bcg_coverage_improvement_DiD

Nasir Umar

2024-07-10

Contents

Load libraries

```r
library(dplyr)
library(ggplot2)
library(survey)
library(mitools)
library(kableExtra)
library(mice)
library(tidyverse)
library(stringmagic)
library(dreamerr)
library(fixest)
library(srvyr)
library(haven)
library(Hmisc)
library(sjPlot)
library(labelled)
library(flextable)
library(gt)
library(modelsummary)
```

Load & prepare MICS data files

- Load MICS data files

```r
mics3_soml <- readRDS("Data/mics3_final_backup.rds")
mics4_soml <- readRDS("Data/mics4_final_backup.rds")
mics5_soml <- readRDS("Data/mics5_final_backup.rds")


#   Make the datasets for parallel trends analysis and DiD
mics_soml_stack_bcg <- bind_rows(mics3_soml, mics4_soml, mics5_soml)
```

- Make core DiD variables

```r
mics_soml_stack_bcg <- mics_soml_stack_bcg %>%
  mutate(
    treated = adopter_type_bin,
    post = if_else(year >= 2012, 1, 0),
    did = treated * post
  )
```

- Select variables for model & imputation

```r
mics_imp_bcg <- mics_soml_stack_bcg %>%
  select(bcg, education_mom, wealth_quintile,
         cluster_id, year, survey_weightwm, state, rural_urban_bin,
         mum_agecat, adopter_type_bin, treated, post, did)
```

Multiple imputation

  - MICE to handle missing values

```r
imp_bcg <- mice(mics_imp_bcg, m = 5, method = "pmm", seed = 123, printFlag = FALSE)

#   All imputed datasets into a list
imp_list_bcg <- complete(imp_bcg, "all")
```

- Defining survey design for each imputed dataset

```r
survey_designs_bcg <- lapply(imp_list_bcg, function(df) {
  df %>%
    svydesign(
      id = ~cluster_id,
      strata = ~state,
      weights = ~survey_weightwm,
      data = .,
      nest = TRUE
    )
})
```

Visualise parallel trend check

  - Set parallel trend check

```r
#   Use one (first one) complete imputed dataset

completed_data_bcg_imp1 <- complete(imp_bcg, 1)

#     Mean bcg coverage for plotting by treatment group and year
bcg_summary_plot <- completed_data_bcg_imp1 %>%
  group_by(treated, year) %>%
  summarise(
    mean_bcg = mean(bcg, na.rm = TRUE),
    .groups = 'drop'
  ) %>%
  mutate(
    treated_label = if_else(treated == 1, "Treated States (SOML Adopters)", "Control States"),
    year_label = as.character(year)
  )
```

- Box plot - mean coverage for each group in each year

```r
ggplot(bcg_summary_plot, aes(x = year_label, y = mean_bcg, fill = treated_label)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8), width = 0.7) +
  geom_text(aes(label = sprintf("%.2f", mean_bcg)),
            position = position_dodge(width = 0.8), vjust = -0.5, size = 3) +
  labs(
    title = "bcg Coverage by Treatment Group and Year",
    subtitle = "Years: 2007 (Pre), 2011 (Pre), 2016 (Post-SOML 2012 Launch)",
    x = "Year",
    y = "Mean bcg Coverage (0-1 scale)",
    fill = "Group"
  ) +
  scale_y_continuous(limits = c(0, max(bcg_summary_plot$mean_bcg) * 1.1)) +
  theme_minimal() +
  theme(legend.position = "bottom")
```

bcg Coverage by Treatment Group and Year
Years: 2007 (Pre), 2011 (Pre), 2016 (Post−SOML 2012 Launch)
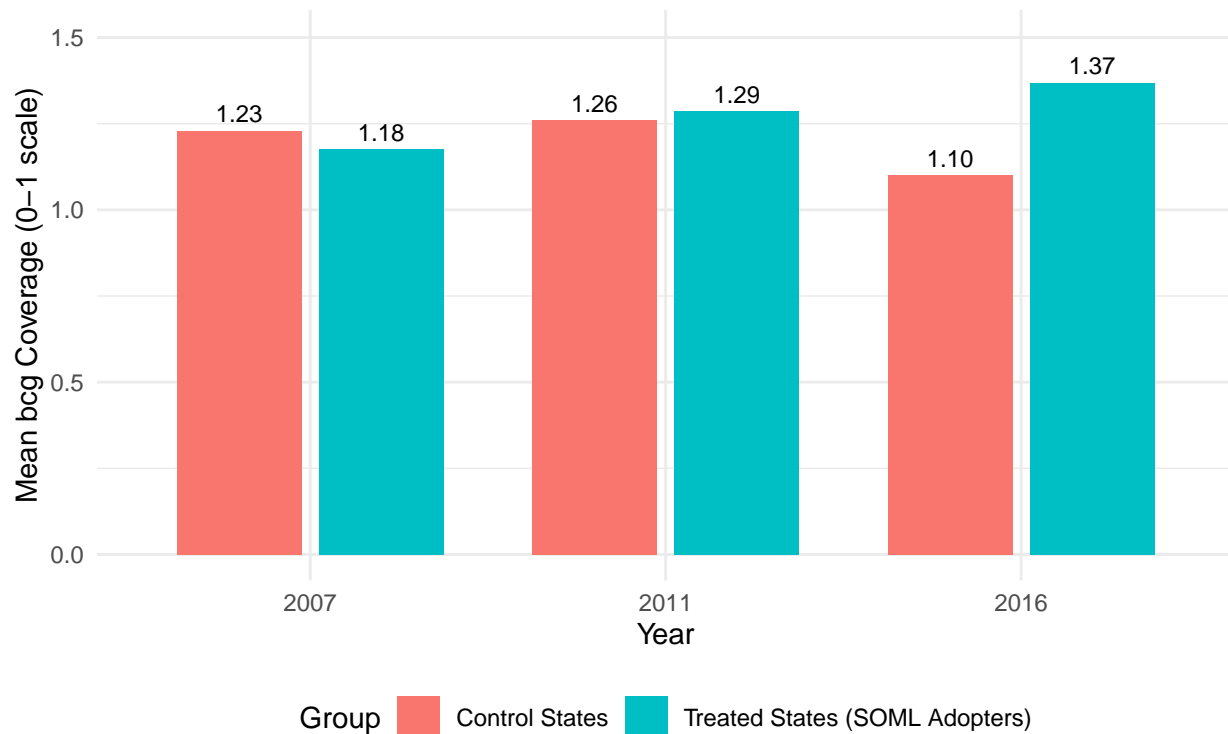
- Line plot - trend of coverage over time for both groups

```
ggplot(bcg_summary_plot, aes(x = year_label, y = mean_bcg, color = treated_label, group = treated_label
  geom_line(size = 1.2) +
  geom_point(size = 3) +

  geom_vline(xintercept = "2011.5", linetype = "dashed", color = "grey50", size = 0.8) +
  annotate("text", x = "2011.5", y = max(bcg_summary_plot$mean_bcg) * 0.95,
           label = "SOML Launch (2012)", hjust = -0.1, vjust = -0.5, size = 3.5, color = "grey50") +
  labs(
    title = "Trend of bcg Coverage by Treatment Group",
    subtitle = "Parallel Trends Check (2007-2011 Pre-Period) and Post-SOML Trends",
    x = "Year",
    y = "Mean bcg coverage (0-1 scale)",
    color = "Group"
  ) +
  scale_y_continuous(limits = c(0, max(bcg_summary_plot$mean_bcg) * 1.1)) +
  theme_minimal() +
  theme(legend.position = "bottom")
```

Trend of bcg Coverage by Treatment Group
Parallel Trends Check (2007–2011 Pre–Period) and Post–SOML Trends
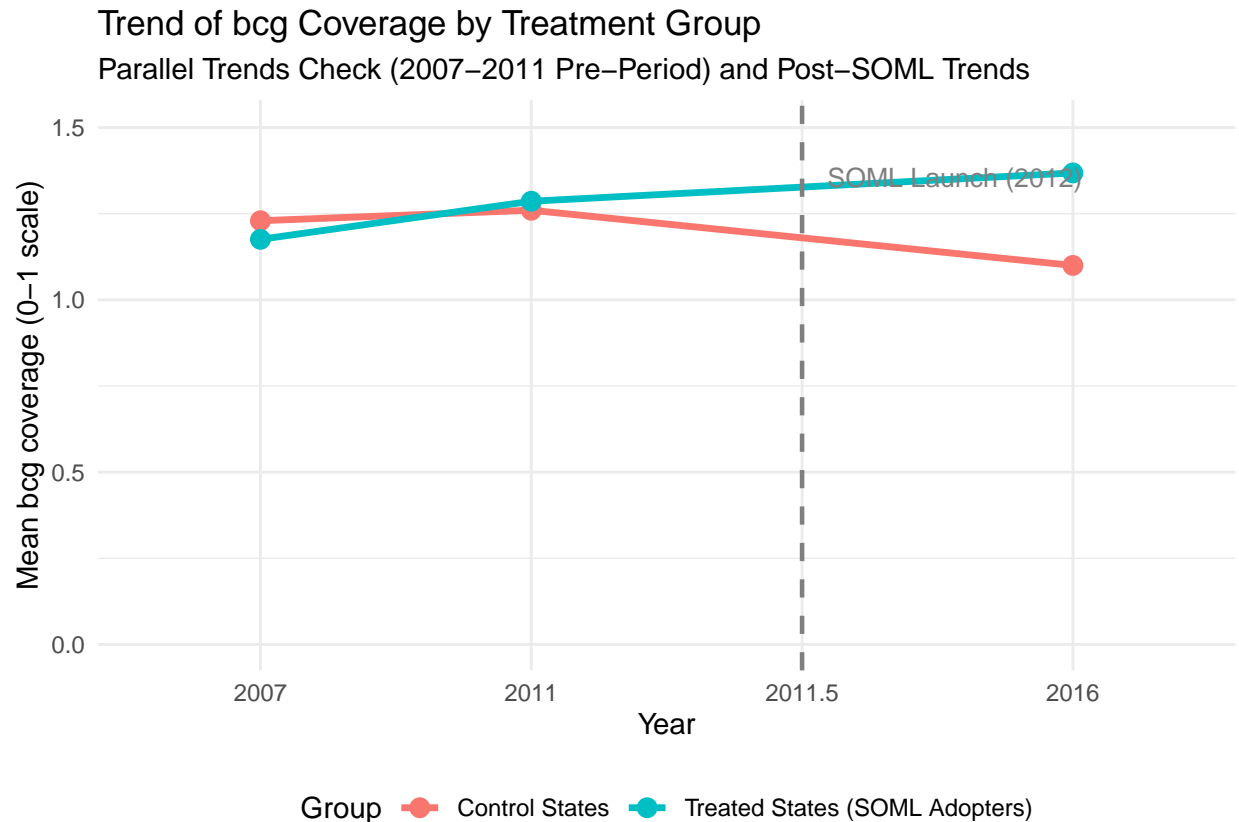
Difference-in-Differences model

- Fit DiD model

```
did_models_bcg <- lapply(survey_designs_bcg, function(design) {
  svyglm(
    bcg ~ treated * post + rural_urban_bin + mum_agecat +
      education_mom + wealth_quintile,
    design = design,
    family = quasipoisson()
  )
})
```

- Pool and format results table

```
library(survey)
did_pooled_bcg <- MIcombine(did_models_bcg)
```

- Exponentiated Coefficients

```
bcg_table <- data.frame(
  Term = rownames(confint(did_pooled_bcg)),
```

```
    IRR = round(exp(coef(did_pooled_bcg)), 4),
    Std.Error = round(sqrt(diag(did_pooled_bcg$variance)), 4),
    `CI 2.5%` = round(exp(confint(did_pooled_bcg)[, 1]), 4),
    `CI 97.5%` = round(exp(confint(did_pooled_bcg)[, 2]), 4),
    `P-value` = round(2 * pnorm(abs(coef(did_pooled_bcg)/sqrt(diag(did_pooled_bcg$variance)))), lower.tail
    check.names = FALSE
) %>%
  mutate(
    Significance = case_when(
      `P-value` < 0.001 ~ "***",
      `P-value` < 0.01 ~ "**",
      `P-value` < 0.05 ~ "*",
      TRUE ~ ""
    )
  ) %>%
  select(Term, IRR, `CI 2.5%`, `CI 97.5%`, `P-value`, Significance)
```

- Name and clarify the terms

```
bcg_table <- bcg_table %>%
  mutate(
    Term = case_when(
      Term == "(Intercept)" ~ "Intercept",
      Term == "treated" ~ "Treated Group (Pre-SOML)",
      Term == "post" ~ "Post-SOML Period",
      Term == "treated:post" ~ "DiD Estimate (SOML Effect)",
      Term == "rural_urban_bin" ~ "Rural (Ref: Urban)",
      Term == "mum_agecat" ~ "Mother's Age (Binary)",
      Term == "education_mom" ~ "Mother's Education (Binary)",
      Term == "wealth_quintile" ~ "Wealth Quintile (Binary)",
      TRUE ~ Term
    )
  )
#print(did_pooled_bcg)
```

```
bcg_table <- bcg_table %>%
  mutate(Interpretation = case_when(
    IRR > 1 & `P-value` < 0.05 ~ "Significant increase",
    IRR < 1 & `P-value` < 0.05 ~ "Significant decrease",
    TRUE ~ "No significant effect"
  ))
```

```
bcg_table %>%
  kbl(caption = "Difference-in-Differences Analysis for bcg Vaccination Coverage") %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"),
                full_width = FALSE)
```

- Plotting Incident Rate Ratios (IRRs)

Table 1: Difference-in-Differences Analysis for bcg Vaccination Coverage

|	Term	IRR	CI 2.5%	CI 97.5%	P-value	Significance	Interpretation
(Intercept)	Intercept	1.6324	1.4733	1.8088	0.0000	***	Significant in
treated	Treated Group (Pre-SOML)	0.9949	0.9425	1.0502	0.8538		No significant
post	Post-SOML Period	0.8972	0.8548	0.9416	0.0000	***	Significant de
rural_urban_bin	Rural (Ref: Urban)	1.0510	1.0016	1.1029	0.0428	*	Significant in
mum_agecat	Mother's Age (Binary)	0.9761	0.9576	0.9950	0.0133	*	Significant de
education_mom	Mother's Education (Binary)	0.9388	0.9166	0.9616	0.0000	***	Significant de
wealth_quintile	Wealth Quintile (Binary)	0.9738	0.9544	0.9937	0.0100	*	Significant de
treated:post	DiD Estimate (SOML Effect)	1.1825	1.0915	1.2811	0.0000	***	Significant in

```r
library(ggplot2)

did_irr <- data.frame(
  Term = "SOML Intervention (DiD)",
  IRR = exp(did_pooled_bcg$coefficients["treated:post"]),
  CI_low = exp(confint(did_pooled_bcg)["treated:post", 1]),
  CI_high = exp(confint(did_pooled_bcg)["treated:post", 2])
)

# Create forest plot
ggplot(did_irr, aes(x = Term, y = IRR, ymin = CI_low, ymax = CI_high)) +
  geom_pointrange(size = 1, color = "#4B9DA3") +
  geom_hline(yintercept = 1, linetype = "dashed", color = "red") +
  labs(
    title = "Impact of SOML on bcg Doses (Incidence Rate Ratio)",
    subtitle = "IRR > 1 indicates increased doses post-intervention",
    x = "",
    y = "IRR with 95% CI"
  ) +
  coord_flip() +  # Horizontal plot
  theme_minimal() +
  theme(plot.title = element_text(face = "bold"))
```

Impact of SOML on bcg Doses (Incidence Rate Ratio)

IRR > 1 indicates increased doses post–intervention

SOML Intervention (DiD)

1.0   1.1   1.2

IRR with 95% CI
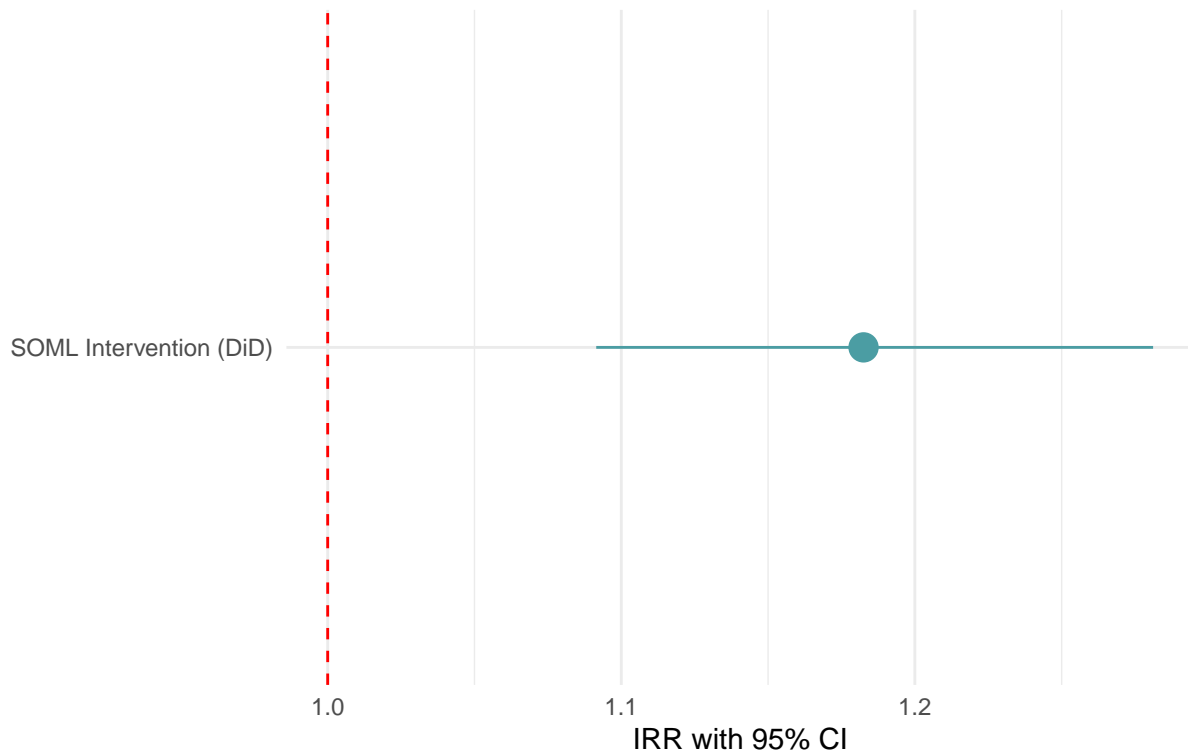
- Event-Study Analysis

Key Findings from BCG Analysis

SOML Program Success
  * 17.8% increase in BCG coverage attributable to SOML (IRR=1.183, 95% CI[1.092-1.281], p<0.001)
  * Strong evidence the program worked as intended

Concerning Background Trend
  * 10.3% decline in control areas during study period (IRR=0.897, p<0.001)
  * Highlights program's importance amid decreasing coverage

Equivalence at Baseline
  * No pre-existing differences between groups (IRR=0.995, p=0.854)
  * Validates parallel trends assumption

Equity Insights
  * Better outcomes in rural areas (+5.1%, p=0.043)
  * Disparities persist for:
  * Older mothers (-2.4%, p=0.013)
  * Less educated mothers (-6.1%, p<0.001)
  * Lower wealth households (-2.6%, p=0.010)