proximately 1.6 gigatonnes in the early 21st century. Regions in East and Central Asia, Central Africa, and North and South America have shown substantial depletion in soil moisture. The findings also indicate that lost terrestrial water has not recovered to previous levels. This persistent decline suggests that the negative shift in soil moisture may be irreversible because of prolonged drought conditions and reduced precipitation in certain regions.

Although Seo *et al*. provide an analysis of global terrestrial water variations over the past two decades, a broad range of factors that influence precipitation and evapotranspiration (the transfer of water from land to atmosphere through evaporation and transpiration) must also be considered. Advanced land surface and hydrological models that can accurately represent these factors under the influence of changing climate are crucial to capturing the evolution of terrestrial water storage.

The findings of Seo *et al*. underscore the urgent need to improve parameterization of land surface models to better understand complex geophysical problems (*14*). Developing next-generation models that incorporate anthropogenic influences such as farming, large dams, and irrigation systems (*3*) is essential. The ongoing advancements of a land surface modeling system by the European Centre for Medium-Range Weather Forecasts represents a promising step forward (*15*). These improvements could reduce uncertainties and enhance our understanding of the impacts of climate change on the global water cycle. ■

REFERENCES AND NOTES

1. L. Samaniego, R. Kumar, M. Zink, *J. Hydrometeorol.* **14**, 47 (2013).
2. G. Blöschl *et al.*, *Science* **357**, 588 (2017).
3. K.-W. Seo *et al.*, *Science* **387**, 1408 (2025).
4. M. Yamaguchi, M. Furuya, *Earth Planets Space* **76**, 1 (2024).
5. J. Kuehne, C. R. Wilson, *J. Geophys. Res.* **96**, 4337 (1991).
6. R. Gross, *J. Geophys. Res.* **27**, 2329 (2000).
7. T. M. Eubanks, in *Contributions of Space Geodesy to Geodynamics: Earth Dynamics*, vol. 24, Geodynamics Series (American Geophysical Union, 1993), pp. 1–54.
8. R. Baatz *et al.*, *Rev. Geophys.* **59**, e2020RG000715 (2021).
9. H. Hersbach *et al.*, *Q. J. R. Meteorol. Soc.* **146**, 1999 (2020).
10. E. F. Wood *et al.*, *Water Resour. Res.* **47**, W05301 (2011).
11. M. F. P. Bierkens *et al.*, *Hydrol. Processes* **29**, 310 (2015).
12. O. Rakovec, R. Kumar, S. Attinger, L. Samaniego, *Water Resour. Res.* **52**, 7779 (2016).
13. O. Rakovec *et al.*, *J. Hydrometeorol.* **17**, 287 (2016).
14. L. Samaniego *et al.*, *Hydrol. Earth Syst. Sci.* **21**, 4323 (2017).
15. S. Boussetta *et al.*, *Atmosphere* **12**, 723 (2021).

ACKNOWLEDGMENTS

The author acknowledges the 4DHydro initiative, which is funded by the European Space Agency (ESA) under contract 4000141141/23/I-EF.

10.1126/science.adw5851

SOCIAL SCIENCE

New data fill long-standing gaps in the study of policing

Data show discrimination, but analysis must be more policy relevant

By **Dean Knox**[1] *and* **Jonathan Mummolo**[2]

Data limitations have long stymied research on racial bias in policing. To persuasively demonstrate bias, scholars have sought to compare officer behavior toward minority versus white civilians while holding constant all other factors in the police-civilian encounter that might provide alternative explanations for enforcement disparities. These comparisons in "similar circumstances" are also critical in litigation concerning discriminatory policing, which can often lead to court-ordered remedies (*1*). Such "all-else-equal" scenarios are elusive in many realms of social science, but two challenges have made them particularly difficult to find in the study of policing. On page 1397 of this issue, Aggarwal *et al.* (*2*) report using data from the ridesharing service Lyft—having obtained vehicle location on more than 200,000 drivers using high-frequency GPS pings from their smartphones—to analyze speeding enforcement by the Florida Highway Patrol (FHP) and to show how such data offer a path forward for addressing both challenges.

One challenge to establishing all-else-equal comparisons in studies of policing is that standard police datasets contain one-sided officer accounts of civilian behavior, which past work has shown do not always accurately measure actual driver behavior. For example, prior research (*3*) has shown that in the same FHP context, officers gave white drivers a "discount" on tickets by reporting lower speeds relative to the speeds reported for minority drivers. Aggarwal *et al.* use the Lyft data to construct an objective measure of speeding behavior.

The other challenge is that traditional police-generated datasets are inherently selective: For example, they do not contain every police-civilian encounter in which an officer could have cited a speeding driver but rather only the subset in which an officer chose to pull vehicles over and therefore had to fill out forms documenting the stop. Research has shown that under reasonable assumptions, if there is racial bias in the initial decision to stop, then analyses that take the resulting stop data at face value—e.g., using them in standard regression analyses—can substantially underestimate racial bias in subsequent decisions, such as whether to issue a citation (*4*). Intuitively, this is because differential selection into the data means that—even if encounters appear to be similarly situated on observed characteristics—stops of minority and white civilians will likely differ on unobservable characteristics that influenced officers' stopping decisions. The approach of Aggarwal *et al.* resolves this challenge too, by allowing researchers to observe all times when Lyft drivers are active rather than only the selected sample of those where officers chose to detain them.

Using these rich data, Aggarwal *et al.* find statistically significant evidence of discriminatory policing even among Lyft drivers incentivized to drive safely—a subgroup where effects are likely to be lower than those observed in the general driving population. How consequential is this discrimination in substantive or legal terms? The answer is more complex than it may appear and hinges on the specifics of how the statistical estimand—the quantity of interest targeted by the research design and analysis—relates to legal standards of evidence in discrimination cases. This connection is one that academic work on policing, including Aggarwal *et al.*, has rarely considered. Nevertheless, it is critical to ensuring that scientific research translates into real-world impact.

On their face, the disparities demonstrated by Aggarwal *et al.* appear substantively small: The largest estimates of racial discrimination translate to roughly one additional speeding citation per 28 years of full-time driving and roughly one additional dollar in fines per year. But these results cannot be correctly interpreted without considering both the circumstances in which drivers are observed and the precise comparisons on which these estimates are based.

[1]Operations, Information and Decisions Department, Wharton School, University of Pennsylvania, Philadelphia, PA, USA. [2]Department of Politics, School of Public and International Affairs, Princeton University, Princeton, NJ, USA. Email: jmummolo@princeton.edu

Aggarwal *et al*. report racial discrimination in policing using vehicle location data on more than 200,000 drivers.

PHOTO: LOWPHOTO/ALAMY STOCK PHOTO

On the former point, drivers observed in the Lyft data reported by Aggarwal *et al.* rarely speed to any notable extent. In 85.3% of pings, drivers are not speeding at all—meaning that FHP cannot defensibly issue a speeding citation, let alone discriminate in doing so. Another 11.3% of pings involve only slight speeding violations, which are rarely enforced. This means that aggregate-level differences that initially seem small are in fact driven by far larger differences in the most relevant 3.4% of observations where notable speeding occurred. In seeking to address discriminatory law enforcement, civil rights practitioners litigating similar issues have therefore focused on scenarios where officers might plausibly exhibit the behavior of interest, rather than reporting estimates that are diluted by the 96.6% of scenarios where discriminatory enforcement is almost mechanically precluded (*5*). For example, a recent US Department of Justice (DOJ) investigation of the Phoenix Police Department in Arizona estimated racial bias in ticketing "Among drivers who speed…" and "Among drivers who engage in low-level moving violations…" [(*6*), p. 60]. Focusing on these potentially enforceable cases requires a different statistical estimand.

The latter point is a more subtle one that touches on the relationship between statistical analysis and discrimination case law, which focuses not on differences between the treatment of two groups in general but rather on demonstrating harm to one protected group specifically. For example, a DOJ legal manual on the topic states, "To prove…systemic discrimination…A plaintiff in a pattern or practice case can…[present] statistical evidence of similarly situated individuals not in the protected class who were treated better than those in the protected class" [(*7*), p. 22]. The focus is on realized harm to actual minority drivers, in the sense that (according to a statistical analysis) they were treated worse than a hypothetical group of white drivers encountered in circumstances similar to those of the actual minority drivers. In statistical terms, this is known as an average treatment effect on the treated (ATT), where minority status may be regarded as the "treatment" in the parlance of causal inference. By contrast, the overall average treatment effect (ATE) that Aggarwal *et al.* appear to target also incorporates an additional comparison that does not involve realized harm to minority drivers—whether actual white drivers were treated better than hypothetical minority drivers encountered in circumstances similar to those of actual white drivers. Both ATT and ATE can be estimated using the same data, though the less-relevant ATE can often be estimated with more statistical precision because it relies more heavily on data for white drivers, who are far more numerous. Notably, the ATT and ATE can diverge substantially because, as Aggarwal *et al.* show, minority drivers tend to be younger and drive vehicles with different characteristics, they likely drive in different geographic areas, and these differing circumstances are among the most important factors influencing police citation decisions.

Moving from academic research to practical applications has other implications for statistical analysis. For one, it requires researchers to recognize that the quantity examined in any particular study—in the case of Aggarwal *et al.*, relating to officers discriminatorily citing minority versus white drivers in comparable locations and times—tells only part of the broader story about how discrimination may manifest in law enforcement. For example, it may be just as important to study bias in how officers are assigned to work those locations and times in the first place—estimands that may reveal patterns of, for example, overdeployment in minority neighborhoods causing disparate impact.

To be clear, academic research on discriminatory policing need not always tailor its statistical analyses for policy settings. And Aggarwal *et al.* have provided a template for using recent technological advances to overcome some of the most challenging obstacles impeding policing research. But to maximize impact on pressing social problems, this study—like decades of research before it (*8*)—could benefit from greater clarity about the specific estimand being targeted and a precise explanation of why it is the most relevant quantity of interest. This clarification is imperative if academics wish to aggregate knowledge across independent studies being conducted on the roughly 18,000 police agencies in the US. To conduct meaningful meta-analyses and comparisons, scholars will need to agree on, and precisely specify, the statistical quantities that they seek to measure. ■

REFERENCES AND NOTES

1. *Yick Wo* v. *Hopkins*, 118 U.S. 356 (1886).
2. P. Aggarwal *et al*., *Science* **387**, 1397 (2025).
3. F. Goncalves, S. Mello, *Am. Econ. Rev*. **111**, 1406 (2021).
4. D. Knox, W. Lowe, J. Mummolo, *Am. Polit. Sci. Rev*. **114**, 619 (2020).
5. Subsetting to instances where a driver is objectively speeding, as proposed by Aggarwal *et al*., is conceptually distinct from subsetting to instances where police subjectively chose to detain a civilian for speeding, as in much prior work that has analyzed police stop data. The former is based on an objective fact that is unlikely to induce selection bias because there is arguably no unobserved factor that jointly determines a civilian's decision to speed and an officer's decision to issue a speeding ticket—especially because analysts can condition on car make, time, location, and other circumstances. In the latter case, an officer's decisions to stop a civilian and issue a citation could be jointly influenced by many factors that remain unobserved.
6. US DOJ Civil Rights Division, "Investigation of the City of Phoenix and the Phoenix Police Department" (2024); https://www.justice.gov/crt/media/1355866/dl?inline.
7. US DOJ Civil Rights Division, "Section VI: Proving Discrimination – Intentional Discrimination," in "Title VI Legal Manual" (2015); https://www.justice.gov/crt/media/1384931/dl?inline.
8. I. Lundberg, R. Johnson, B. M. Stewart, *Am. Sociol. Rev*. **86**, 532 (2021).

ACKNOWLEDGMENTS

The authors thank J. Grimmer for helpful feedback on this article. This work was supported by Arnold Ventures grant 22-06762 (D.K. and J.M.) D.K. and J.M. have worked as paid consultants for the US DOJ Civil Rights Division, the American Civil Liberties Union, and the NAACP Legal Defense Fund.

10.1126/science.adw3618