Impact of PM2.5 Exposure on Low Birth Weight in Ulaanbaatar (2016–2025)

Ahmedi, Barua, Bhuyan, Karayel

2025-04-29

Contents


```
knitr::opts_knit$set(root.dir = here::here())
```

List of Tables

List of Figures

1 Rationale and Research Questions

Exposure timing plays a key role in identifying how air pollution affects birth outcomes. Our initial analysis showed a weak and non-significant association between monthly PM2.5 levels and low birth weight rates ($p = 0.113$). This limited result likely occurred because measuring air pollution in the same month as birth ignores the critical periods during pregnancy when exposure actually affects fetal growth.

To address this, we will use lagged regression methods. These align PM2.5 exposure data with biologically relevant periods—like the second and third trimesters—when exposure most impacts fetal development. This approach helps reduce confounding from seasonal variations and reveals the true health impacts of PM2.5 on birth outcomes.

Research Questions: 1. Does prenatal exposure to PM2.5 affect the incidence of low birth weight (LBW) in Ulaanbaatar? 2. Which gestational periods (measured as lags) show the strongest associations between PM2.5 and LBW? 3. Are cumulative exposures over multiple months more predictive than single-month lags?

2 Dataset Information

```
# Read birth weight and live births data
birth_weight_low <- read.csv(here("Data/Raw/BIRTH WEIGTH LOWER THAN 2500 GRAMS.csv"), stringsAsFactors =

live_births <- read.csv(here("./Data/Raw/LIVE BIRTHS.csv"), stringsAsFactors = TRUE)

# Clean live births
live_births_clean <- live_births
for (col in names(live_births_clean)[-1]) {
  live_births_clean[[col]] <- as.numeric(gsub(",", "", live_births_clean[[col]]))
}

# Convert wide to long format
birth_weight_low_long <- birth_weight_low %>%
  pivot_longer(-Aimag, names_to = "Month", values_to = "Low_Birth_Weight")

live_births_long <- live_births_clean %>%
  pivot_longer(-Aimag, names_to = "Month", values_to = "Live_Births")

# Remove 'X' from month names
birth_weight_low_long <- birth_weight_low_long %>% mutate(Month = gsub("^X", "", Month))
live_births_long       <- live_births_long %>% mutate(Month = gsub("^X", "", Month))

# Merge datasets and create Date column
births_merged <- left_join(birth_weight_low_long, live_births_long, by = c("Aimag", "Month")) %>%
  mutate(Date = ym(Month)) %>%
  select(Aimag, Date, Low_Birth_Weight, Live_Births)
```

```
# Load and clean PM2.5 data
years <- 2015:2025
pm25_files <- paste0(here("Data","Raw"), "/Ulaanbaatar_PM2.5_", years, "_YTD.csv")
names(pm25_files) <- years
pm25_all <- map_dfr(pm25_files, read_csv, show_col_types = FALSE) %>%
  mutate(across(where(is.numeric), ~ na_if(., -999))) %>%
  clean_names() %>%
  rename(DateTime = date_lt) %>%
  mutate(DateTime = parse_date_time(DateTime, orders = "ymd IMp"), Date = date(DateTime))
```

```
# Aggregate PM2.5 data
pm25_daily <- pm25_all %>%
  mutate(Date = date(DateTime)) %>%
  group_by(Date) %>%
  summarize(
    raw_conc_daily = mean(raw_conc, na.rm = TRUE),
    aqi_daily = mean(aqi, na.rm = TRUE),
    hours_reported = n(),
    hours_missing_raw = sum(is.na(raw_conc)),
    hours_missing_aqi = sum(is.na(aqi)),
    .groups = "drop"
  ) %>%
  mutate(DateTime = as_datetime(Date))
```

```r
pm25_monthly <- pm25_daily %>%
  mutate(Month = floor_date(Date, "month")) %>%
  group_by(Month) %>%
  summarize(
    raw_conc_monthly = mean(raw_conc_daily, na.rm = TRUE),
    aqi_monthly = mean(aqi_daily, na.rm = TRUE),
    days_reported = n(),
    days_missing_raw = sum(is.na(raw_conc_daily)),
    days_missing_aqi = sum(is.na(aqi_daily)),
    .groups = "drop"
  ) %>%
  mutate(DateTime = as_datetime(Month))

# Merge with birth data
full_data <- births_merged %>%
  left_join(pm25_monthly, by = c("Date" = "Month")) %>%
  arrange(Date)
```

3 Exploratory Analysis

```
ggplot(pm25_daily, aes(x = Date, y = raw_conc_daily)) +
  geom_line() +
  labs(
    title = "Daily PM2.5 Concentrations (µg/m³)",
    x = "Date",
    y = "Daily mean PM2.5"
  ) +
  theme_minimal()
```

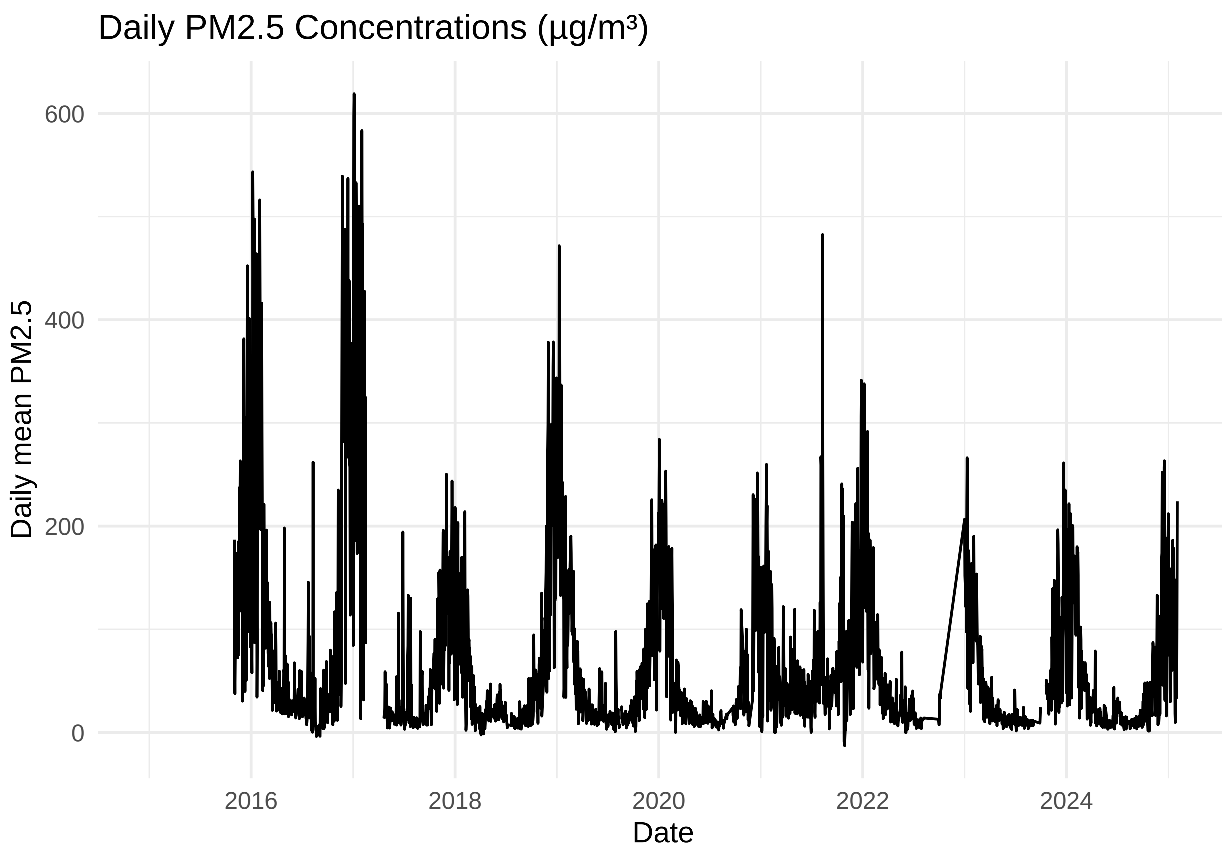

Figure 1: Daily PM2.5 Concentrations (µg/m³)

```
pm25_yearly_missing <- pm25_monthly %>%
  mutate(Year = year(Month)) %>%
  group_by(Year) %>%
  summarize(
    total_months = n(),
    months_with_missing_days = sum(days_missing_raw > 0),
    total_missing_days = sum(days_missing_raw),
    .groups = "drop"
  )
```

```r
ggplot(pm25_yearly_missing, aes(x = Year, y = months_with_missing_days)) +
  geom_col(fill = "tomato") +
  labs(
    title = "Number of Months with Missing PM2.5 Data by Year",
    x = "Year",
    y = "Months with  1 Missing Day"
  ) +
  theme_minimal()
```

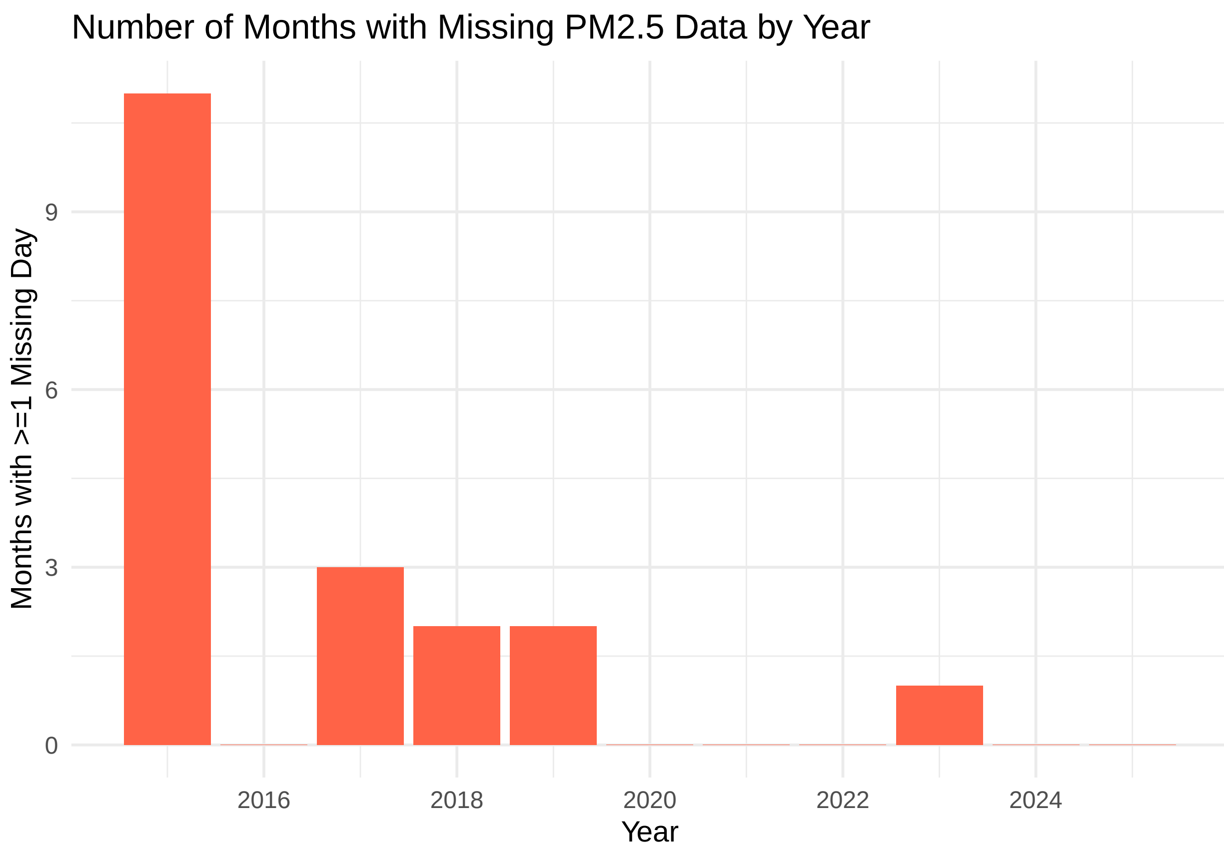

Figure 2: Number of Months with Missing PM2.5 Data by Year

```r
ggplot(pm25_monthly, aes(y = raw_conc_monthly)) +
  geom_boxplot(outlier.colour = "red", outlier.shape = 1) +
  labs(
    title = "Distribution of Monthly PM2.5",
    y = "Monthly mean PM2.5 (µg/m³)"
  ) +
  theme_minimal()
```

```r
# Compute low birth weight rate (Percentage)
full_data <- full_data %>% mutate(LBW_rate = 100 * Low_Birth_Weight / Live_Births)

ggplot(full_data, aes(x = raw_conc_monthly, y = LBW_rate)) +
```

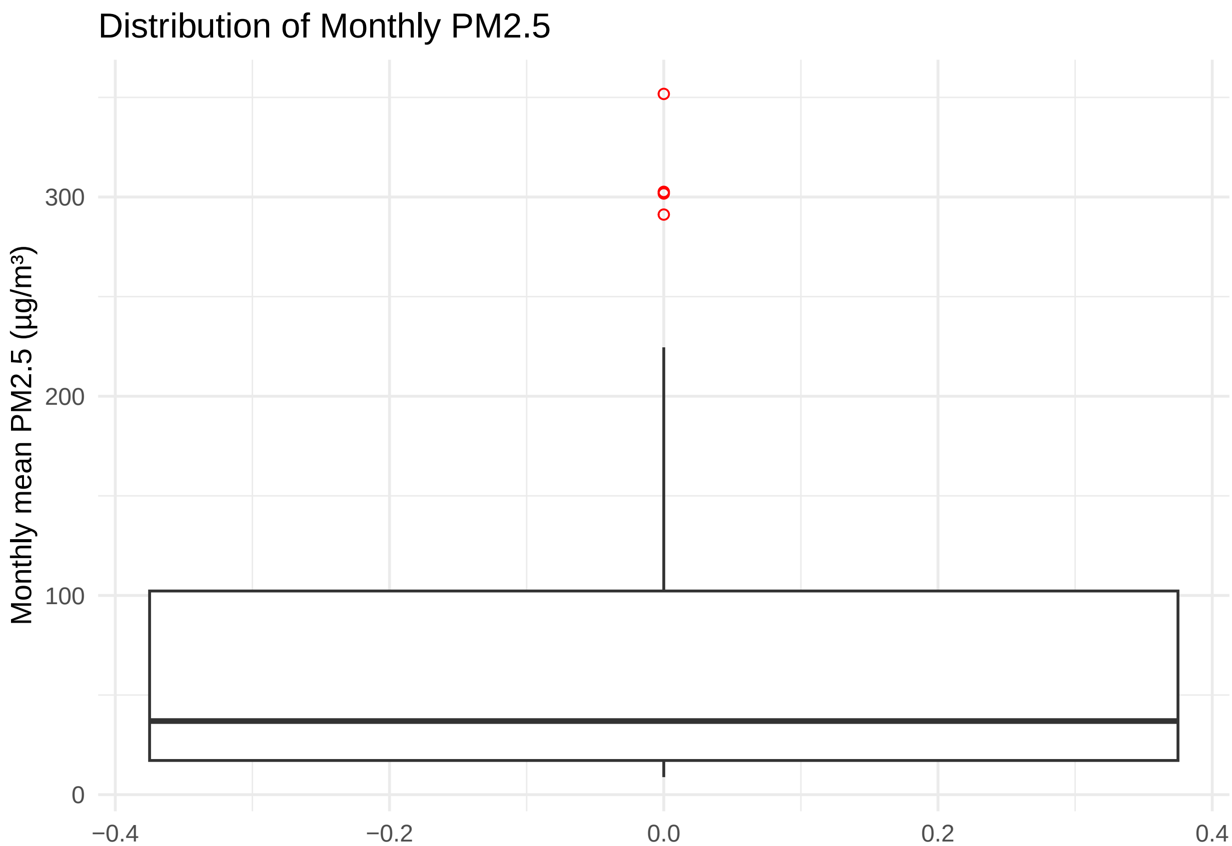

Figure 3: Distribution of Monthly PM2.5 Concentrations

```r
  geom_point() +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  labs(
    title = "Low Birth Weight Rate vs. Monthly PM2.5",
    x = "PM2.5 (µg/m³)",
    y = "LBW Rate (Percentage)"
  ) +
  theme_minimal()
```

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

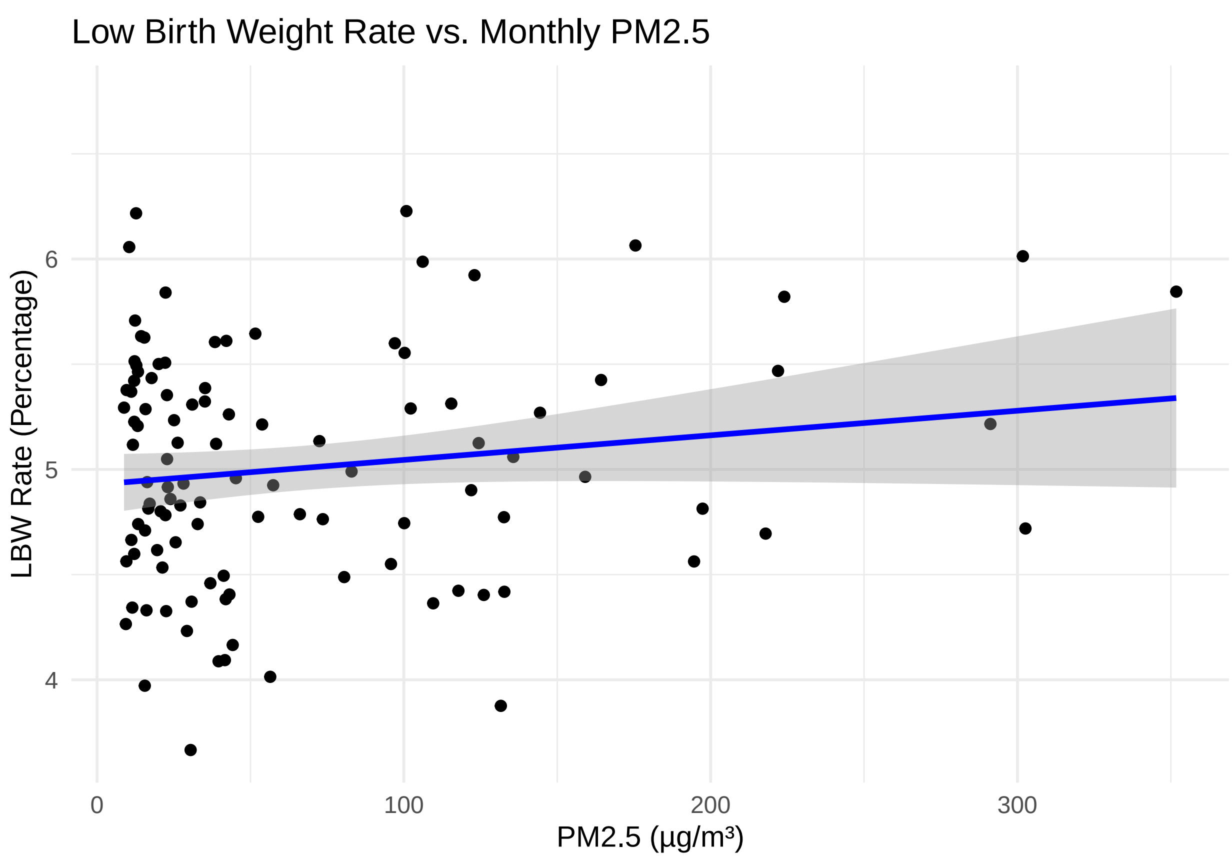

Figure 4: Low Birth Weight Rate vs. Monthly PM2.5

```r
# Summary statistics for birth outcomes and PM2.5
births_summary <- full_data %>% summarise(
  Mean_LBW = mean(Low_Birth_Weight, na.rm = TRUE),
  Median_LBW = median(Low_Birth_Weight, na.rm = TRUE),
  Min_LBW = min(Low_Birth_Weight, na.rm = TRUE),
  Max_LBW = max(Low_Birth_Weight, na.rm = TRUE),
  SD_LBW = sd(Low_Birth_Weight, na.rm = TRUE),
  N_LBW = sum(!is.na(Low_Birth_Weight)),
  Mean_Live = mean(Live_Births, na.rm = TRUE),
  Median_Live = median(Live_Births, na.rm = TRUE),
  Min_Live = min(Live_Births, na.rm = TRUE),
```

```
  Max_Live = max(Live_Births, na.rm = TRUE),
  SD_Live = sd(Live_Births, na.rm = TRUE),
  N_Live = sum(!is.na(Live_Births))
)

pm25_summary <- full_data %>% summarise(
  Mean_PM25 = mean(raw_conc_monthly, na.rm = TRUE),
  Median_PM25 = median(raw_conc_monthly, na.rm = TRUE),
  Min_PM25 = min(raw_conc_monthly, na.rm = TRUE),
  Max_PM25 = max(raw_conc_monthly, na.rm = TRUE),
  SD_PM25 = sd(raw_conc_monthly, na.rm = TRUE),
  N_PM25 = sum(!is.na(raw_conc_monthly)),
  Mean_AQI = mean(aqi_monthly, na.rm = TRUE),
  Median_AQI = median(aqi_monthly, na.rm = TRUE),
  Min_AQI = min(aqi_monthly, na.rm = TRUE),
  Max_AQI = max(aqi_monthly, na.rm = TRUE),
  SD_AQI = sd(aqi_monthly, na.rm = TRUE),
  N_AQI = sum(!is.na(aqi_monthly))
)

# Summary tables
births_summary %>%
  t() %>% as.data.frame() %>%
  rownames_to_column("Statistic") %>%
  rename(Value = V1) %>%
  kable(caption = "Summary of Birth Outcomes", digits = 2) %>%
  kable_styling(full_width = FALSE)
```

Table 1: Summary of Birth Outcomes

Statistic	Value
Mean_LBW	155.97
Median_LBW	153.00
Min_LBW	88.00
Max_LBW	214.00
SD_LBW	21.87
N_LBW	111.00
Mean_Live	3110.86
Median_Live	3187.00
Min_Live	1934.00
Max_Live	3737.00
SD_Live	360.42
N_Live	111.00

```
pm25_summary %>%
  t() %>% as.data.frame() %>%
  rownames_to_column("Statistic") %>%
  rename(Value = V1) %>%
  kable(caption = "Summary of Monthly PM2.5 Exposure", digits = 2) %>%
  kable_styling(full_width = FALSE)
```

Table 2: Summary of Monthly PM2.5 Exposure

Statistic	Value
Mean_PM25	67.28
Median_PM25	35.22
Min_PM25	8.80
Max_PM25	351.76
SD_PM25	72.90
N_PM25	107.00
Mean_AQI	111.86
Median_AQI	92.67
Min_AQI	31.73
Max_AQI	274.00
SD_AQI	66.15
N_AQI	107.00

4 Analysis

5 Results and Interpretation

6 Conclusion and Policy Implications

7 References