

# 1 Final Paper Format

You will submit this as a group (not individually). Below we explain the general format, including section names and content. The use of generative AI to generate text for this paper is prohibited. If you use outside sources, you must cite them. This is due Monday, November 9th at 11:59pm. Please use the provided .Rmd template that is on the website. Papers should be at least 5 pages long, including the prompts in the template but not including code. Use the “Knit to PDF” option and submit the knitted PDF on Canvas.

## 1.1 Introduction

1. Describe your causal question in a way that someone who has not taken this class would understand. Why are you interested in this question? How could answering this question allow for better decision making? Include any necessary background or context. Cite outside sources you use.
2. Describe your causal question in the language of causal inference we’ve learned in this course: What is the treatment? What is the outcome? What are the potential outcomes? Write these out in words and in the math notation we have used in class.

## 1.2 DAG

1. Draw a DAG representing your causal question that includes at least three relevant variables besides treatment and outcome that are included in your dataset. You may include more than three variables. You may include variables that are not in your dataset, but at least 3 of your variables (excluding treatment and outcome) must be included in your dataset. If you use letters to denote variables, make sure they are clearly defined.
2. Explain your DAG: tell us in words what is meant by each edge in your DAG.
3. Discuss your DAG. How realistic is it? Are there variables or edges you excluded from your DAG that someone else might argue should be included? Playing devil’s advocate, how would you critique the reliability of your DAG?

## 1.3 Discussion: Method and Identification

1. What method are you using to estimate a causal effect? What causal effect are you estimating (ATE vs LATE vs ATT)? What assumptions are required to identify the causal effect via your chosen method?
2. Explain what conditional exchangeability means in the context of your causal question. Is it important? Why or why not? How do sufficient adjustment sets relate to conditional exchangeability?
3. Assuming your DAG is true, list out all non-causal paths between treatment and outcome and list one sufficient adjustment set to identify the causal effect of the treatment on the outcome. If a sufficient adjustment set does not exist, add additional variables to your DAG so that one does exist.
4. Discuss the plausibility of conditional exchangeability in your setting. If your sufficient adjustment set contains variables that are not in your dataset, discuss the implications.
5. Discuss any other identification assumptions for your method here, such as positivity and consistency. What do they mean in the context of your causal question and are they plausible?

## 1.4 Discussion: Analysis and Results

1. Give some context for your dataset. Who is included in your dataset? How was the data collected? When was the data collected? Make sure to cite the dataset.

2. Discuss any choices you made regarding data cleaning and processing: Did your data have missing values or outliers? How did you handle them? Were there any variables you dichotomized (i.e. made binary), or variables that you changed the format (e.g. yes/no to 1/0)?
3. Discuss the impact of any choices you made regarding your dataset, such as choices you made in data cleaning or processing.
4. Explain how you estimated a causal effect.
  - If you used matching, explain and discuss your choices. What formula did you use and why? What matching strategy did you use and why? Are there any advantages or drawbacks to the strategy you chose? How many units did your matching drop? How was the covariate balance in your matched sample? Discuss the implications of any choices you made and the quality of your matching.
  - If you didn't use matching, explain any choices you made related to the method you used and discuss their implications. Think about advantages or drawbacks to any choices you made, possible bias-variance trade-offs, and assessing how well your method did.
5. Report your causal effect estimate and interpret it in the context of your causal question.
6. Discuss the limitations of your analysis: what are the limitations of your dataset? Is there other data you would have wanted to have to bolster your analysis? Playing devil's advocate, how would you critique the reliability of your causal estimate?

## 1.5 Reproducible Code

1. This is where your code goes for the data cleaning/processing and analysis.
2. Make your code clean and easy-to-follow. Add short comments to explain what you are doing. If a classmate who wasn't as familiar with R were to read through this section, would they be able to follow along?

## 2 Rubric (100 Points Total)

### Overall Paper Structure and Formatting (10 points)

- 5 points: Clear organization, adheres to provided structure, appropriate page length
- 3 points: Proper formatting (e.g., .Rmd template, Knitted to PDF, proper citations).
- 2 points: Overall readability (grammar, spelling, and flow of writing).

### Introduction (15 points)

- 4 points: Clear explanation of the causal question understandable to a general audience.
- 4 points: Background/context is well-explained and relevant; sources cited if appropriate.
- 7 points: Accurate representation of the causal question using causal inference terminology, including mathematical notation from class.

### DAG (Directed Acyclic Graph) (20 points)

- 5 points: DAG is correctly drawn with all required variables (at least three besides treatment and outcome) clearly labeled.
- 5 points: Comprehensive explanation of the DAG, including meaning and relevance of edges.
- 10 points: Critical discussion of DAG plausibility / reliability.

### Discussion: Method and Identification (20 points)

- 4 points: Accurate explanation of the chosen method, causal effect type (ATE, LATE, ATT), and relevant assumptions.
- 4 points: Conditional exchangeability explained with clarity and in problem context; connection to adjustment sets addressed.
- 2 points: Correct sufficient adjustment set and non-causal paths.
- 10 points: Plausibility of assumptions like conditional exchangeability, positivity, and consistency (or whichever are relevant for their method) critically discussed.

**Discussion: Analysis and Results (20 points)**

- 2 points: Dataset context provided, with proper citation and brief description of collection methods.
- 3 points: Thoughtful discussion of data cleaning and processing choices; implications of these choices addressed.
- 3 points: Explanation of the causal effect estimation method, rationale for choices, and analysis quality.
- 2 points: Causal effect estimate reported and interpreted accurately.
- 10 points: Thorough discussion of the limitations of the analysis.

**Reproducible Code (15 points)**

- 10 points: Code is clear, well-organized, and includes explanatory comments for accessibility.
- 5 points: Code execution produces reproducible results consistent with the report.