

Students Jacoya Thompson (ID 2992892) and Victor S. Bursztyn (ID 3046613). Professor Jennie Rogers.

Project Proposal - [Revised](#)

Revision: This document has been revised following Professor's feedback. Please note the list of checkpoint-related questions starting on Page 2 (highlighted in blue).

Background: Minority residents of Chicago are up to 14 times more likely to be targeted for excessive and deadly force, according to the latest statistics [1]. Seven black people have been killed by police in 2019, and the fight for justice for people of color does not end after Jason Van Dyke. Our group attempts to investigate how socio-demographic features such as race and ethnicity could play a role in the **outcomes** of police misconduct investigations, rather than focusing on the **occurrences** of such misconducts.

Main Research Question: *When a case of police misconduct victimizes groups that are historically marginalized, how unfavourable is its investigation?*

This single question unfolds over a series of hypothetical, measurable effects on the investigation. When compared to non-marginalized groups, to some extent it could:

- Effect 1: Slow down a disciplinary investigation;
- Effect 2: Decrease the likelihood of the officer(s) involved being disciplined;
- Effect 3: Decrease the likelihood of the state granting compensation.

Methodology: It's challenging to use observational data to study causality. For instance, we could find that a particular socio-demographic feature is predictive of investigations going forward, but they could be highly correlated to the severity of the misconduct — which could be the true cause here. As seen in Fig. 1, we should try to control for lurking variables while designing our checkpoint-specific questions.

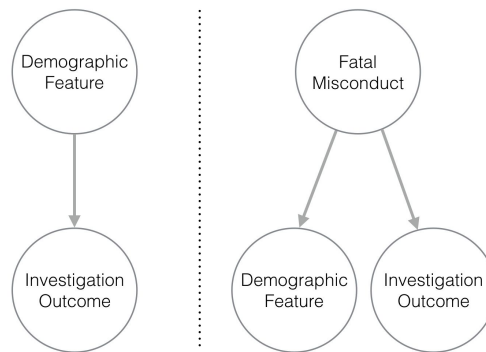


Fig. 1 — A lurking variable as the true cause of the phenomenon.

Checkpoints and Checkpoint-Specific Questions: This proposal is organized around five checkpoints, each with its own set of questions that contribute to the pursuit of the main research question.

1. Checkpoint #1 — Relational Analytics:

- a. *Question #1: What's the percentage of allegations that are sustained?*
- b. *Question #2: What's the average investigation time measured from its start to the first decision?*
- c. *Question #3: What's the most common type of disciplinary action? Implementation details: Retrieve the mode.*
- d. *Question #4: What's the average compensation amount measured from settlements data?*

2. Checkpoint #2 — Visualization:

- a. *Tableau Chart #1: What's the distribution of investigation times?*
- b. *Tableau Chart #2: What's the frequency of each disciplinary action?*
- c. *Tableau Chart #3: What's the distribution of compensation amounts?*
- d. *Interactive Visualization #1: What's the influence of victim race on the first two distributions? Implementation details: Add an interactive dimension "victim race" to Tableau Charts #1 and #2. We can't address Tableau Chart #3 before data cleaning and integration (see Checkpoint #3).*

3. Checkpoint #3 — Data Cleaning and Integration:

- a. *Question #1: What's the average compensation amount by victim race?*
- b. *Question #2: What's the average compensation amount by allegation type?*
- c. *Question #3: What's the average compensation amount by victim race and allegation type combined?*
- d. *Question #4: What's the average compensation amount by victim race, allegation type and case's judge combined?*

4. Checkpoint #4 — Graph Analytics:

- a. *Question #1: How allegation types connect in a graph where edges represent allegations committed by the same officer? Can we identify different levels of severity in the communities (clusters) of allegations?*
- b. *Question #2: Do the communities match scales for crime severity discussed in the literature?*

5. Checkpoint #5 — Machine Learning and Text Analytics:

- a. *Question #1: How predictive the features victim race, allegation type and case's judge are in predicting the investigation time?*
- b. *Question #2: How predictive the features victim race, allegation type and case's judge are in predicting the disciplinary action?*
- c. *Question #3: How predictive the features victim race, allegation type and case's judge are in predicting the disciplinary action?*

References:

[1] "Chicago Police are 14 times more likely to use force against black men than against whites." — <https://theintercept.com/2018/08/16/chicago-police-misconduct-racial-disparity/>