

Milestone 4

2025-03-05

Setup

Install packages and data

Installing important packages for the project and downloading the dataset

```
install.packages("tidyverse")
```

```
##  
## The downloaded binary packages are in  
## /var/folders/4r/kdq7r6r51dx66ggpy4zs5cfh0000gn/T//RtmpBamqet/downloaded_packages
```

```
install.packages("gt")
```

```
##  
## The downloaded binary packages are in  
## /var/folders/4r/kdq7r6r51dx66ggpy4zs5cfh0000gn/T//RtmpBamqet/downloaded_packages
```

```
install.packages("tinytex")
```

```
##  
## The downloaded binary packages are in  
## /var/folders/4r/kdq7r6r51dx66ggpy4zs5cfh0000gn/T//RtmpBamqet/downloaded_packages
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.4      v readr      2.1.5  
## v forcats    1.0.0      v stringr    1.5.1  
## v ggplot2    3.5.1      v tibble     3.2.1  
## v lubridate  1.9.3      v tidyr      1.3.1  
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(gt)  
library(tinytex)  
dhs <- read_csv("~/Documents/GitHub/bios/453/bios453/IAIR7EFL.csv")
```

```
## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)

## Rows: 724115 Columns: 423
## -- Column specification -----
## Delimiter: ","
## chr   (1): caseid
## dbl (277): v001, v002, v003, v004, v005, v012, v015, v020, v021, v022, v023,...
## lgl (145): midx_4, midx_5, midx_6, m3a_4, m3a_5, m3a_6, m2n_2, m2n_3, m2n_4,...
##
## 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.
```

1) Description of data management utilized and cleaning

As part of the data management process for this project, I

```
dhs_clean <- dhs %>%
  select(m60_1, v481, m18_1, m19_1, m17_1, v214, v457) %>%
  na.omit(dhs) %>%
  rename(
    paradrug = m60_1,
    insured = v481,
    sizechild = m18_1,
    bwkg = m19_1,
    csect = m17_1,
    duration = v214,
    anemia = v457
  ) %>%
  filter(
    paradrug != 8,
    bwkg < 5000,
    sizechild != 8
  ) %>%
  mutate(
    paradrug = as.factor(paradrug),
  )
```

2) Missing data report

3) Concept map

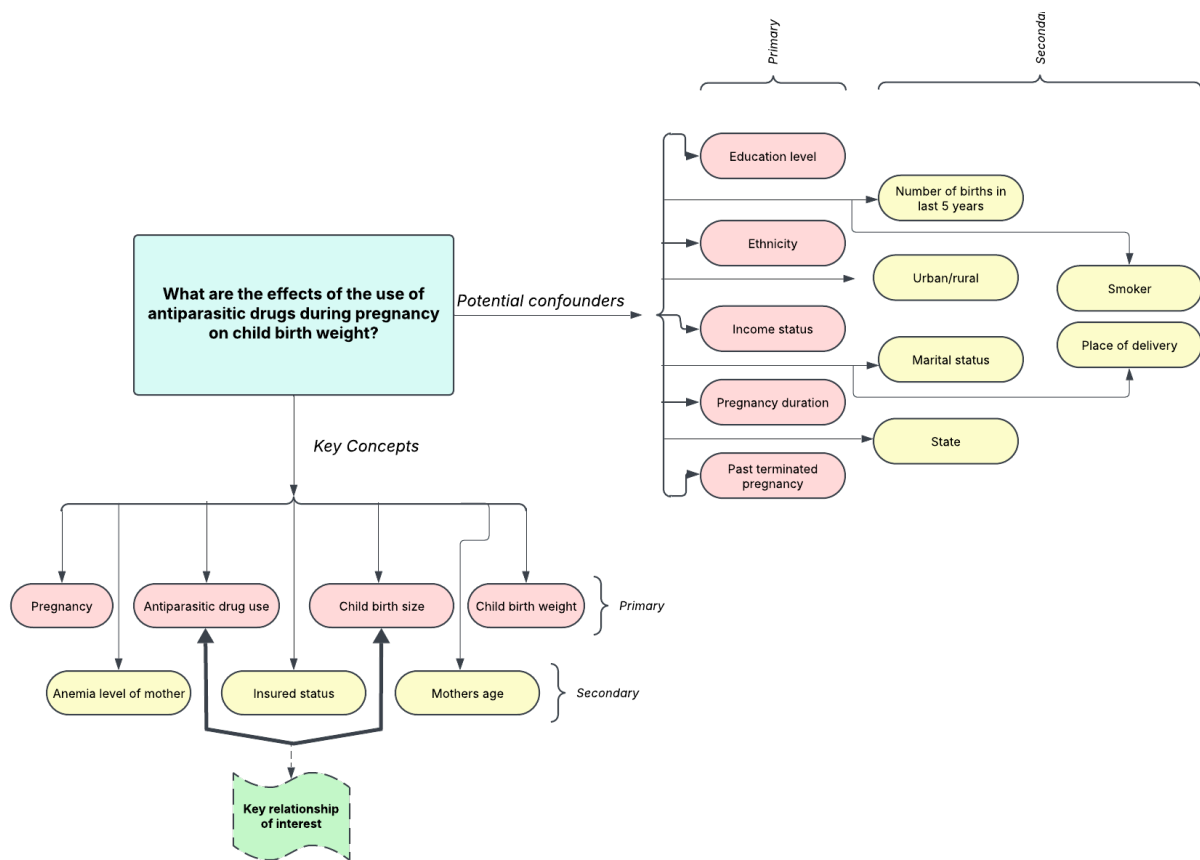


Figure 1: Concept Map