Project 3: Counterfactual Disparity Measure for Education, Smoking, and Preeclampsia in the NuMOM2b Study

Ashley I Naimi, PhD
Associate Professor
Director of Graduate Studies
Dept of Epidemiology
Emory University

✉ ashley.naimi@emory.edu

EMORY | ROLLINS SCHOOL OF PUBLIC HEALTH

Overview

Preeclampsia is a serious adverse pregnancy outcome resulting from high blood pressure and protenuria. Preeclampsia can lead to both fetal and maternal death. There are known and strong associations between preeclampsia and pre-pregnancy smoking status. Ever smokers are at a much higher risk of preeclampsia compared to never smokers. Furthermore, compared to women with less than a high school education, any education is associated with a much lower risk of preeclampsia. Finally, smoking status is also associated with education, with higher smoking rates among individuals with lower levels of education.

As such, a natural question that arises is whether the educational disparity in preeclampsia would be reduced if we somehow prevented women from smoking. Alternatively, we can ask how much of the educational disparity in preeclampsia is due to smoking status?

Overview

- We'll use a simulated dataset based on the NuMoM2b Study, a prospective observational cohort study of roughly 10,000 women experiencing their first pregnancy across the use. These data contain:

  - Exposure: Education, measured categorically as less than HS grad, high school, assciates/tech degree, some college, completed college, and degree work beyond college.
  - Mediator: Prepregnancy ever smoker
  - Outcome: Preeclampsia
  - Confounders: marital status, income, maternal age, pre pregnancy BMI, diet quality, physical activity, diabetes status, and hypertension status

Overview

- Three Phases - data management and exploration; analysis; interpretation

- Seven steps

    - Setup Project Folder
    - Data Management and Exploration
    - Import, Transform, Explore Raw Data
    - Regression Analysis
    - CDM Analysis
    - Interpretation

Overview: The Data

```
## # A tibble: 6 × 14
##   pree_acog education   momra…¹ dadra…² married cigsp…³ cigsd…⁴ v1_in…⁵ momage bmipr…⁶ hei20…⁷ pa_to…⁸ predi…⁹
##      <dbl> <chr>        <chr>   <chr>      <dbl>   <dbl>   <dbl>   <dbl>  <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1       0 [Degree wo… [NH Ot… [NH Ot…        1       0       0      14     35    32.5   0.651  0.0110       0
## 2       0 [HS grad o… [NH Wh… [NH Wh…        1       0       0      11     28    23.8   0.326  0.0148       0
## 3       0 [Degree wo… [NH Wh… [NH Wh…        1       0       0      11     30    21.5   0.657  0.0305       0
## 4       0 [Completed… [NH Wh… [NH Wh…        1       0       0       8     24    21.3   0.359  0.0264       0
## 5       0 [Assoc/Tec… [NH Wh… [NH Wh…        1       0       0      12     31    24.4   0.504  0.0247       0
## 6       0 [Degree wo… [Hispa… [NH Wh…        0       0       0      11     27    29.3   0.528  0.0137       0
## # … with 1 more variable: prehtn1 <dbl>, and abbreviated variable names ¹momrace4, ²dadrace4, ³cigsprepreg,
## #   ⁴cigsduring, ⁵v1_income, ⁶bmiprepreg, ⁷hei2015_total, ⁸pa_totmetwk_v1, ⁹prediab1
```

```
## # A tibble: 6 × 2
##   education                   n
##   <chr>                   <int>
## 1 [Assoc/Tech degree]      1081
## 2 [Completed college]      3266
## 3 [Degree work beyond college] 2889
## 4 [HS grad or GED]          738
## 5 [Less than HS grad]       445
## 6 [Some college]           1617
```

```
## # A tibble: 2 × 2
##   pree_acog     n
##       <dbl> <int>
## 1         0  9178
## 2         1   858
```

```
## # A tibble: 2 × 2
##   `as.numeric(cigsprepreg > 0)`     n
##                           <dbl> <int>
## 1                             0  8656
## 2                             1  1380
```

Phase 1: Data Management and Exploration

Step 1: Construct a Project Folder

- Create a folder called `Disparities_Project3`

- Create Subfolders: `data`, `code`, `figures`, `misc`, `sandbox`, and `report`

- Create an RStudio Project in the `Disparities_Project3` folder

Step 2: A Data Management and Exploration File

- In the code subfolder, create two code files:

  - data_man.R
  - main_analysis.R

- Open the data_man.R file and include the **<u>preamble</u>** at the top:

```r
packages <- c("data.table","tidyverse","skimr","here", "lmtest", "sandwich")

for (package in packages) {
  if (!require(package, character.only=T, quietly=T)) {
    install.packages(package, repos='http://lib.stat.cmu.edu/R/CRAN')
  }
}

for (package in packages) {
  library(package, character.only=T)
}
```

Step 3: Import, Explore, and Transform Data

- Use the `read_csv()` function to import the data.

- There is little need to transform variables in these data.

- Conduct a basic exploratory analysis of the variables in the numom data.

Step 3: Import, Explore, and Transform Data

- explore the data you just imported using:

  - the `head()` and/or `tail()` functions
  - `skimr::skim()`
  - `GGally::ggpairs()`
  - `summary()`
  - basic descriptives and table functions: `mean(), median(), sd(), table()`
  - other

Step 3: Import, Explore, and Transform Data

- Conduct some basic regression modeling of the numom data

  - Look at the relationship between preeclampsia and education, preeclampsia and smoking, and education and smoking
  - Examine a table or a histogram of the income variable. Can you determine what might be a useful threshold to dichotomize this variable?

Step 4: Export the Data

- Create an analytic dataset with any transformed variables that you may use

```
write_csv(here("data","analytic_numom.csv"))
```

Phase 2: Analysis

Step 1: Set Up and Import

- Open the `main_analysis.R` file and include the **preamble** at the top. Add the `lmtest`, `sandwich`, and the `VGAM` packages to the list.

- Import the `analytic_numom.csv` data, and explore. For example:

```
head(a)
tail(a)

skimr::skim(a)
```

Step 2: Basic Regression

- Fit an outcome model regressing `pree_acog` against all other variables in the data

- Fit a propensity score model for the categorical education variable adjusting for `momrace4` and `dadrace4`

- You'll need to use the `vglm()` function from the `VGAM` package

  - Use the `family = multinomial` argument

- You should extend the code used in project 2 for the multicategory education variable

Step 2: Basic Regression

- Generate a propensity score for the six category education variable

  - Create a PS overlap plot for income and save the figure using `ggsave()`
  - Construct stabilized IP weights for the exposure and look at the distribution using `summary()`

- Interpret these models, figures, and summaries

Step 3a: Exposure - Outcome Association

a. Fit a marginally adjusted regression model (g computation) for the association between `education` and `preeclampsia`. Use the bootstrap to obtain standard errors.

- Use this marginally adjusted approach to estimate the **risk difference** and the **risk ratio** for the association between education and preeclampsia. Contrast the risk that would be observed if everyone had some college, versus if everyone had less than high school.

b. Create an indicator (dummy) variable for whether a woman smoked any cigarettes prior to pregnancy. Make the referent level "never smoker"

c. Conduct a likelihood ratio test for the interaction between education and cigarette smoking, with preeclampsia as the outcome.

Step 3a: Exposure - Outcome Association

a. Fit a regression model for `preeclampsia` regressed against `education`. Weight this regression using the IP weights constructed with the `vglm` model in a previous step.

- Fit the model using an identity link and a log link to obtain the **risk difference** and the **risk ratio** for the association between education and preeclampsia. Contrast the risk that would be observed if everyone had some college, versus if everyone had less than high school.
- Use the robust variance estimator to obtain standard errors.

Step 3b & 3c: CDM, Structural Transformation and IP weighting

Estimate the association between education and preeclampsia that would be observed if all women were never smokers. Use the structural transformation method and IP weighting.

Step 5: Combine the Results

a. Combine all the results into a single table

b. Export the table to a csv file

Step 5: Combine the Results

	Estimate	Std Error	LCL	UCL
ATE RD: Marginally Standardized	-0.05	0.01	-0.09	-0.02
ATE RD: IP Weighted	-0.06	0.02	-0.11	-0.02
CDM RD: Structural Transformation	-0.12	0.03	-0.19	-0.06
CDM RD: IP Weighted	-0.16	0.04	-0.24	-0.08
----------------------------------	--------	-----------	------	-----
ATE RR: Marginally Standardized	0.58	0.15	0.43	0.77
ATE RR: IP Weighted	0.54	0.17	0.39	0.76
CDM RR: Structural Transformation	0.38	0.83	0.07	1.93
CDM RR: IP Weighted	0.32	0.18	0.22	0.46

Phase 3: Interpretation

Step 6: Why are the CDMs different from the ATEs?

	mu1	mu0
ATE: Marginally Standardized	0.074	0.12
CDM: Structural Transformation	0.041	0.16
CDM: IP Weighted	0.077	0.21

Step 7: Interpretation

- Exercise: Interpret the risk differences and risk ratios for the ATEs and the CDMs in the Table above.

Q&A

Project 2: Counterfactual Disparity Measure for Race, Income, and High Blood Pressure in the NHEFS

Ashley I Naimi, PhD
Associate Professor
Emory University

✉ ashley.naimi@emory.edu
⏺ ainaimi

EMORY | ROLLINS SCHOOL OF PUBLIC HEALTH