# R Lab 3 Part 2 - Understanding Time Dependent Confounding and Identifiability in Longitudinal Context

Advanced Topics in Causal Inference

**Assigned:** September 28, 2021
**Lab due:** October 05, 2020 on bCourses. Please answer all questions and include relevant `R` code. You are encouraged to discuss the assignment in groups, but should not copy code or interpretations verbatim. Upload your own completed lab to bCourses.

**Last lab:**
Translate causal questions into target causal parameters and intervene on the data generating processes described by Structural Causal Models (SCMs) to evaluate the true value of these parameters.

**This lab:**
1. Review the concept of identifiability.
2. Determine which assumptions let us achieve identifiability.
3. Write the target causal parameter (a function of the counterfactual or "post intervention" distribution) as a function of the observed data distribution.
4. Obtain the value of the statistical estimand.
5. Understand the challenges posed by time-dependent confounding.

**Next lab:**
Implementation of longitudinal IPTW to estimate the intervention specific mean and the parameters of a marginal structural model.

---

## Data Structure 2: $O = (L(1), A(1), L(2), A(2), L(3), A(3), L(4), A(4), Y)$

1. SCM:

   $U = (U_{L(t)}, U_{A(t)}, U_Y), t = 1, ..., 4 \sim P_U$. Assume $U$s are jointly independent.
   Structural equations, $F$:

$$\begin{aligned} L(1) &= f_{L(1)}(U_{L(1)}) \\ A(1) &= f_{A(1)}(L(1), U_{A(1)}) \\ L(t) &= f_{L(t)}(\bar{L}(t-1), \bar{A}(t-1), U_{L(t)}) \text{ for } t = 2, ..., 4 \\ A(t) &= f_{A(t)}(\bar{A}(t-1), \bar{L}(t), U_{A(t)}) \text{ for } t = 2, ..., 4 \\ Y &= f_Y(\bar{L}(4), \bar{A}(4), U_Y) \end{aligned}$$

   **Is the true $P_{U,X}$ compatible with the SCM presented?** Refer back to the second data structure of `R` Lab 1 for the true $P_{U,X}$.

2. Target causal parameter:

$$\Psi^F(P_{U,X}) = E_{U,X}[Y_{\bar{a}(4)}]$$

   **Is the target causal parameter (a parameter of $P_{U,X}$) identified (as a parameter of $P_0$) under the standard, point treatment randomization assumption/back door criteria? Why or why not?**

3. **If the target parameter is not identified in the previous question, what are the alternative assumptions under which the parameter would be identified?**

4. **What is the corresponding statistical estimand, $\Psi(P_0)$, under these assumptions?**

5. **Bonus!** Suppose instead your target causal parameter is $E_{U,X}[Y_{a(1)}]$. What is the interpretation of this target parameter? What assumptions are needed for identifiability here? What is the statistical estimand?

## Data Structure 4: $O = (L(1), C(1), A(1), Y(2), L(2), C(2), A(2), Y(3))$

1. SCM:

   $U = (U_{L(t)}, U_{C(t)}, U_{A(t)}, U_{Y(t+1)}), t = 1, 2 \sim P_U$. Assume $U$s are jointly independent.
   Structural equations, $F$:

$$
\begin{aligned}
L(1) &= f_{L(1)}(U_{L(1)}) \\
C(1) &= f_{C(1)}(L(1), U_{C(1)}) \\
A(1) &= f_{A(1)}(L(1), C(1), U_{A(1)}) \\
Y(2) &= f_{Y(2)}(L(1), C(1), A(1), U_{Y(2)}) \\
L(2) &= f_{L(2)}(L(1), C(1), A(1), Y(2), U_{L(2)}) \\
C(2) &= f_{C(2)}(\bar{L}(2), C(1), A(1), Y(2), U_{C(2)}) \\
A(2) &= f_{A(2)}(\bar{L}(2), \bar{C}(2), A(1), Y(2), U_{A(2)}) \\
Y(3) &= f_{Y(3)}(\bar{L}(2), \bar{C}(2), \bar{A}(2), Y(2), U_{Y(3)})
\end{aligned}
$$

   **Is the true $P_{U,X}$ an element of the SCM presented?** Refer back to the fourth data structure in `R` Lab 1 for the true $P_{U,X}$.

2. Target causal parameter:

$$\Psi^F(P_{U,X}) = E_{U,X}[Y(3)_{\bar{a}(2)=1,\bar{c}(2)=0}]$$

   **Is the target causal parameter (a parameter of $P_{U,X}$) identified (as a parameter of $P_0$) under the standard, point treatment randomization assumption/back door criteria? Why or why not?**

3. **If the target parameter is not identified in the previous question, what are the alternative assumptions under which the parameter would be identified?**

4. **What is the corresponding statistical estimand, $\Psi(P_0)$, under these assumptions?**

# 1 For Your Project: Identification

Think through the following questions and apply them to the dataset you will use for your final project.

1. Under what assumptions is the target causal parameter you came up with in the previous lab identified as a function of the observed data distribution?
2. What is your $\Psi(P_0)$, the statistical estimand?
3. Optional: confirm that in your simulation, the value of your estimand equals the value of your target causal parameter.

## 2 Feedback

Please attach responses to these questions to your lab. Thank you in advance!

1. Did you catch any errors in this lab? If so, where?
2. What did you learn in this lab?
3. Do you think that this lab met the goals listed at the beginning?
4. What else would you have liked to review? What would have helped your understanding?
5. Any other feedback?