CHANCE

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ucha20

How to Use Causal Inference to Study Use of Force

Jennifer Wyatt Bourgeois, Anna Haensch, Saatvik Kher, Dean Knox, Gregory Lanzalotto & Tian An Wong

How to Use Causal Inference to Study Use of Force

Jennifer Wyatt Bourgeois, Anna Haensch, Saatvik Kher, Dean Knox, Gregory Lanzalotto, and Tian An Wong

In the fields of health, social, and behavioral sciences, most research is focused on understanding what causes something to happen rather than just identifying patterns of association. In the context of law enforcement, causal inference can be used to assess factors impacting the use of force by police officers while rigorously accounting for challenges in the collection of policing data. Generally, law enforcement officers are allowed to use force when it is deemed necessary, such as for self-defense or to protect another person or group. There is not a standardized use of force definition; however, according to the International Association of Chiefs of Police[1], use of force is considered "effort required by police to compel compliance by an unwilling subject." Based on the amount of force used, it is categorized as either justified or excessive. One widely used guideline is a "force matrix" determining the appropriate amount of force to use, on a continuum ranging from verbal commands to lethal force, in response to various civilian behaviors, which can range from cooperative compliance to noncompliant actions that could harm others[2]. The intensity of force used by an officer depends on the specifics of the situation, such as the officer's training and experience.

Before applying causal inference techniques to the study of police use of force, it is useful to frame the discussion within the established legal precedents. The framework for evaluating police conduct is shaped by landmark US Supreme Court decisions such as *Tennessee v. Garner* (1985)[3] and *Graham v. Connor* (1989)[4]. Both cases establish the legal standards that guide the assessment of police actions. *Tennessee v. Garner* limits the use of deadly force to circumstances where there is a substantial threat to the officer or others, creating a foundational standard for the application of lethal force. Additionally, *Graham v. Connor* utilizes the "objective reasonableness" standard, evaluating an officer's decisions to use force based on what a reasonable officer might do under similar circumstances, without the clarity provided by hindsight.

Within this established framework, it becomes meaningful to study factors impacting the use of force, that is, the extent to which force is actually used (i.e., how often, and how much force) relative to how often it could have been used. This is a question that is at the heart of causal inference. By comparing these quantities, it is possible to understand instances where use of force is excessive or discriminatory and also how individual officer behavior and department policies impact use of force.

As an illustrative example, consider the case of Portland, Oregon. According to a 2019 report, 45 out of every 1,000 arrests of Black people involved the use of force, while only 31 out of every 1,000 arrests of white people involved use of force[5]. Additionally, 29 percent of all use of force incidents were carried out on Black people, while only 8 percent of Portland's residents identify as Black. If we consider the variable of interest, or outcome, as experiencing a use of force event, the report suggests differential treatment across demographic groups. However, it does not immediately suggest the reason for these differences. To understand the differential treatment of different groups, in this case racial groups, it is necessary to understand the causal mechanism behind these different

[1] International Association of the Chiefs of Police. 2001. Police Use of Force in America, 2001. Alexandria, Virginia.
[2] *https://nij.ojp.gov/topics/articles/use-force-continuum*
[3] *Tennessee v. Garner*, 471 U.S. 1 (1985)
[4] *Graham v. Connor*, 490 U.S. 386 (1989)
[5] Training Advisory Council. 2020. Patterns in Portland Police Bureau Force Data Summary Reports. *https://s3.documentcloud.org/documents/7016734/PPB-Report-Patterns-in-the-Use-of-Force-7-2020.pdf*

outcomes. More formally, it is critical to identify and correctly control for factors that might conceal true relationships between an observed event (e.g. the application of force by a law enforcement officer) and its potential causes (e.g. behavior of the suspect, existence of a perceived threat, or environmental factors). By using techniques such as propensity score matching or regression discontinuity designs, researchers can more accurately isolate the effects of specific actions or policies on both the probability and characteristics of force use.

In the realm of public policy, causal inference is also useful for impact assessment to determine whether a policy achieves its stated goals. Specifically, causal inference is a process to determine the difference in outcomes that are directly due to a policy implementation, by separating these outcomes from counterfactuals (i.e., outcomes that would have occurred in the absence of the policy intervention). For example, a department hoping to decrease the prevalence of force incidents might consider alternative deployment and dispatching strategies, such as deploying more experienced officers to neighborhoods with high rates of violent crime or diverting calls to ancillary emergency agencies such as EMT or behavior health units. A successful policy would be one that simultaneously minimized inappropriate police use of force and other public safety tradeoffs. Causal inference methods for policy evaluation can help assess whether these goals are attained by disentangling effects of the policy from other confounding factors (e.g., pre-existing trends or other characteristics of treated and control neighborhoods). These techniques can be applied to study the effects of police training programs, community policing initiatives, and use-of-force policies within the framework of the described legal standards. By systematically evaluating these factors, we can pinpoint what truly works in reducing unnecessary and excessive use of force, thereby informing more effective policing strategies and ensuring that legal guidelines are not only followed but supported by evidence-based practices.

Available Data

Before we focus on the specific statistical methods used to conduct causal inference using police data, we will outline the data sources used in these analyses. Robust causal inference relies not only on sophisticated statistical techniques but also on high-quality, comprehensive data. These data sources enable researchers to make precise comparisons and draw valid conclusions about the factors influencing police behavior.

The main source of data is police record management system data, with a 2020 estimate indicating that 87.2 percent of US police departments use an RMS system[6]. Police track many of the actions they perform each day, creating detailed records that researchers can exploit. Kept by each department, RMS data includes time- and location-stamped data sets for each arrest, citation issued, and incident of force. These data also include information about the civilian(s) and officer(s) involved in each record, and they usually have some indication of civilian and officer behavior. The granularity of these records facilitates precise comparisons, thereby enhancing the validity of inferences.

Beyond RMS incident tracking, much of an officer's day is also tracked in calls for service data. CFS data includes responses to dispatch-initiated calls (answering 911 calls) and self-initiated calls, such as when an officer makes a traffic stop. This data includes officer IDs, start and end times for each call, location, and call type/call code, which indicates what the officer is doing. This data can usually be linked to the incident records described above, further allowing close comparisons.

At a coarser level, shift records include start and end times, where the officer is deployed (usually a beat, or an area the officer is supposed to patrol), and sometimes additional information about assigned officer tasks.

These three types of data sets (incident, call, and shift records) form the bulk of the data used in a causal analysis, but there are still many other types of policing data that can be used. For example, roster data can provide additional identifying information about officers, and internal affairs records can show civilian complaints and disciplinary information.

Additionally, body-worn camera data has become a common source of police data over the past decade. Many departments now utilize BWCs, and the audiovisual data these devices generate is ripe for use in causal inference research. Further discussion on inference using BWCs appears in the future directions section.

The data described thus far is mostly kept by individual departments, and thus different departments have different standards for sharing it. Many departments will require a research collaboration to share, but due to public pressure to increase transparency, more departments are becoming open to releasing data publicly[7].

[6] US Department of Justice. LEMAS 2020 survey, *https://bjs.ojp.gov/document/lpdppt20st.pdf*

[7] *www.policedatainitiative.org*

A crucial consideration for police-generated data is its inherent slant, showing only the police perspective by default. Police may recall events differently (e.g., descriptions of civilian behavior) compared to bystanders or civilians involved. Moreover, it is standard practice for officers to report events selectively. Documentation rarely exists for encounters where officers choose not to initiate a stop; similarly, stops that end in a verbal warning (as opposed to a citation or arrest) often go unreported as well. Body-worn camera data introduces an added accountability layer, though it is computationally challenging to study at scale. Civilian records of police interactions are less common than police records; however, the Police-Public Contact Survey, a supplement to the National Crime Victimization Survey[8], provides insights into citizens' perceptions of police behavior and responses during these encounters.

Beyond department-provided data, other types of data can supplement an analysis. For instance, mobile location data collected by companies such as SafeGraph[9] can help contextualize use of force within specific places and community dynamics. Additionally, crime data from hospitals, emergency rooms, victimization surveys, and gunshot detection[10] data provide objective measures of violent incidents and their outcomes. Census data such as the American Community Survey can be used to identify the demographic composition of different areas, adding another layer of context to the analysis. Furthermore, traffic sensors offer ground-truth measures of traffic violations.

By integrating these varied data sources, researchers can conduct robust causal analyses to uncover the complex dynamics underlying police use of force, ultimately contributing to more informed and effective policy decisions.

Limitations of Existing Statistical Methods

A key challenge in statistical analysis is that researchers typically possess only a small portion of the data needed to estimate the causal quantities of interest reliably, which include the following:

- the proportion of police encounters in which force was excessive, compared to a department's stated use-of-force policy
- the extent to which minority civilians are subject to discriminatory force, compared to similarly situated white civilians
- the difference in how often one officer uses force, compared to peer officers facing similar pools of civilian behavior
- changes in any of the above, either over time or because of a policy intervention

Each of these quantities involves not only how often force was *actually* used (a numerator) but also how often force *could have* been used (a denominator). This poses a problem for statistical analysis, because police data-recording practices often focus on recording numerators rather than denominators.

For example, police agencies have extensive requirements for documenting the circumstances around uses of force that meet some (agency-specific) reporting threshold, particularly any civilian behavior that might be used to justify officers' decisions. But data collection on non-uses of force is far laxer: In some agencies, officers are not required to document officer-initiated pedestrian or traffic stops that do not involve force or other enforcement actions (e.g. citation, arrest). This can lead to serious errors, as illustrated by the retraction of a prominent *Proceedings of the National Academy of Sciences*[11]. Using only force records, analysts estimate demographics among civilians subjected to force, $p(\text{race} \mid \text{force}, X)$, which represents the likelihood of a civilian's race given that force was used and some additional information such as civilian behavior (X). However, questions about racial discrimination revolve around the rate at which force is used against similarly situated civilians, which are disparities in $p(\text{force} \mid \text{race}, X)$. Without knowing the denominator of how often analysts encountered civilians of each group, $p(\text{race} \mid X)$, analysts are limited in what they can learn about discrimination in force. However, this concern can be addressed by careful statistical thinking about the quantity of interest, for example by focusing on cases with a clear denominator (such as the civilian-initiated calls for service) or by using proxies for the appropriate denominator with clearly stated assumptions and appropriate caveats

[8] *https://bjs.ojp.gov/data-collection/police-public-contact-survey-ppcs*
[9] *www.safegraph.com*
[10] *www.soundthinking.com/law-enforcement/leading-gunshot-detection-system*
[11] Retraction for Johnson et al., Officer characteristics and racial disparities in fatal officer-involved shootings. *Proceedings of the National Academy of Sciences* 116:15877–15882. 10.1073/pnas.1903856116.

(such as mobile location data as a proxy for the demographics of civilians in a particular place and time, as well as traffic crashes as a proxy for the demographics of dangerous drivers that might be subject to a traffic stop).

Even data on police stops does not fully address this issue because officers often have discretion in whether a stop is initiated. Because every police–civilian encounter represents one potential use of force, distortions are introduced into the denominator when officers choose to let a civilian pass without a formal stop (with an informal warning or simply by taking no action at all). Knox, Lowe, and Mummolo[12] show how selective documentation can distort regression analyses and lead to underestimates of key quantities, because officers that discriminate in force use are also likely to discriminate in the initial decision to stop. To see why, suppose officers encounter 100 minority and 100 white civilians engaged in disorderly conduct. Further suppose that there are undocumented aggravating circumstances in 50 encounters with each group, and officers discriminate by using force in all 50 aggravated encounters with minority civilians but only 25 aggravated encounters with white civilians. If officers stop minority civilians for any disorderly conduct, then the minority use-of-force rate would be 50/100 = 0.5. If they discriminate by only stopping white civilians for the higher aggravated threshold, the white force rate in the selectively documented stop data would appear to suggest no discrimination, at 25/50 = 0.5. However, the correct comparison set would either be 50/100 = 0.5 minority to 25/100 = 0.25 white (among all encounters) or 50/50 = 1.0 minority to 25/50 = 0.5 white (among aggravated encounters), both of which would reveal discrimination. Thus, analysts who were not aware of selection bias may inadvertently draw the wrong conclusions.

To address these and other limitations of police data, Knox (2021)[13] calls for the broader use of causal-inference techniques to address data inaccuracies, selective reporting, and omitted variables. These may include data fusion with third-party sources (e.g. traffic sensors or mobile location data), follow-up data collection (e.g. using random body-worn camera audits to validate self-reported officer data), as well as sensitivity analyses and sharp bounds that capture best- and worst-case scenarios while transparently acknowledging that analysts possess incomplete information.

Future Directions

The preceding discussion clarifies that much work remains to be done in applying causal inference methods to the study of use of force by law enforcement. We conclude with some indications as potential future directions for research in this area.

At a broad level, a more expansive approach to these research questions might be carried out by combining the analysis of police-generated data sets with other sources mentioned above, such as public surveys such as the Bureau of Justice Statistics' Police-Public Contact Survey. At the same time, each new data set brings with it new methodological concerns regarding the internal validity and generalizability of conclusions across various contexts and scales. Nonetheless, drawing critically from new sources can provide new information that may be used to bolster analysis.

At a methodological level, careful implementation of statistical methods is necessary in order to be able to draw accurate conclusions about the evaluation in question. Besides the issues discussed above, other common sources of error include in the use of statistically biased estimators for time-series data[14,15], and in the handling of potentially non-random missingness in administrative data sets.

For these and many other reasons, randomized controlled trials remain the gold standard for evaluating interventions. However, buy-in from departments is not a given and requires consideration[16]. Given the high stakes and difficulty of implementation, experimental design is crucial. For example, multiple policing reforms

[12] Knox, D., Lowe, W., Mummolo, J. 2020. Administrative records mask racially biased policing. *American Political Science Review* 114(3):619–637.
[13] Knox, D. 2021. Revealing racial bias. *Science* 374(6568):701–702.
[14] Imai, K., Kim, I.S., Wang, E. H. 2023. Matching methods for causal inference with time-series cross-sectional data. *American Journal of Political Science* 67(3):587-605.
[15] Arkhangelsky, D., Imbens, G.W. 2023. Fixed effects and the generalized Mundlak estimator. *Review of Economic Studies*, rdad089. Imai, K., Kim, I. S. 2019. When should we use unit fixed effects regression models for causal inference with longitudinal data? *American Journal of Political Science* 63(2):467-490.
[16] Goerger, S., Mummolo, J., Westwood, S.J. 2023. Which police departments want reform? Barriers to evidence-based policymaking. *Journal of Experimental Political Science* 10(3):403-412.

About the Authors

Jennifer Wyatt Bourgeois is a postdoctoral fellow at the Center for Justice Research at Texas Southern University.

Anna Haensch is affiliated with the Tufts University Data Intensive Studies Center.

Saatvik Kher is a PhD student of computer science at the University of California, Irvine.

Dean Knox is an assistant professor of operations, information, and decisions at The Wharton School of the University of Pennsylvania.

Gregory Lanzalotto is a PhD student of operations, information, and decisions at the Wharton School of the University of Pennsylvania.

Tian An Wong is an assistant professor of mathematics and statistics at the University of Michigan-Dearborn.

implemented simultaneously can obscure policy evaluations (e.g. changing use of force definitions while implementing reforms to reduce force). Moreover, experiments that deliver a policy or information treatment to some squads or precincts may have spillover effects if officers communicate with peers from other units. Numerous causal inference techniques have been developed to address such issues, but their use often requires RCTs to be designed in particular ways that introduce additional complexity in implementation. Finally, it is worth noting that police use of force is a setting where policy changes can lead to unintended consequences, including possible harm to civilians—not only through inadvertent increases in the use of force but also through adverse effects on public safety through other avenues—an ethical consideration that must be carefully weighed against potential benefits before proceeding with RCTs.