

Discussion Assignment 2

Introduction to Causal Inference

1 Instructions

For the two studies discussed in the previous assignment **think** through the following questions. **Please provide brief *written* answers to the questions for your group's assigned study (study 1 or 2).** Use the notation developed in class. You are encouraged to discuss with your group, but please bring your own written responses to class. **Be prepared to *present* your answers in class.**

2 Background (Steps 1-2 of the roadmap)

• Consider the following structural causal models \mathcal{M}^F for studies 1 and 2. Other causal models from assignment 1 may also be correct, but to simplify discussion we will work with these.

- Study #1: *Physical activity and mortality in the elderly*

Endogenous variables: $X = (W, A, Y)$

Exogenous variables: $U = (U_W, U_A, U_Y) \sim P_U$

Structural equations F :

$$\begin{aligned}W &= f_W(U_W) \\A &= f_A(W, U_A) \\Y &= f_Y(W, A, U_Y)\end{aligned}$$

where W ={smoking, comorbidities, body fat}, A ={energy expenditure} and Y ={survival}. By collapsing the baseline variables into a single node W , we lose the information that smoking was measured first. However, there are no consequences for our statistical estimation problem.

- There are no independence assumptions on P_U .

- Study #2: *Effect of male circumcision on risk of HIV acquisition*

Endogenous variables: $X = (W, A, Z, Y)$

Exogenous variables: $U = (U_W, U_A, U_Z, U_Y) \sim P_U$

Structural equations F :

$$\begin{aligned}W &= f_W(U_W) \\A &= f_A(W, U_A) \\Z &= f_Z(W, A, U_Z) \\Y &= f_Y(W, A, Z, U_Y)\end{aligned}$$

where W ={tribe, religion}, A =male circumcision, Z ={sexual behavior, STI}, and Y =HIV status.

- There are no independence assumptions on P_U .

• **Suppose we are interested in the average treatment effect of causal risk difference.**

- *Study #1: Physical activity and mortality in the elderly*

The target causal parameter is the difference in the counterfactual probability of survival if all elderly had a high energy expenditure ($A = 1$) and the counterfactual probability of survival if all elderly had a low energy expenditure ($A = 0$)

$$\Psi^F(P_{U,X}) = E_{U,X}(Y_1) - E_{U,X}(Y_0)$$

where the counterfactual outcome Y_a is the seven-year survival for an individual if, possibly contrary to fact, the individual had energy expenditure $A = a$ for this two week period.

- *Study #2: Effect of male circumcision on risk of HIV acquisition*

The target causal parameter is the difference in the counterfactual risk of HIV acquisition if all males were circumcised ($A = 1$) and the counterfactual risk of HIV acquisition if all males were not circumcised ($A = 0$)

$$\Psi^F(P_{U,X}) = E_{U,X}(Y_1) - E_{U,X}(Y_0)$$

where the counterfactual outcome Y_a is the two-year HIV status for an individual, if possibly contrary to fact, the individual had circumcision status $A = a$.

3 Questions to be answered

1. Specify your observed data.
 - (a) What notation do we use to refer to the distribution of the observed data?
 - (b) Specify the link between the SCM and the observed data.
 - (c) What is the statistical model \mathcal{M} ? Does the SCM place any restrictions on \mathcal{M} ?
2. Using the backdoor criteria, assess identifiability of $\Psi^F(P_{U,X})$.
 - (a) If not identified, under what assumptions would it be? Are some of these sets of additional assumptions more plausible than others? Are there additional measurements you could make so that the needed identifiability assumptions are more plausible?
 - (b) What notation do we use to denote the original SCM, augmented with additional assumptions needed for identifiability?
3. Specify the target parameter of the observed data distribution (i.e. the statistical estimand).
4. What is the relevant positivity assumption? Are you concerned about violations of the positivity assumption in your study?

4 Study-specific bonus (optional) questions:

1. **Study #1.** The investigators assume no unmeasured common causes of (W, A, Y) is this necessary? Is it sufficient?
2. **Study #2.** The study investigators adjust for (condition on) $W = \{\text{tribe, religion}\}$ and $Z = \{\text{sexual behavior, STI}\}$. Under what causal structure would W would satisfy the back door criterion, but (W, Z) would not? Under what causal structure would (W, Z) satisfy the back door criterion, but W alone would not?