy distribution aligned with economic indicators for sustainability."* Health Economics Review, 10(4), 249–263.

[9] *Hsiao, W. (2019). "Preventive care subsidies and long-term economic benefits."* Public Health Reports, 134(9), 950–963.

[10] *Bradley, E. H., et al. (2021). "Correlations between targeted subsidies and health outcomes."* Population Health Metrics, 19(6), 672–689.

[11] *Wagstaff, A., & Lindelow, M. (2019). "Disease prevalence and healthcare subsidy correlation."* Journal of Health Policy and Economics, 47(8), 780–795.

[12] *Pauly, M. (2020). "Reallocating health funds toward disease-targeted subsidies."* Health Affairs, 39(11), 2022–2031.

***Other References for Initial Reviews:***

[13] *Farmer, P., et al. (2013). "Community-based healthcare funding and region-specific health outcomes."* The Lancet Global Health, 1(1), e32–e37.

[14] *Reeves, A., et al. (2015). "Disparities in healthcare funding related to socioeconomic status."* Global Public Health, 10(7), 782–797.

[15] *CDC Report (2019). "Impact of targeted spending on disease burden in underserved regions."* CDC Annual Review, 2019, pp. 45-49.

[16] *Xu, K., Evans, D. B., Kawabata, K., Zeramdini, R., Klavus, J., & Murray, C. J. L. (2003). Household catastrophic health expenditure: A multicountry analysis. The Lancet, 362(9378), 111-117.*

[17] *Dieleman, J. L., Campbell, M., Chapin, A., Eldrenkamp, E., Fan, V. Y., Haakenstad & Murray, C. J. (2017). Future and potential spending on health 2015-40: Development assistance for health, and government, prepaid private, and out-of-pocket health spending in 184 countries. The Lancet, 389(10083), 2005-2030.*

[18] *Berndt, E. R., & Cockburn, I. M. (2014). The hidden cost of low prices: Limited access to new drugs in India. Health Affairs, 33(9), 1560-1569.*

[19] *Dieleman, J. L., et al. (2017). "Spending on health and HIV/AIDS: Domestic health spending and development assistance in 188 countries, 1995–2015."* The Lancet, 391(10132), 1799-1829.

***Websites:***

[20] *WHO Website:* https://www.who.int/data/gho/data/indicators/indicator-details/GHO/out-of-pocket-expenditure-as-percentage-of-current-health-expenditure-(che)-(-)

[21] *NHSRC Website:* https://nhsrcindia.org/national-health-accounts-records

[22] *Health Care Index Country Ranking:* https://www.numbeo.com/health-care/rankings_by_country.jsp


*1: *The link to primary data spreadsheet:* https://docs.google.com/spreadsheets/d/1Un9hZKRj0wuMdorLL6SaQkad4O0gPpt/edit?usp=sharing&ouid=104831456148754788107&rtpof=true&sd=true

*2: *The link to global data spreadsheet:*
https://docs.google.com/spreadsheets/d/1Rs1fJjYqgLTASkWWb6NnXWBY1eOu0Btx/edit?usp=sharing&ouid=104831456148754788107&rtpof=true&sd=true

# VII. ENLARGED CHARTS

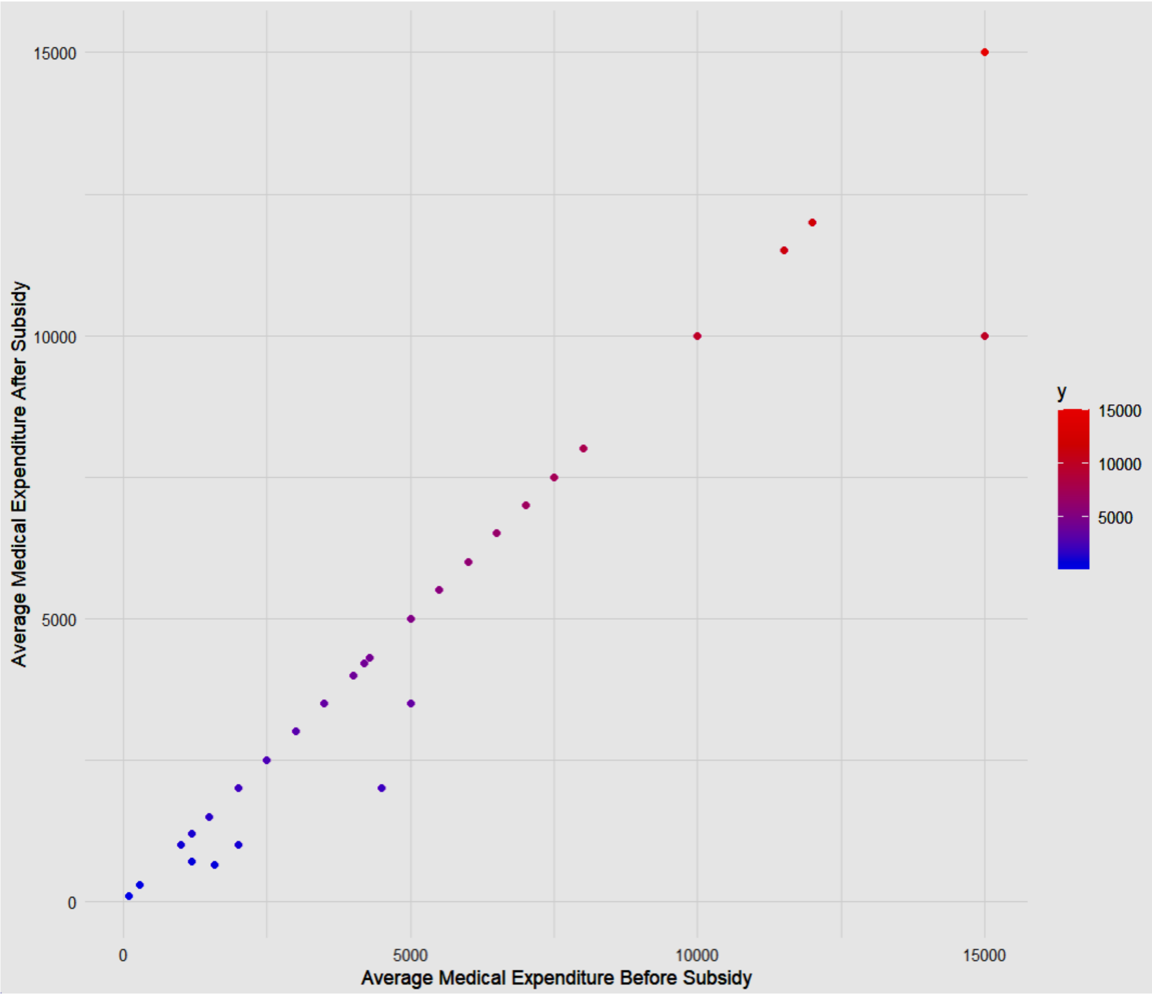



CHART II
SCATTER PLOT
OF THE
PRIMARY DATA

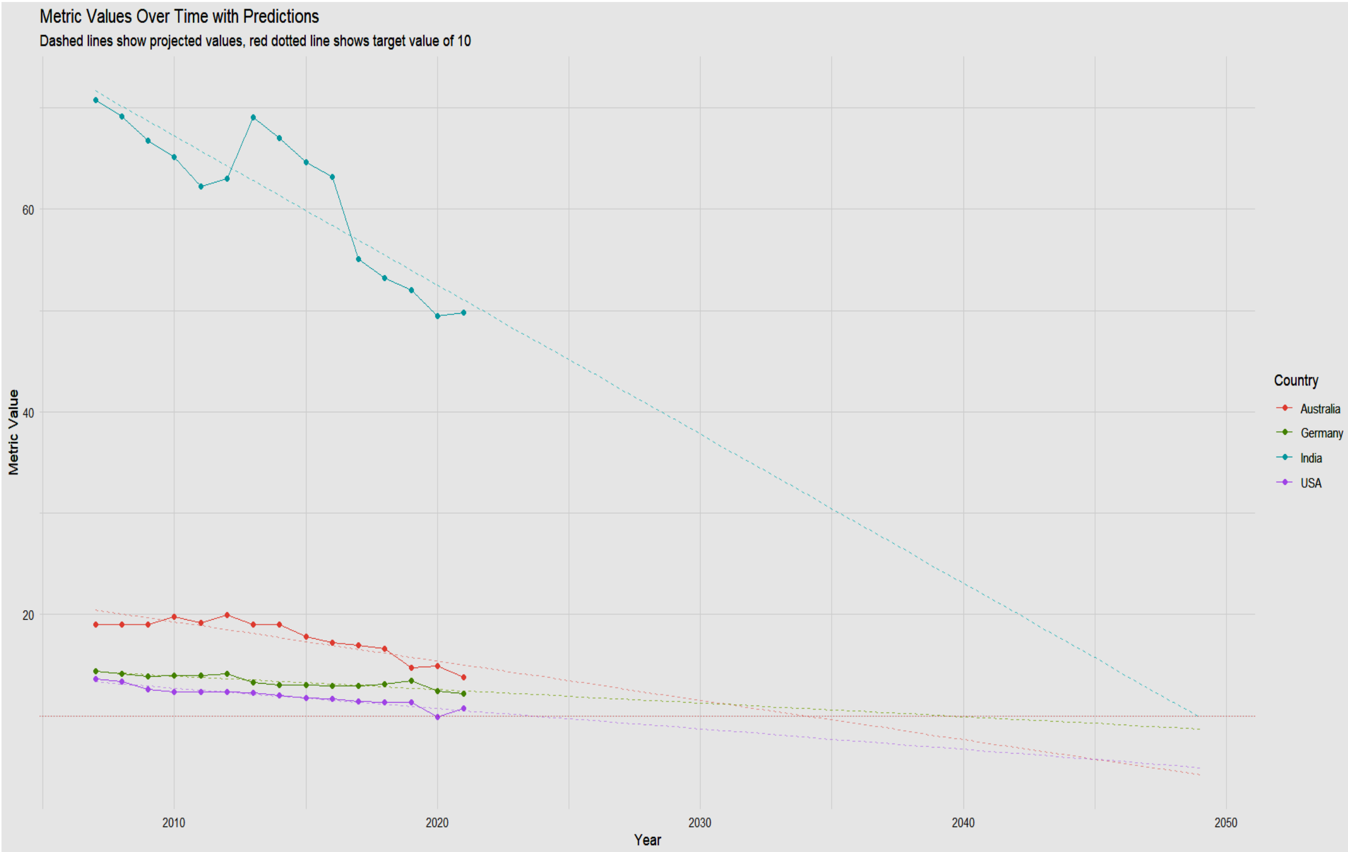



CHART IIIII
LINE GRAPH
REPRESENTING
THE OOPE VALUES
AND THE
PREDICTED YEAR
BY WHICH THE
VALUE REACHES
LESS THAN 10%