Body cameras are seen as a new means of documenting police-civilian interactions, but their use—and, at times, convenient disuse—can reflect bias as well.

PRIZE ESSAY

SOCIAL SCIENCES

Revealing racial bias

Causal inference can make sense of imperfect policing data

By **Dean Knox**

For decades, high-profile incidents of excessive force against minorities have fueled allegations of abusive policing in the United States and demands for reform. Yet one of the main drivers of today's policing crisis remains unchanged: massive racial disparities in law enforcement.

Courts and city councils struggle to measure the severity of racial bias in policing, let alone to identify the means to address such bias. Solutions are difficult to identify because the policing data landscape is fraught with inconsistent record-keeping and incomplete, task-specific datasets. In examining the dizzying array of analytic approaches used in this context, my colleagues and I found many to be mutually incompatible or even misleading, producing contradictory results and impeding knowledge accumulation (*1*–*3*). Making use of formal statistical frameworks for drawing causal inferences (*4*, *5*)—that is, reliable conclusions about how and why events occur, given explicitly stated assumptions and observed data—we have shown the importance of measuring and accounting for the long chain of events from officer deployment to contact, detainment, and violence.

NOMIS & Science YOUNG EXPLORER AWARD

Policing presents substantial challenges for statistical analysis. For almost 100 years, police agencies have been the sole source of data on police-civilian interactions (*6*). Administrative datasets only document incidents that were required to be reported—historically, violent and/or property crimes. Agencies now increasingly also report stops, frisks, arrests, and uses of force against civilians. Still, only a smattering of interactions are documented.

Why does this matter? The application of off-the-shelf statistical methods to "datasets of convenience"—that is, datasets focused solely on obtainable information, without considering what variables or observations might not be obtainable—often leads to fragile conclusions hinging on implausible or unstated assumptions (*7*). Similar challenges arise when analyzing datasets acquired through open record requests (*1*) or labor-intensive crowdsourcing (*2*).

An increasingly important subfield of statistics and computer science, causal inference, aims to address this issue. Causal inference focuses on a deceptively simple question: Where do our data come from? From this starting point, we can build frameworks (*4*, *5*, *8*, *9*) for analyzing datasets contaminated by inaccuracies, selective reporting, and omitted variables. Rather than ignoring imperfections in our data, we ask what the range of possible interpretations is and what new information must be collected to further narrow this list.

In disciplines facing similar challenges, such as medicine, where patients rarely

Operations, Information and Decisions Department, University of Pennsylvania, Philadelphia, PA 19014, USA. Email: dcknox@upenn.edu

PHOTO: SEBASTIAN KAHNERT/PICTURE ALLIANCE/GETTY IMAGES visit doctors unless something has gone wrong, the causal-inference toolkit has proven invaluable for academics and policy-makers (*10*). Yet in policing research, careful causal analysis remains the exception, not the rule.

Failure to account for unobserved causal processes has frequently undermined our understanding of policing. Decades of research have analyzed police detainment records to compare treatment of white and minority detainees (*11*, *12*). Scholars frequently conclude that there is surprisingly little evidence of discrimination. However, our work shows that without accounting for an officer's initial decision to detain, an act that may itself reflect bias, analysts have little hope of recovering accurate estimates of discrimination (*1*).

This initial decision is relevant because minorities may be detained in situations where white individuals would not be detained, for example, jaywalking encounters. By contrast, individuals engaged in behaviors like assault will typically be detained irrespective of race. Comparisons of minority detainments to white detainments therefore fail to achieve the apples-to-apples conditions needed to demonstrate disparate treatment (*7*). Studies that ignore the underlying processes inherent in the formation of these datasets can thus sharply underestimate discrimination in police violence or mask it entirely.

To aid others studying policing datasets, we have developed techniques for estimating bounds—best- and worst-case levels of discrimination—consistent with the available, imperfect data. Bounding approaches consider the ways that unobserved phenomena may manifest in a dataset, for example, the decision to detain; thus, they can yield conclusions that are robust to those unobserved parameters. Similar causal bounding approaches are widely used in other domains, like epidemiology (*13*), but have appeared in policing only recently (*3*).

Crowdsourced datasets of police killings (*14*, *15*) increasingly provide an objective check on officer self-reports. Yet our work shows that challenges arise even when studying civilian-collected datasets, because most nonviolent incidents are not recorded. The consequences of this lack of data were recently highlighted by a flawed, high-profile study that erroneously concluded "if anything, [we] found anti-White disparities" in police violence (*16*). The study, which went on to dismiss widely proposed police diversification reforms, bolstered antireformers (*17*), but our analysis showed that its claims were mathematically baseless, and the paper was ultimately retracted (*2*, *18*).

Our own study of diversity in the Chicago Police Department illustrates the difficulty of conducting research in this area and how causal reasoning can be of aid. Using data on 2.9 million officer shifts, we found that officers from marginalized groups engage in substantially less violence, particularly toward minority civilians. These officers also engage in less discretionary enforcement for minor offenses (*19*). Such an analysis is feasible thanks to detailed patrol records that let us compare officers rotating through similar places at similar times with similar numbers and types of encounters. When simulating typical data constraints in prior work, which relied on enforcement records alone, we found that a failure to account for uneventful shifts could lead to inaccurate inferences based on an incomplete dataset.

As concerns about policing have exploded, policy-makers and the general public are turning to academics to make sense of its complexities. However, a literature filled with contradictory results has often led scholars to say that "we simply do not know" (*20*). Our causal analyses of common research methods reveal the roots of this confusion. More importantly, they suggest a path forward as the discipline grows.

We are now developing a system for reconstructing encounter timelines from body-worn camera footage, building on my prior work on conversation and vocal-tone analysis in audio data (*21*, *22*) and incorporating our collaborators' expertise in computer vision (*23*, *24*). We continue to develop new causal methods for enabling the systematic evaluation of officers and agencies by fusing existing, imperfect datasets (*25*). Above all, we demonstrate that a shared language—a coherent causal framework—is needed to evaluate evidence, adjudicate contradictory claims, and accumulate knowledge in this critically important domain (*3*). ■

WINNER

Dean Knox

Dean Knox received his undergraduate degree from the University of Illinois at Urbana-Champaign and a PhD from the Massachusetts Institute of Technology. After completing his postdoctoral fellowship at Microsoft Research, he joined the Politics Department at Princeton University as an assistant professor in 2018. Dean is currently an assistant professor at the Operations, Information and Decisions Department at the Wharton School of the University of Pennsylvania, where he co-founded the Research on Policing Reform and Accountability group with Jonathan Mummolo. His research develops statistical methods for analyzing imperfect social science data.

FINALIST

Geoffrey Supran

Geoffrey Supran received his undergraduate degree from Trinity College, University of Cambridge, and a PhD from the Massachusetts Institute of Technology (MIT). After completing joint postdoctoral fellowships at MIT and Harvard University, Geoffrey became a research fellow in the Department of the History of Science at Harvard in 2019 and also Director of Climate Accountability Communication at the Climate Science Social Network in 2020. His research focus is the quantitative historical analysis of climate change disinformation and propaganda by fossil fuel interests.

science.org/doi/10.1126/science.abm3434

REFERENCES AND NOTES

1. D. Knox, W. Lowe, J. Mummolo, *Am. Polit. Sci. Rev*. **114**, 619 (2020).
2. D. Knox, J. Mummolo, *Proc. Natl. Acad. Sci. U.S.A*. **117**, 1261 (2020).
3. D. Knox, J. Mummolo, *J. Polit. Inst. Polit. Econ*. 2020, **341** (2020).
4. J. Pearl, *Biometrika* **82**, 669 (1995).
5. D. B. Rubin, *J. Educ. Psychol*. **66**, 688 (1974).
6. International Association of Chiefs of Police, Uniform Crime Reports (1930).
7. D. Knox, W. Lowe, J. Mummolo, SSRN 3940802 (2020).
8. P. W. Holland, *J. Am. Stat. Assoc*. **81**, 945 (1986).
9. J. Neyman, "Sur les applications de la theorie des probabilites aux experiences agricoles: Essai des principes," thesis, Univ. of Warsaw (1923).
10. J. M. Robins, *Math. Model*. **7**, 1393 (1986).
11. R. G. Fryer Jr., *J. Polit. Econ*. **127**, 1210 (2019).
12. D. A. Smith, C. A. Visher, L. A. Davidson, *J. Crim. Law Criminol*. **75**, 234 (1984).
13. C. E. Frangakis, D. B. Rubin, *Biometrics* **58**, 21 (2002).
14. D. B. Burghart, Fatal encounters. Database (2012); https://fatalencounters.org/.
15. *Washington Post*, Fatal force. Database (2015); www.washingtonpost.com/graphics/investigations/police-shootings-database/.
16. D. J. Johnson, T. Tress, N. Burkel, C. Taylor, J. Cesario, *Proc. Natl. Acad. Sci. U.S.A*. **116**, 15877 (2019).
17. H. Mac Donald, Written testimony before the Committee on the Judiciary of the United States House of Representatives, Oversight Hearing on Policing Practices, 19 September 2019; https://docs.house.gov/meetings/JU/JU00/20190919/109952/HHRG-116-JU00-Wstate-MacDonaldH-20190919.pdf.
18. D. Knox, J. Mummolo, "Prominent claims that policing is not racially biased rest on flawed science," Medium (2020); https://medium.com/@jon.mummolo/prominent-claims-that-policing-is-not-racially-biased-rest-on-flawed-science-6f66535dc7e5.
19. B. A. Ba, D. Knox, J. Mummolo, R. Rivera, *Science* **371**, 696 (2021).
20. D. A. Sklansky, *J. Crim. Law Criminol*. **96**, 12091244 (2005).
21. D. Knox, C. Lucas, *Am. Polit. Sci. Rev*. **115**, 649 (2021).
22. S. A. Mehr *et al*., *Science* **366**, eaax0868 (2019).
23. O. Russakovsky *et al*., *Int. J. Comput. Vis*. **115**, 211 (2015).
24. S. Yeung *et al*., *Int. J. Comput. Vis*. **126**, 375 (2018).
25. G. Duarte, N. Finkelstein, D. Knox, J. Mummolo, I. Shpitser, arXiv:2109.13471 (2021).

10.1126/science.abm3432

Revealing racial bias

Dean Knox

Science, 374 (6568),

View the article online
https://www.science.org/doi/10.1126/science.abm3432
Permissions
https://www.science.org/help/reprints-and-permissions

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.