UAW-GM Cohort Study

Left truncation

October 22, 2021

Motivation

- Although FU for mortality starts in 1941, FU for cancer incidence does not start until 1985
- The cohort in 1985 is the result of both selection and left censoring
  - Left censoring: cancer incidence information is missing prior to 1985
  - Selection: Person-time starting in 1985 not representative of person-time starting in 1941

Selection bias

- Person-time after 1984 has undergone a period of selection where those most susceptible to cancer have either died or gotten cancer
- Hence, if our analysis starts in 1985, then we need to down-weight individuals with low susceptibility and up-weight individuals with high susceptibility
  - This can be done by inverse probability weighting

Pseudo-population by inverse probability weighting

- **Main idea.** Weight units by the inverse of their probability of selection/entry such that those in the study stand-in for those not in the study
- **Example of use.** In an analysis accounting for attrition due to death, Weuve et al. (2012) weighted units by the inverse of their estimated survival probability[1]

$$w_i = \frac{1}{\mathbb{P}\left(C = 1 \mid X_i\right)}$$

where $C = 1$ indicates that an individual was selected ie **in the cohort** ie still alive, and X are covariates used to predict selection

[1] Simplified for convenience.

Review of IPW

- IP weighting for selection bias is analogous to IP weighting for estimating counterfactual quantities
- Let X be covariates preceding election; $g(X) = \mathbb{P}(C = 1 \mid X)$ be the probability of selection given covariates X; and Y be some outcome measured at a later time
- Assume
  - $C \perp\!\!\!\perp Y_c \mid X$ for $c = 0, 1$ (strong ignorability)
  - $0 < g(X) < 1$ (positivity)
- Then $C \perp\!\!\!\perp X \mid g(X)$ and

$$\boxed{\mathbb{E}[Y_1] = \mathbb{E}\left[\frac{CY}{g(X)}\right]}, \qquad \mathbb{E}[Y_0] = \mathbb{E}\left[\frac{(1-C)Y}{1-g(X)}\right]$$

Susceptibility and estimated mortality

- Weuve et al. (2012) up-weighted those with a high estimated risk of death to account for those who died before a certain FU interval
  - The validity of this strategy rests in the belief that the selection mechanism (mortality probability) is correctly modeled
- It would be reasonable for us to follow the approach of Weuve et al. (2012) if we were able to model mortality status in 1985 as a function of past susceptibility to MWF
  - Since we have no measure of susceptibility, we assume the extreme case that those with a high estimated risk of death survived *because* they had low susceptibility
- So, we can estimate $\mathbb{P}\left(C = 1 \mid X\right)$ using $1 - \hat{S}(1985 \mid X)$ where $\hat{S}(t)$ is the predicted survival through time t

In other words,

Why does the inverse-weighted estimator work?

$$
\begin{aligned}
\mathbb{E}\left[\frac{CY}{g(X)}\right] &= \mathbb{E}\left[\frac{CY_1}{g(X)}\right] \\
&= \mathbb{E}\,\mathbb{E}\left[\frac{CY_1}{g(X)} \mid X\right] \\
&= \mathbb{E}\left[\frac{1}{g(X)}\mathbb{E}\left[CY_1 \mid X\right]\right] \\
&= \mathbb{E}\left[\frac{\mathbb{E}\left[C \mid X\right]}{g(X)}\mathbb{E}\left[Y_1 \mid X\right]\right] \\
&= \mathbb{E}\left[\frac{\mathbb{P}\left(C = 1 \mid X\right)}{g(X)}\mathbb{E}\left[Y_1 \mid X\right]\right] \\
&= \mathbb{E}\left[\frac{g(X)}{g(X)}\mathbb{E}\left[Y_1 \mid X\right]\right] \\
&= \mathbb{E}\,\mathbb{E}\left[Y_1 \mid X\right] = \mathbb{E}\left[Y_1\right]
\end{aligned}
$$

Notes

- We believe that every subject in the 1985 cohort has undergone some selection process, so we have $C = 1$ for all
  - This means we can only estimate $\mathbb{E}[Y_1]$ the expected value of the outcome when everyone had low enough susceptibility to have made it to 1985 (can do this in a MSM)
- Weighting by $1 - g(X)$ rather than $1/g(X)$ would lead to a breakdown in the derivation shown above
- In the Weuve et al. (2012) example,

$$\text{estimate } \mathbb{P}(C = 1 \mid X) \text{ with } \hat{S}(t_0 \mid X)$$

- In our proposal,

$$\text{estimate } \mathbb{P}(C = 1 \mid X) \text{ with } 1 - \hat{S}(t_0 = 1985 \mid X)$$

Citations

Weuve, Jennifer, Eric J Tchetgen Tchetgen, M Maria Glymour, Todd L Beck, Neelum T Aggarwal, Robert S Wilson, Denis A Evans, and Carlos F Mendes de Leon. 2012. "Accounting for Bias Due to Selective Attrition: The Example of Smoking and Cognitive Decline." *Epidemiology (Cambridge, Mass.)* 23 (1): 119.