# R Lab One

**Katherine Wolf**
Introduction to Causal Inference (PH252D)
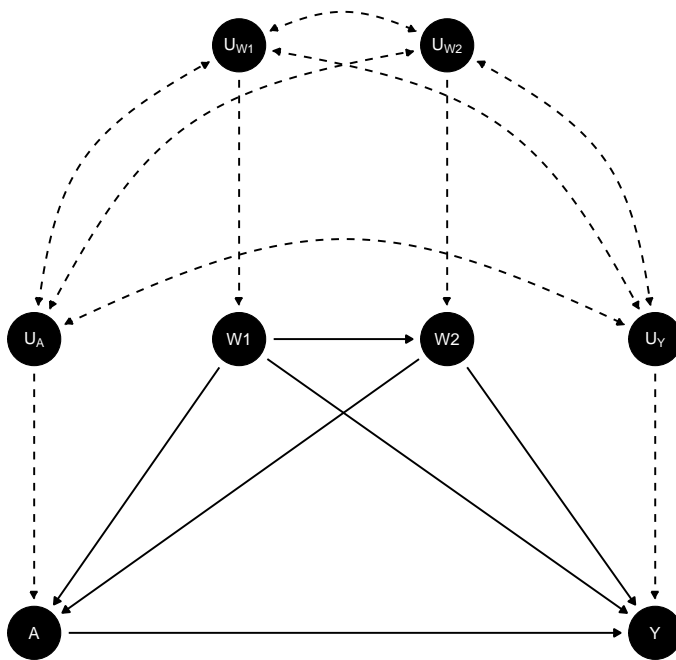February 26, 2020

## 1 Background Story

## 2 Steps 1-5 of the Roadmap

### 2.1 Step 1: Causal model representing real knowledge

a. **Draw the accompanying directed acyclic graph (DAG).**



b. **Are there any exclusion restrictions? Recall we are working with recursive (time-ordered) structural causal models.**

c. **Are there any independence assumptions on the distribution of unmeasured factors $\mathbb{P}_U$?**

### 2.2 Step 2: Counterfactuals and causal parameter

a. **Define the counterfactual outcomes of interest with formal notation and in words.**

b. **How are counterfactuals derived?**

   c. Suppose we are interested in the average treatment effect. Specify the target causal parameter. Use formal notation as well as explain in words.

## 2.3   Step 3: Observed data and link to causal

   a. Specify the link between the SCM and the observed data.

   b. What restrictions, if any, does the SCM place on the allowed distributions for the observed data? (Recall d-separation.)

   c. What notation do we use to denote the true (but unknown) distribution of the observed data and the statistical model?

## 2.4   Steps 4-5: Identification and statistical estimand

   a. Using the backdoor criterion, assess identifiability.

   b. If the target causal parameter is not identified, under what assumptions would it be?

   c. What notation is used to denote the original SCM augmented with additional assumptions needed for identifiability?

   d. Specify the target parameter of the observed data distribution (the statistical estimand).

   e. What is the relevant positivity assumption? Is it reasonable here?

# 3  Bonus: Identifying the Mean Outcome Under a Dynamic Intervention

1. Explain why (1) holds using properties of conditional expectations. Given access to the full population and the ability to implement intervention $d$, what does (1) tell you about how you could compute $\mathbb{E}_{U,X}[Y_d]$?

2. Explain why (2) holds using properties of conditional expectations and the fact that $Y_d \perp\!\!\!\perp A | W_1, W_2$ under our convenience assumptions for the backdoor criterion made in Question 4 of Section 2.

3. Explain why (3) holds. What does this mean in terms of the RUTF example?

4. Explain why (4) holds. What does this mean in terms of the RUTF example?

# 4   A Specific Data-Generating Process

## 4.1   Closed form evaluation on the target parameter

1. Evaluate the target causal parameter $\psi^F(\mathbb{P}_{U,X})$ in closed form (i.e., by hand) for this data generating process.

2. Interpret $\psi^F(\mathbb{P}_{U,X})$.

## 4.2   Translating this data generating process for $\mathbb{P}_{U,X}$ into simulations, generating counterfactual outcomes and evaluating the target causal parameter.

1. First set the seed to 252.

2. Set $n = 50,000$ as the number of independent and identically distributed draws from the data-generating process.

3. Simulate the background factors $U$.

4. Evaluate the structural equations $F$ to deterministically generate the endogenous nodes $X$.

5. Intervene to set the supplement to RUTF ($A = 1$) and generate counterfactual outcomes $Y_1$ for $n$ units. Then intervene to set the supplement to the standard ($A = 0$) and generate counterfactual outcomes $Y_0$ for $n$ units.

6. Create a data frame $X$ to hold the values of the endogenous factors $(W_1, W_2, A, Y)$ and the counterfactual outcomes $Y_1$ and $Y_0$. The rows are the n children and the columns are their characteristics. Use the head and summary to examine the resulting data.

7. Evaluate the causal parameter $\psi^F(\mathbb{P}_{U,X})$ for this population of 50,000 units.

# 5 Defining the Target Causal Parameter with a Working Marginal Structural Model

1. **For** $n = 5,000$ **children, generate the exogenous factors** $U$ **and the pre-intervention covariates** $(V, W1, W2)$. **Then set** $A = 1$ **to generate the counterfactual weight gain under RUTF** $Y_1$. **Likewise, set** $A = 0$ **to generate the counterfactual weight gain under the standard supplement** $Y_0$.

2. **Create a data frame** `X.msm` **consisting of age** $V$, **the set treatment levels** $a$, **and the corresponding outcomes** $Y_a$.

3. **Evaluate the target causal parameter.**

4. **Interpret the results.**