PNAS Nexus

PNAS Nexus, 2024, **3**, 1–10

https://doi.org/10.1093/pnasnexus/pgae026

Advance access publication 20 February 2024

Research Report

The deadliest local police departments kill 6.91 times more frequently than the least deadly departments, net of risk, in the United States

Josh Leung-Gagné[a,*]

[a]Center on Poverty & Inequality, Stanford University, Stanford, CA 94305, USA

*To whom correspondence should be addressed: Email: jgagne@stanford.edu

Edited By: Taylor Jaworski

Abstract

I use data linking counts of homicides by police to police department (PD) and jurisdiction characteristics to estimate benchmarked (i.e. risk-adjusted) police homicide rates in 2008–2017 among the 711 local PDs serving 50,000 or more residents, a sample with demographics resembling all mid-to-large Census places. The benchmarked rate estimates capture PD deadliness by comparing PDs to peers whose officers face similar risks while adjusting for access to trauma care centers to account for differential mortality from deadly force. Compared to existing estimates, differences in benchmarked estimates are more plausibly attributable to policing differences, speaking to whether the force currently used is necessary to maintain safety and public order. I find that the deadliest PDs kill at 6.91 times the benchmarked rate of the least deadly PDs. If the PDs with above-average deadliness instead killed at average rates for a PD facing similar risks, police homicides would decrease by 34.44%. Reducing deadliness to the lowest observed levels would decrease them by 70.04%. These estimates also indicate the percentage of excess police homicides—those unnecessary for maintaining safety—if the baseline agency is assumed to be optimally deadly. Moreover, PD deadliness has a strong, robust association with White/Black segregation and Western regions. Additionally, Black, Hispanic, foreign-born, lower income, and less educated people are disproportionately exposed to deadlier PDs due to the jurisdictions they reside in. Police violence is an important public health concern that is distributed unevenly across US places, contributing to social disparities that disproportionately harm already marginalized communities.

Keywords: social demography, criminology, spatial inequality, public health

Significance Statement

Many contend that US police officers use appropriate force, attributing community differences in the rates of police killing people to officers accurately tailoring their use of force to differently risky communities. Suppose we accept this premise for the average US police department (PD): the average PD kills at the frequency necessary to ensure officer and public safety. Given this assumption, I find that roughly one in three police homicides could have been avoided without endangering officers and the public. This is because there is a wide range of police homicide rates among PDs that face similar risks. The differences are also associated with jurisdiction demographics with already marginalized groups tending to reside in jurisdictions with more deadly policing.

Introduction

US police kill far more frequently than police in peer countries, accounting for 8% of all homicides of US adult males (1–3). The frequency is particularly high for Black men, who are over 3 times more likely than White men to be killed by police, such that police homicides are a leading cause of death for young Black men (1, 4). In addition to the deceased, police homicides harm the health of decedents' families and communities, as well as the health of the involved officers, and are increasingly recognized as a critical public health concern (5–9). It is critically important to avoid excess police homicides, which might be thought of as those that do not prevent bystanders and officers from incurring severe injuries.

This study grapples with the possibility of excess police homicides in the United States by focusing on police department (PD) tendencies. Excess police homicides are more likely when officers have low thresholds for using deadly force, overestimate threats, are overconfident in their perceptions of threat, or are underconfident in their perceptions of safety. They may also be more likely if the approach to policing fosters more hostile interactions (10). These factors are all likely to be shaped by the place and PD in which an officer works. Police officers across the United States

Competing Interest: The author declares no competing interest.
Received: May 11, 2023. **Accepted:** January 10, 2024

are subject to different state laws and PD policies, strategies, trainings, norms, and oversight arrangements (11–21). Even PDs with similar organizational conditions may nonetheless have different behavioral tendencies due to contextual differences in political economy and demographics that elicit more or less officer suspicion, fear, respect, or care (15, 22–27).

Recent advancements in the collection of police homicides data have enabled researchers to demonstrate that police kill far more frequently in some parts of the country compared to others (1, 28, 29). Though it is common to frame these findings as differences in police violence, it remains unclear whether the differences are due to excess police homicides or variation in policing contexts (30).

A common explanation for community variation in police homicides is that contextual differences, particularly in the risks faced by officers, necessitate different amounts of force. Early articulations of the Community Violence Hypothesis in the criminology literature attribute community variation in police use of force to officers accurately tailoring their approach to the communities they police in order to be safe and effective (31–33). This perspective also appears in public perceptions: in 2016, 75% of Whites agreed that police use "the right amount of force in each situation" (34). Shedding light on police violence requires a focus on discretionary uses of force.

Another possible explanation for the observed differences is variation in mortality risk from a given use of force, whether due to officer aim or medical care (30). One important factor is trauma care access, which has been identified as an important variable determining the likelihood of surviving gunshot wounds (35–37). Adequate benchmarks that account for policing context are essential to understanding variation in police violence.

The recently released National Officer-Involved Homicide Database (NOIHD) has made such benchmarking feasible by linking high-quality police homicide data from the Fatal Encounters database (FE) to several datasets with information on PDs, injuries to officers, crimes and arrests, and contextual characteristics like trauma care access (3, 38, 39). This facilitates comparisons of PDs that face similar risks to better identify pockets of policing violence with respect to deadly force.

This paper uses the NOIHD to describe variation in PD deadliness—defined as the extent to which a PD kills more often than peers that face similar risks—in 2008–2017 among the 711 local PDs with jurisdictions of 50,000 or more residents. This sample has a close demographic resemblance to the full set of US Census places with populations of 50,000 or more, but is not generalizable to the whole United States (SI Appendix, Table S1).

By estimating differences more plausibly attributable to differences in policing and less plausibly attributable to variation in nondiscretionary police homicides, the analysis speaks to whether current levels of violence used by police are in excess of what is necessary to maintain officer and public safety. I operationalize PD deadliness as the PD's counterfactual police homicide rate per capita under average conditions of (i) trauma care access (per capita and per square mile), (ii) violence against officers (differentiating deadly and nonfatal injurious assaults), and (iii) officer encounters that imply risks of violence against officers and bystanders (noninjurious assaults against officers and arrest counts for nonviolent crime and three types of violent crime). The estimates focus on nonvehicular intentional use of force (IUF) homicides as coded by the FE team; in order of frequency, this includes deaths from officers shooting, tasering, asphyxiating, bludgeoning, stabbing, using chemical agents, or burning people.

I use multilevel negative binomial models and report Bayesian PD estimates and maximum likelihood (ML) between-PD variance estimates that account for uncertainty-induced variance inflation in estimates of rare events. The model assumes that policing violence occurs strictly under the measured conditions, which may not hold due to two continuing challenges in the national data infrastructure (i) the reliance on police homicide data, rather than all uses of deadly force and (ii) weaknesses in the national policing data I use to capture policing context. However, a battery of sensitivity analyses indicate that my findings are robust to the addition of omitted confounders, alternative functional forms, using ambient rather than residential population, focusing on police homicides by gunshots, expanding or narrowing the time period, using listwise deletion to purge potential "data dumping" PDs rather than using multiple imputation, and alternative data cleaning approaches. Additional detail on the data, sample, methods, assumptions, limitations, and sensitivity analyses can be found in Materials and methods below and SI Appendix.

My results show that PDs facing similar risks kill at substantially different rates, with the deadliest PDs killing 6.91 times more frequently than the least deadly PDs. If PD deadliness was reduced such that no PDs had benchmarked (i.e. risk-adjusted) homicide rates higher than the average, the number of police homicides in the sample would decrease by 34.44%. If the average mid-to-large local PD kills as frequently or more frequently than necessary to maintain officer and public safety, this indicates that at least one in three police homicides by mid-to-large local PDs are excess police homicides.

I also consider whether variation in PD deadliness contributes to demographic disparities in police homicides. Nationwide, unadjusted rates of police homicide vary substantially by the race/ethnicity of the decedent and the racial/ethnic and economic composition of the neighborhood (1, 40). Compared to Whites, Black decedents are more likely than Whites to be unarmed and Black people have tended to die at lower rates when shot by police, suggesting that unadjusted rates could understate racial disproportionality in police use of deadly force (41, 42). Though a rich body of research details the intersection of geographic, race/ethnicity, and class inequalities in experiences with the US criminal justice system (1, 22, 28, 29, 42–46), it remains unclear whether disparate treatment by PDs is exacerbated by differential exposure to more violent PDs. I find that Black, Hispanic, foreign-born, low income, and less educated people are segregated into jurisdictions with more deadly PDs.

Additionally, I describe the variation in PD deadliness in a correlational analysis using a battery of variables describing state and agency policies, staff composition, agency location, jurisdiction composition, and jurisdiction inequality in accordance with the literature's identification of policy variation, organizational features, and local social context as determinants of police behavior (10–17, 19, 22–27, 44, 47). I identify White/Black segregation and Western location as strong correlates of PD deadliness robust to the inclusion of other features.

The findings are inconsistent with widely held beliefs that variation in policing violence is an adequate response given differential police exposure to commonly cited risk factors like crime and violence against officers. Instead, they suggest that there are opportunities for intervening in PD practices, policing writ large, and social policy to reduce police homicides without reducing officer safety and effectiveness.

Results

Figure 1 displays the range of PD-specific deadliness estimates, measured as the counterfactual rate (per 1 million jurisdiction residents per year) of police homicides under sample average conditions of risk to officers and access to trauma care. PD deadliness estimates are benchmarked rates based on PDs' estimated deviations from model predictions; the ratio of PD deadliness to the sample average is equivalent to the ratio of the PD police homicide rate to rates at which PDs facing similar risks tend to kill. The deadliness estimates in the highlighted PD jurisdictions are significantly different ($P < 0.05$) from the sample average of 3.39 (3.31, 3.47; 95% uncertainly level) per million residents per year.

The estimated 95% plausible interval for PD deadliness ranges from 1.29 in the least deadly PDs (estimated at the 2.5 percentile) to 8.91 in the deadliest PDs (estimated at the 97.5 percentile), indicating that the deadliest PDs kill at 6.91 times the benchmarked rate of the least deadly PDs (see SI Appendix, Table S6 for full regression estimates). This is substantially greater than what would occur randomly; across 100 simulations of random police homicide rates across PD-year observations, all estimated ratios of most to least deadly PDs were <1.02.

Previous spatial comparisons of police homicides used no adjustment, potentially overstating differences in discretionary police violence (1, 28–30). Following that approach, one would estimate that the PDs that kill most often kill at 9.39 times the rate of those that kill least often, as compared to my risk-adjusted estimate of 6.91. The rank-ordered distribution of unadjusted estimates is included in Fig. 1 for reference. The benchmarked comparisons do find less inequality across PDs, primarily due to lower estimates among the deadliest PDs, but the overall variation in policing deadliness remains stark. Moreover, benchmarking the comparison modestly alters how PDs rank in comparison to one another (SI Appendix, Fig. S3).

Figure 2 lists the 20 most and least deadly PDs in the sample with their PD deadliness estimates. Benchmarked rates in the three deadliest PDs are roughly 3 times the sample average; I estimate that under average conditions, St. Louis Metro PD (MO), Kansas City PD (MO), and Columbus Division of Police (OH) would, respectively, kill 10.64 (9.36, 12.00), 10.04 (8.89, 11.34), and 9.74 (8.66, 10.95) people per million annually. In comparison, I estimate that the least deadly PDs—Raleigh PD (NC), Metro Nashville PD (TN), and Worcester PD (MA)—would, respectively, kill 1.37 (1.06, 1.77), 1.56 (1.31, 1.87), and 1.65 (1.23, 2.21) people per million per year under average conditions. That the least deadly PD estimates are all above the lower bound of the 95% plausible interval reflects the greater degree of uncertainty—and therefore greater Bayesian adjustment—in PD-specific estimates of infrequent events compared to estimates of variance between PDs (see Materials and methods below and SI Appendix for more discussion of uncertainty adjustments).

Figure 3 demonstrates that the deadliest PDs killed far more frequently in 2008–2017 than the least deadly PDs despite comparable estimated levels of risk. The figure compares PD residuals to fitted values. Substantively, this means the horizontal axis—the fitted values—is the rate at which PDs policing similarly risky jurisdictions tended to kill in the study period, while the vertical axis—the residual ratio—is a ratio comparing this to my estimates of what the PD's actual police homicide rate was during that time. Note that the vertical axis is an alternative presentation of the PD deadliness estimates above. A PD can be compared to peer PDs with comparably risky jurisdictions by scanning vertically at a given fitted value. There is substantial variation in police homicide rates among PDs facing similar risks. Moreover, the fitted values of the 20 most and least deadly PDs are unexceptional, indicating that their exceptional estimated deadliness is not driven by the model extrapolating to account for outliers. These findings replicate when comparing PD deadliness using unimputed rates of officer injuries from assaults rather than model-fitted values (SI Appendix, Fig. S10).

Given that PD deadliness varies substantially among comparably risky jurisdictions, it stands to reason that some PDs kill more frequently than necessary to maintain officer and public safety. Perhaps, then, PD deadliness could be reduced to the average level observed among peers. How many lives would this save? If every PD with above-average deadliness—those above the 1:1 ratio line in Fig. 3—instead had average deadliness—falling along the 1:1 line—34.44% of the people killed by police in this sample would have survived their encounter with police.

This can also be thought of as a possible estimate of the proportion of police homicides that are excess police homicides. Suppose we restate the claim that US police appropriately tailor use of force to account for risk as "the average US PD kills at the frequency necessary to maintain officer and public safety." Under this conservative assumption, excess police homicides are 34.44% of all police homicides. Less conservative assumptions yield far greater estimates. For example, if Raleigh PD is a good benchmark for how deadly PDs need to be, 70.04% of police homicides were excess polices homicides (see SI Appendix, Table S9 for estimates using alternative benchmarks).

Figure 4 considers how the differences in PD deadliness relate to demographic disparities in the likelihood of being killed by police. Municipalities and counties are important sites of residential segregation by race and class such that these groups potentially experience policing differently partly as a result of living in different jurisdictions (48, 49). Using the PD deadliness estimates and 2015–2019 American Community Survey (ACS) data, I estimate the average PD deadliness in the jurisdictions in which demographic groups reside. Note that these are differences between jurisdictions that do not account for within-jurisdiction disparities.

I find that Black people reside in jurisdictions with 15.31% (12.27, 17.39) more deadly policing than those of White people after accounting for jurisdiction differences in risk to police and trauma care access. Hispanic people are 7.59% (4.92, 9.26) more exposed to deadly policing than White people, and foreign-born people are 3.39% (0.78, 6.01) more exposed than those born in the United States. Households making $10,000 or less are in jurisdictions with 11.37% (9.10, 13.03) more deadly policing than households making $200,000 or more, while people without high school degrees are 6.61% (5.11, 7.58) more exposed than college graduates.

One implication of Fig. 4 is that reducing PD deadliness to Raleigh PD levels would not only reduce police homicides in the sample by 70.04% but also reduce the racial disparity between Black and White people, among other demographic disparities. In SI Appendix, I simulate a counterfactual where the sample PDs all had Raleigh-level deadliness in 2017, while within-PD racial disproportionality remained unchanged, and all out-of-sample PDs are unaffected. The national Black:White ratio of per capita police homicide rates reduced from 2.78 to 1.62 and the Black–White gap in rates (per million) reduced from 3.47 to 0.96 (see SI Appendix, Table S10 for all estimates).

Figure 5 describes the pattern of PD deadliness by correlating it with a battery of agency and jurisdiction characteristics, which I have categorized to provide insights into the types of characteristics that characterize variation in policing violence. An extensive literature attributes variation in police behavior to patterns of officer discretion, policy variation, and organizational features, as well as policing's role in the social structure (10–17, 19, 22–27, 47, 50). I group 57 variables drawn from the NOIHD, the

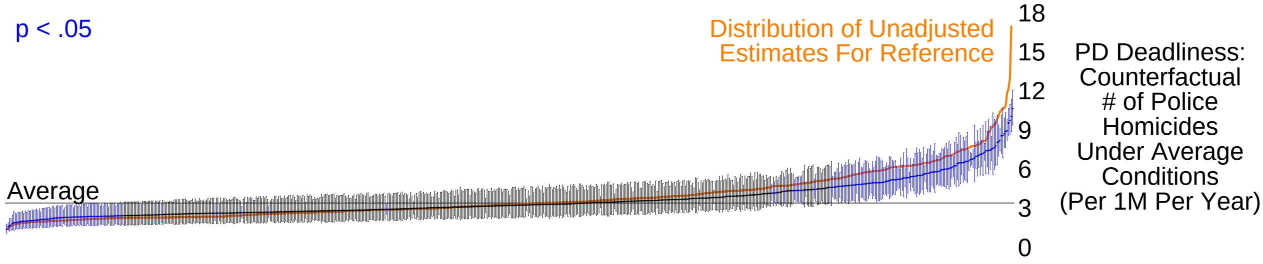

Fig. 1. Rank-ordered distribution of PD deadliness, measured as the counterfactual rate (per 1 million jurisdiction residents per year) of police homicides under sample average conditions of risk to officers and access to trauma care. The sample includes 711 agencies with jurisdictions of 50,000 or more in 2008–2017. Markers indicate Bayesian PD estimates. Vertical dashes indicate 95% uncertainty intervals. The rank-ordered distribution of unadjusted Bayesian PD estimates is provided as a reference.

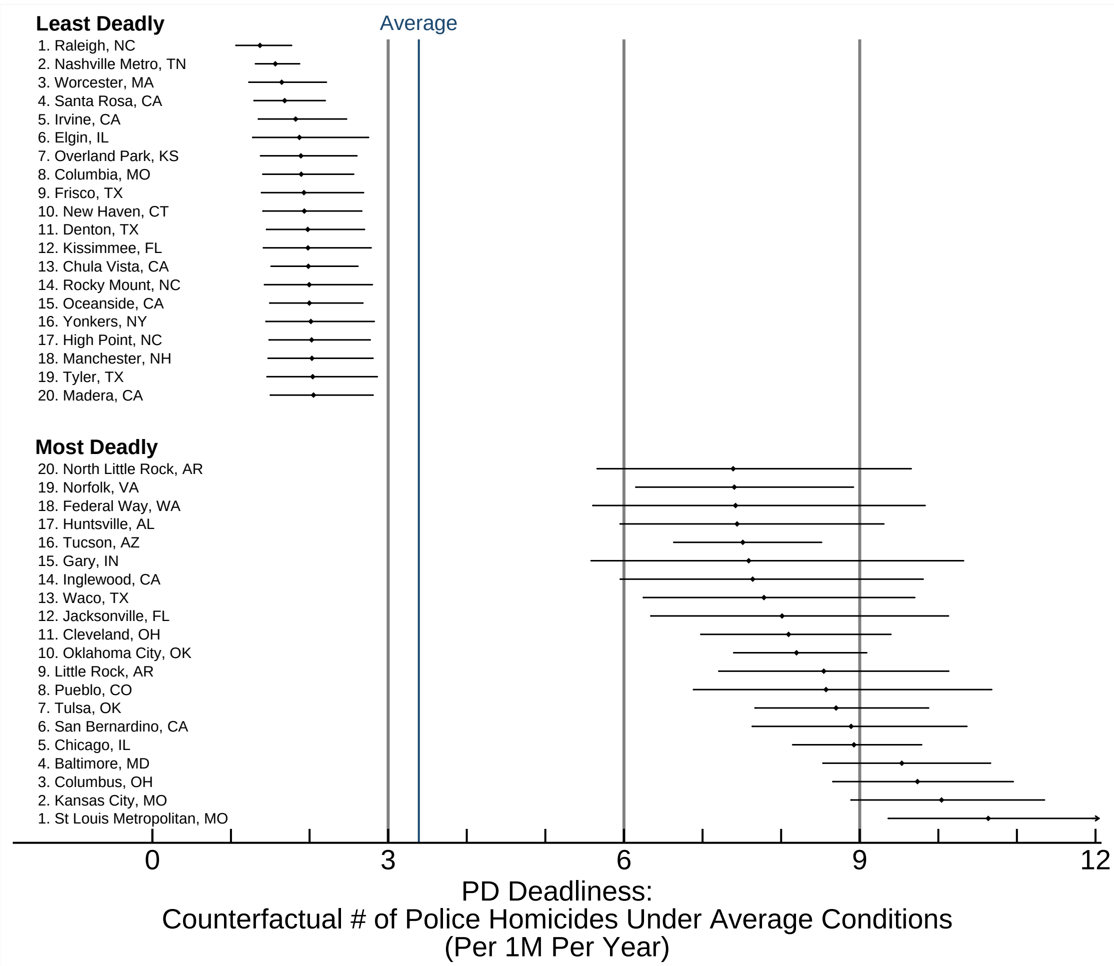

Fig. 2. The PDs with the lowest (least deadly) and highest (most deadly) counterfactual rates (per 1 million jurisdiction residents per year) of police homicides under sample average conditions of risk to officers and access to trauma care. The sample includes 711 agencies with jurisdictions of 50,000 or more in 2008–2017. Markers indicate Bayesian PD estimates. Horizontal dashes indicate 95% uncertainty intervals.

2015–2019 ACS, and qualitative descriptions of state policies (20, 21) into 11 categories: staff composition; training, restrictions, oversight, and gun regulation policy; location; jurisdiction composition by race/ethnicity, socioeconomic status (SES), and other; and jurisdiction inequality by race/ethnicity, SES, and SES by race.

Group-specific and bivariate analyses assess which types of factors characterize more deadly PDs, while the full model analysis using all the variables asses which types of factors explain the variance independent of the others. All models use precision-weighted Ordinary Least Squares (OLS) to account for differential error in deadliness estimates across PDs. Variables are coded a priori to predict greater PD deadliness based on past research findings (see SI Appendix for details). I consider an association strong when a 1 SD increase in a continuous variable (or 0 to 1 change in a binary variable) is associated with a statistically significant ($P < 0.05$) 0.20 SD change in PD deadliness.

Altogether, the correlates explain 35% of the observed variation in PD deadliness (in the full model, $R^2 = 0.35$). I find that PD

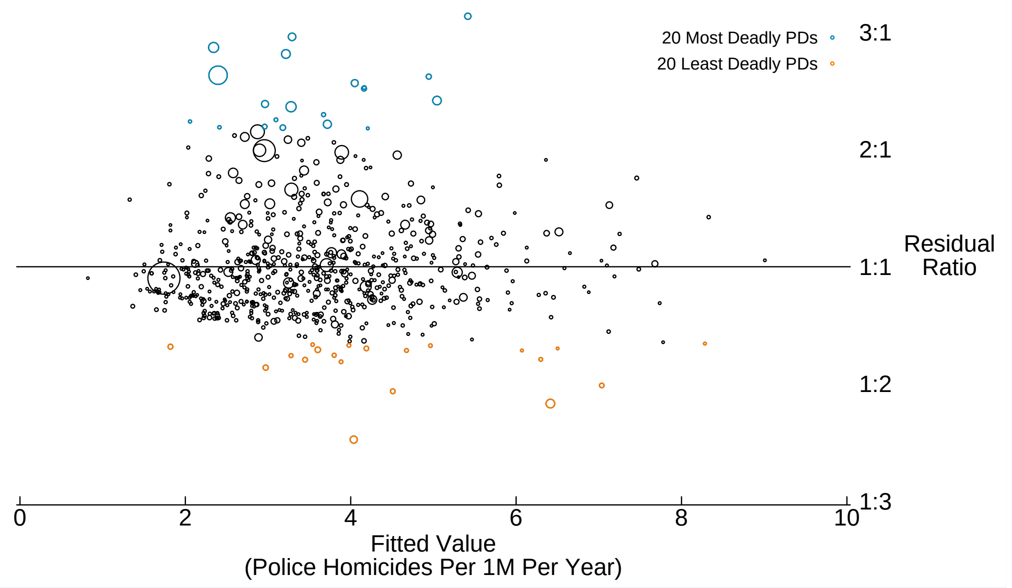

Fig. 3. Bayesian residual ratios vs. fitted values of PD police homicide rates (per 1 million jurisdiction residents per year) as a function of risk to officers and trauma care access in 2008–2017. Markers are weighted by jurisdiction population and highlighted to indicate the 20 most and least deadly PDs. Bayesian residual ratios are ratios of Bayesian PD police homicide rate estimates to PD fitted values and are the basis for the PD deadliness estimates in Figs. 1 and 2.

deadliness is strongly associated with location in Western Census divisions and more racially segregated jurisdictions, and these strong associations persist independent of the associations between PD deadliness and the other characteristics. Western location exhibits a particularly strong independent association with PD deadliness which holds across all Census divisions West of the Mississippi. Among racial segregation measures, Black/White neighborhood segregation stands apart as a strong, robust correlate, while other types of racial segregation are only strong correlates prior to accounting for other characteristics.

Other strong associations that do not persist independent of other characteristics include a tendency for greater PD deadliness when the jurisdiction has lower median income, more single-mother-headed households, more inequality and segregation by income, and White-favoring White–Black income disparities. The associations with race/ethnic composition are in the predicted direction, but not as strong and not robust to including other characteristics. Perhaps counterintuitively, PD deadliness tends to be somewhat lower in PDs with more White and male officers, but these tendencies are not independent of the associations of PD deadliness. Among the 20 policy variables, only policies indicating weaker oversight are strongly associated with PD deadliness and explain the variation therein, particularly in full models, but the direction of association is inconsistent across models and variables. Variable-specific estimates are presented in SI Appendix, Table S11.

Discussion

Policing is far more deadly in some PDs than in others. The deadliest PDs kill 6.91 times more frequently than the least deadly PDs after accounting for variation in risk to officers and trauma care access. This estimate is lower than those found in previous studies that do not adjust for contextual risk factors, but remains large and stark. The deadly and least deadly PDs largely patrol areas of moderate risk, indicating that their estimated rates are not driven by extrapolation in the tails of the risk distribution. Moreover, differences in PD deadliness are associated with jurisdiction demographics such that Black, Hispanic, foreign-born, lower income, and less educated people tend to reside in jurisdictions policed by more deadly PDs. I identify White/Black segregation and location in the Western United States as strong, robust predictors of more deadly policing. Police violence is an important public health concern that is distributed unevenly across US places, contributing to social disparities in already marginalized communities.

Many believe that officers accurately gauge the amount of force they use and that the different amounts of force used in different communities reflects tailoring policing to more or less risky places in order to be safe and effective (31–34). This view is compatible with previous research documenting community variation in unadjusted police homicide rates because the variation could be attributed to factors, like risk to officers, that promote different rates of nondiscretionary police homicide rates. However, my findings indicate that PD differences are not readily explained by such factors. Suppose the average US PD accurately gauges their use of deadly force; even under this conservative assumption, I find that roughly one in three police homicides are excess police homicides. Using the least deadly PD in the sample as a less conservative benchmark increases the estimated rate of excess police homicides to roughly two in three. Moreover, PD deadliness is negatively correlated with clearance rates (both on violent crimes and for crime overall), key indicators of PD effectiveness, which suggests that less deadly PDs are not sacrificing effectiveness or officer safety (SI Appendix, Table S12).

This is consistent with an expansive literature attributing variation in police behavior to patterns of officer discretion, policy variation, and organizational features, as well as policing's role in the social structure (10–17, 19, 22–27). The substantial variation suggests that there are opportunities for intervening in PD practices, policing writ large, and social policy to reduce police homicides without reducing officer safety and effectiveness. Reducing policing violence such that no PDs are more deadly than average would have reduced police homicides among mid-to-large local PDs by 34.44% in 2008–2017. Reducing PD deadliness to the lowest

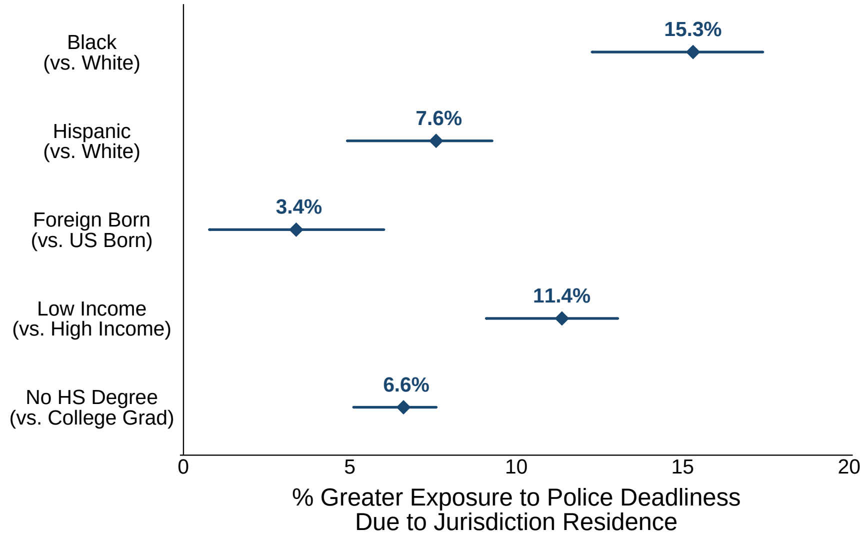

Fig. 4. Demographic disparities in jurisdiction of residence PD deadliness estimates. PD deadliness is estimated as the counterfactual rate (per 1 million jurisdiction residents per year) of police homicides under sample average conditions of risk to officers and access to trauma care. Disparities compare group averages of Bayesian PD deadliness estimates within individuals' jurisdictions of residence where percentages are relative to the comparison group average. The sample includes 711 agencies with jurisdictions of 50,000 or more in 2008–2017. Horizontal dashes indicate 95% uncertainty intervals.

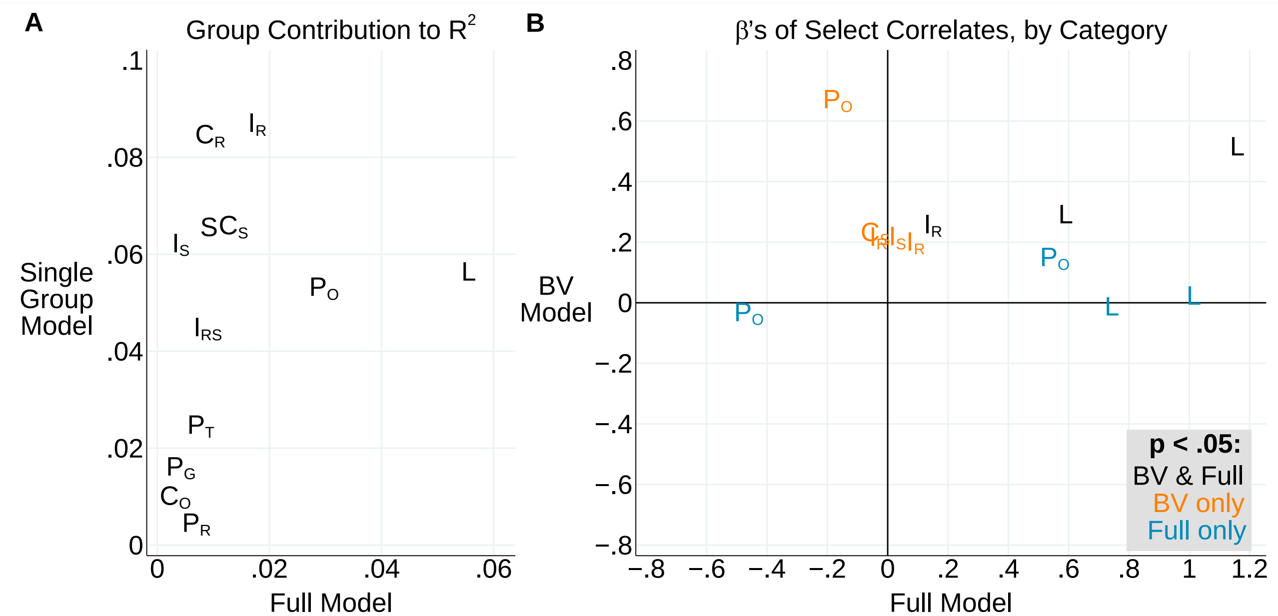

Fig. 5. Correlates of standardized PD deadliness estimates, organized by category. Panel A plots each correlate category's contribution to R^2 in (i) a single-group model including only the correlates in the category, along the vertical axis and (ii) the full model with all correlates, measured as the difference in R^2 between the full model and a model excluding the category, along the horizontal axis. Panel B plots the linear regression coefficients of correlates, by category, in (i) a bivariate (BV) model, along the vertical axis and (ii) the full model, along the horizontal axis, displaying only correlates with significant correlations in at least one model and estimated coefficients of magnitude0.2 or greater in at least model. All correlates are either binary or standardized continuous variables. The sample includes 711 agencies with jurisdictions of 50,000 or more in 2008–2017. Correlate categories: S, staff (variables coded in the direction of more White, male); P_T, policies regarding training (less training, qualifications); P_R, policies restricting conduct (weaker); P_O, policies regarding oversight (limiting); P_G, policies regulating guns (weaker); L, location (Western, lower population density); C_R, jurisdiction composition of race/ethnicity (less White, foreign-born); C_S, jurisdiction composition of SES (lower); C_O, other jurisdiction composition factors (more children, residential instability); I_R, inequality by race/ethnicity (segregated); I_S, inequality by SES (segregated, higher Gini coefficient); I_{RS}, inequality by race/ethnicity and SES (White-favoring income disparities).

observed levels would have reduced homicides by 70.04%. It would also dramatically reduce Black–White racial disparities, even if within-PD racial disproportionality remained the same.

Future research can leverage estimates of PD deadliness to identify PDs for case study and investigate how policies, practices, and contexts have shaped the landscape of policing deadliness

(SI Appendix, Dataset S1). My findings point to White/Black segregation as an important contextual feature and identify greater deadliness among PDs West of the Mississippi, indicating that these factors—oft-noted in the literature (1, 22, 28, 42, 43)—are associated with benchmarked estimates in addition to unadjusted rates. My analyses were less successful at identifying effective policy levers compared to past research (12, 14, 17, 19); future scholarship on determinates of PD deadliness would benefit from better causal purchase and integrating more policy details and data on organizational culture and training approaches into the NOIHD.

Rather than relying on unadjusted local rates of police homicides, PDs can be better identified to more optimally target analyses and interventions using benchmarked estimates like those presented here. The adjusted estimates more accurately and more credibly capture the potential public welfare benefits of reducing each agency's use of deadly force. However, one should still mind the caveat that the PD deadliness estimates are representations of a complex microlevel process that available data cannot speak to. Some special or changing circumstances in individual PDs may not be accounted for. Rather than providing definitive evidence of the performance of specific PDs, the estimates are most illuminating as evidence of substantial across-PD variation in policing deadliness and are only first indications of the potential presence of exacerbating or intervening factors in specific PDs.

My findings are also consistent with the long-documented sociological finding that "places matter"; the political and administrative boundaries of communities (e.g. municipalities, counties, jurisdictions, etc.) configure political economy and institutions, critically impacting the lives of their residents (48, 49, 51–54). This study contributes to this literature by documenting place-based inequality in deadliness, with implications for community public health (5, 6). Patterns of PD deadliness are also associated with patterns of place-based residential segregation such that jurisdictional differences expose disadvantaged racial and economic groups to more deadly PDs than advantaged groups. Future work should consider whether this arises from heightened police violence in response to the racial and economic composition of jurisdictions (22–24, 26).

This study is not without limitations. As discussed in Materials and methods below, I avoid geographic misalignment of the PD- and jurisdiction-specific measures by restricting the sample such that PDs and jurisdictions map 1-to-1. However, this removes 10.63% of the police homicides in agencies that otherwise meet the sample requirements, primarily county sheriff's departments and metropolitan area PDs. For example, Los Angeles County Sheriff's Office killed 294 people during the study period, but had to be excluded because it is only the primary PD for parts of LA County, creating issues for interpreting contextual characteristics. The resulting sample captures 89.4% of the IUF police homicides among local PDs of mid-to-large jurisdictions and 91.0% of the residents in all mid-to-large Census places, where it closely resembles on demographic measures (SI Appendix, Table S1). This sample does not generalize well to the nation, however. Additionally, there was insufficient power with 711 PDs to consider temporal trends within PDs; it is likely that the estimates smooth over important PD deadliness trends and variation therein. Future analyses would benefit from finding less restrictive solutions.

This study improves our understanding of police violence by benchmarking estimates of police homicide rates, an approach recently made possible due to data improvements. However, there are limitations in the data available for benchmarking, particularly because there is no widespread tracking of uses of deadly force in the United States such that benchmarking must account for estimating jurisdiction risk of dying from uses of deadly force. Though FE and the NOIHD are marked improvements in tracking and investigating police homicides, the widespread adoption of administrative encounters-level use of force data remains sorely needed (41). Additionally, it remains difficult to measure jurisdiction risk to officers without more precise data on police encounters, whether or not they result in arrest. It is also unclear how to measure jurisdiction risk to bystanders to the extent that it differs from risk to officers. Some steps toward improving researcher capacity to capture these features and reducing reliance on proxies include nationally tracking use of force incidents, measuring how aid is rendered after use of force incidents, and improving the granularity of measures of violence against officers.

While the benchmarks used here rely on imperfect contextual data, a sensitivity analysis described in SI Appendix finds substantial robustness to omitting confounders, indicating that the PD estimates are useful first indications of the potential presence of exacerbating or intervening factors in specific PDs (see SI Appendix, Table S8). It also suggests that data improvements are unlikely to meaningfully alter the key finding that there is substantial across-PD variation in policing deadliness not attributable to commonly cited risk factors like crime and violence against officers. This is inconsistent with widely held beliefs that variation in police use of force is an adequate response given differential police exposure to such risk factors.

Finally, my statistical approach relies on within-US comparisons of PDs, potentially masking room for improvement among less deadly PDs. International comparisons also suggest that the average US PD is a conservative benchmark and much less deadly policing may be feasible (2).

Materials and methods

The measures used in this analysis are drawn from the NOIHD. The NOIHD uses the Law Enforcement Agency Identifiers Crosswalk to link FE to several official and unofficial datasets at the PD, Census place, and state level. By linking FE to several datasets with contextual information, the NOIHD has made it possible to make benchmarked (i.e. risk-adjusted) comparisons of rates of police homicides and speak more directly to differences in police violence.

The annual PD counts of IUF officer-involved homicides originate in FE, a journalist-led project documenting officer-involved homicides using systematic searches of public records, online news reports, and social media to track all deaths in which an officer on duty is present. FE has better coverage and is less prone to bias than existing official and unofficial sources (3, 39). Though FE has better coverage and is less prone to bias than existing official and unofficial sources, it is subject to undercounting when officer-involved deaths do not appear in news reports and are not documented in searchable public records (3, 39). To account for this concern, the analysis focuses on 2008–2017 and limits the sample to PDs that serve a jurisdiction of at least 50,000 residents. The 2008 start date was selected based on prior research found that FE coverage meaningfully improves prior to this year, as online news media grew, and stopped improving thereafter (4). Excluding small jurisdictions further reduces potential undercounting bias due to limited online news reporting.

Whereas prior studies often describe variation across geographic units that are large (e.g. states) and/or nonpolitical (e.g. census divisions) (1, 28, 29), this study's estimates focus on PDs because they play a primary role in determining local law enforcement policies, strategies, and norms in their respective jurisdictions (11–15, 17, 18). The sample includes the 711 US PDs that serve a jurisdiction of at least 50,000 residents as the local, parent, and primary PD for the entirety of the jurisdiction. This restriction ensures that the PD and its jurisdiction match one-to-one. Police homicides involving multiple agencies are split evenly among them, creating fractional police homicides.

The sample restrictions ensure high data quality at a cost, removing 56.41% of the IUF police homicides identified by FE in 2008–2018 across 34,719 local, subsidiary, state, and federal agencies, ultimately constraining the generalizability of the findings to mid-to-large local PDs and mid-to-large Census places. Among them, I exclude 10.63% of all police homicides due to the restriction by PD type, retaining a large sample for an analysis of this kind. The full analytic sample includes 4921.44 police homicides over 10 years, averaging 0.69 per PD per year with a range of 0 per year in 114 PDs to 22.75 per year by Los Angeles PD. The sample captures 91.0% of the residents in all mid-to-large Census places, resulting in a sample of residents that closely resembles those of mid-to-large Census places on demographic measures. More descriptive statistics are provided in SI Appendix, Tables S1–S3.

The control variables selected from NOIHD seek to capture excess police homicides by proxying for the likelihood that a given use of deadly force incident (i) resulted in death and (ii) was not discretionary due to the violence of the citizen. All measures are averaged over the period within each PD. I proxy for risk of dying from uses of deadly force using measures of trauma care density per capita and per square mile originating in restricted proprietary data. Though trauma care access has been identified as an important variable determining the likelihood of surviving gunshot wounds (35–37), it is not the only factor determining survival of uses of deadly force. I consider omitted confounding below.

I proxy for nondiscretionary uses of deadly force using PD-reported measures constructed from items originating in the FBI's Uniform Crime Reports (UCR) and the Bureau of Justice's Law Enforcement Officers Killed and Assaulted (LEOKA). Two of these measures track violence against officers: injuries and mortalities from assaults while on duty, per officer. These are coarse measures of risk; to supplement them, I screened several indicators of how frequently officers have encounters with higher risks of violence against officers and bystanders. The included measures are noninjurious assaults while on duty, per officer, and four measures of arrests per officer, each focused on one of the four UCR crime categories: parts 1 and 2 violent crimes against persons, part 1 violent property crimes, and nonviolent crimes.

I include indirect measures of risk to officers and bystanders to better control for pathways of nondiscretionary uses of deadly force. However, these measures might also control for pathways of excess police homicides, particularly if they capture officer misperceptions of risk. This could lead to underestimating PD differences. I limit this threat by excluding indicators, like violent crimes per capita, that are not positively associated with violence against officers but are positively associated with police exposure to social deprivation as measured by the frequency of officers encountering deaths from overdoses, suicides, medical emergencies, drownings, and falls. See SI Appendix for a more detailed discussion of the screening process.

I produce Bayesian PD deadliness estimates with robust SEs using the following two-level negative binomial model in which the outcome, P_{tj}, is the per capita rate of IUF homicides by police in year t within PD j:

$$\ln\left(\frac{P_{tj}}{1-P_{tj}}\right) = \beta_{00} + \overline{\overline{\mathbf{X}}}_{\mathbf{j}} + \overline{\overline{\mathbf{X}}}_{\mathbf{j}}^{2} + e_{tj} + \upsilon_{0j}$$

$$e_{tj} \sim N(0, N_{tj} \times P_{tj}(1-P_{tj})), \qquad \upsilon_{0j} \sim N(0, \tau) \tag{1}$$

where $\overline{\overline{\mathbf{X}}}_{\mathbf{j}}$ is the vector of grand-mean-centered control variables and N_{tj} is the jurisdiction residential population, measured annually. Robust SEs reduce model sensitivity to violations of HLM assumptions and using them did not substantially alter the errors in the preferred model (55). I use residential population in keeping with the literature, but using ambient population instead has little effect on PD deadliness estimates (SI Appendix, Fig. S9). I use quadratic polynomials of the covariates to account for potentially changing marginal responses to increases in risk factors because greater frequency can shift norms or induce a sense of crisis. Using linear or cubic polynomials instead has little effect on PD deadliness estimates (SI Appendix, Fig. S8).

The estimated policing deadliness in PD j is the Empirical Bayes estimate of $\beta_{00} + \upsilon_{0j}$ in which υ_{0j} is shrunken toward β_{00} in inverse proportion to its reliability, minimizing the mean squared error of the estimates (55). The model excludes PD-year intercepts due to insufficient power, necessitating estimates of PD deadliness that are pooled over years, so I used all 10 available years since FE coverage stabilized (4). I estimate between-PD variance in policing deadliness, τ, using the HLM ML estimator, which is equivalent to the variance of the PD estimates after shrinking υ_{0j} toward β_{00} in inverse proportion to the square root of its reliability (55). The Bayesian PD estimates and ML variance estimates are preferable to OLS estimates which would overstate PD differences and the between-PD variance due to uncertainty in the estimates (see SI Appendix for more details). This is critical when analyzing rates of police homicides, the rarity of which creates substantial uncertainty in PD-specific estimates.

Another source of uncertainty is missing data in the covariates. For a given PD and variable, missingness occurs when the PD has missing data in all years or the PD partially reported data, either annual data with some years missing or monthly data with some months missing. Inconsistent reporting and persistent noncompliance are a challenge for working with UCR and LEOKA because of the added noise, underestimated errors, and unclear generalizability that this can induce. I overcome this challenge using multiple imputation with chained equations (MICE) to impute ($M = 5$) PD means when PDs are missing data in all years using chained equations (56). MICE models include all potential covariates that were considered prior to screening covariates. To reflect uncertainty in PD means when PDs have partial data for a given covariate, I estimate the error variance in their observed means and vary their means accordingly by adding noise across the imputed datasets. I pool model parameters using scalar inference, producing estimates that generalize to mid-to-large local PDs and inflating SEs to appropriately reflect the noise due to missingness (57). The Bayesian estimation procedure, by shrinking noisy estimates toward the mean, also ensures that noise in individual PD estimates resulting from missing data does not distort estimates of the national distribution.

Two additional challenges with the PD survey data are that PDs may differently interpret survey items, particularly related to officer injuries and assaults, and some PDs engage in data-distorting "data dumps" (e.g. skipping some reporting periods, then reporting numbers aggregated across the current period and the past, skipped periods) (58). The former would distort PD estimates in that PDs using lower thresholds will have underestimated PD deadliness due to the exaggeration of risks, which could bias the variance down (if more deadly PDs use lower thresholds) or up (if less deadly PDs use lower thresholds) and may increase omitted confounding due to noise in the control variables. However, model estimates are robust to excluding any one of the officer injury and assault measures (SI Appendix, Table S8), providing some evidence that any such distortions are unlikely driving the findings.

Regarding "data dumps," they would induce downward bias in estimates for PDs that engage in the practice due to inflating the apparent risk and this could also distort the model, affecting the estimates of PDs that diligently report. I attempt to purge apparent dumps and false zeroes through judicious data censoring, slightly increasing missingness rates (see SI Appendix for more details). However, the substantive findings are unaltered whether I am more permissive (avoiding censoring) or when I use the far more aggressive approach of listwise deletion to ensure that the analysis is purged of any PDs that do data-distorting "data dumps" (SI Appendix, Figs. S5 and S7).

Other important threats to estimating policing deadliness are misalignment of PD- and jurisdiction-specific measures, inaccurate counts of IUF homicides by police, and omitted confounders. I avoid geographically misaligned measures by restricting the sample to PDs are the local, parent, and primary PD for the entirety of the jurisdiction such that PDs and jurisdictions map 1-to-1 (see SI Appendix for more details). I limit the sample to 2008–2017 and to jurisdictions of at least 50,000 residents in 2017 to reduce potential bias from FE undercounts when there is limited online news reporting. Estimated disparities and rankings are robust to focusing on gunshot instead of all IUF homicides and expanding or shrinking the 2008–2017 sample window (SI Appendix, Table S7, and Figs. S5 and S6).

Omitted confounding is a particularly pressing concern. The existing national data is limited in scope and imprecise; UCR for example collapses police encounters that are rich with information, whether or not they lead to arrest, into coarse arrest and officer injury and assault counts aggregated over all encounters. Even more concerning is that police homicide data by its nature samples on an outcome, mortality that is determined by more than the use of deadly force (30, 41). Trauma care access is one key factor (35, 37), but there are others that are unobserved in the existing data that has a national scope, such as officer shooting accuracy. These issues plague the field, which would benefit tremendously from the widespread adoption of administrative data on police encounters and use of force, with California providing a promising model.

However, a sensitivity analysis described in SI Appendix indicates that including an omitted confounder similar to any one of the control variables used in my analyses would reduce PD disparities by <10% and would produce rankings 97% or more correlated with my estimated rankings (SI Appendix, Table S8). This suggests that risk is adequately captured; data improvements equivalent to adding a measure tracking officers being injured by assaults (or any of the other model covariates) would not meaningfully alter the findings. It remains unclear whether including implied risk covariates (i.e. arrest rate data) controls for officer perceptions in addition to true risk, potentially introducing conservative bias. This sensitivity check also cannot simulate the omission of confounders orthogonal to those that are observed, which is particularly a concern for omitted predictors of mortality after uses of deadly force like officer shooting accuracy. Though this analysis improves upon past research with respect to accounting for the selection on the outcome problem, it may not be possible to confidently overcome this issue in nationwide analyses without the mass adoption of administrative data that captures nonfatal shootings.

Acknowledgments

This article benefited from the feedback and advice of Brian K. Finch, Audrey Beck, David Klinger, Joel Garner, D. Brian Burghart, Kyla Thomas, and Richard Johnson.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

The author declares no funding.

Author Contributions

J.L.-G. performed the research and wrote the paper.

Data Availability

The public version of the NOIHD is available online (38); it includes all variables used in the analyses above with the exception of the trauma center prevalence variables which rely on proprietary data and are only available in the restricted-use version of the NOIHD. The SI Appendix includes detailed explanations of all data preparation and modeling decisions. The code used in the analysis and the PD deadliness estimates are also available as supplementarymaterial.

References

1 Edwards F, Esposito MH, Lee H. 2018. Risk of police-involved death by race/ethnicity and place, United States, 2012–2018. *Am J Public Health*. 108:1241–1248.
2 Lartey J. 2015, May 25. By the numbers: US police kill more in days than other countries do in years. The Guardian.
3 Finch BK, *et al.* 2019. Using crowd-sourced data to explore police-related-deaths in the United States (2000–2017): the case of fatal encounters. *Open Health Data*. 6:1.
4 Edwards F, Lee H, Esposito M. 2019. Risk of being killed by police use of force in the United States by age, race–ethnicity, and sex. *Proc Natl Acad Sci U S A*. 116:16793–16798.
5 Alang S, McAlpine D, McCreedy E, Hardeman R. 2017. Police brutality and Black health: setting the agenda for public health scholars. *Am J Public Health*. 107:662–665.
6 Cooper HLF, Fullilove M. 2016. Editorial: excessive police violence as a public health issue. *J Urban Health*. 93(Suppl 1):1–7.
7 Komarovskaya I, *et al.* 2011. The impact of killing and injuring others on mental health symptoms among police officers. *J Psychiatr Res*. 45:1332–1336.
8 Loo R. 1986. Post-shooting stress reactions among police officers. *J Human Stress*. 12:27–31.
9 Solomon RM, Horn JH. 1986. Post-shooting traumatic reactions: a pilot study. In: Reese JT. Goldstein HA, editors. *Psychological services for law enforcement officers*. Washington (DC): US Goverment Printing Office. p. 383–393.
10 Stoughton S. 2014. Law enforcement's warrior problem. *Harv Law Rev Forum*. 128:225–234.

11 Eitle D, D'Alessio SJ, Stolzenberg L. 2014. The effect of organizational and environmental factors on police misconduct. *Police Q*. 17:103–126.

12 Jennings JT, Rubado ME. 2017. Preventing the use of deadly force: the relationship between police agency policies and rates of officer-involved gun deaths. *Public Adm Rev*. 77:217–226.

13 Sierra-Arévalo M. 2021. American policing and the danger imperative. *Law Soc Rev*. 55:70–103.

14 Silver JR, Roche SP, Bilach TJ, Bontrager Ryon S. 2017. Traditional police culture, use of force, and procedural justice: investigating individual, organizational, and contextual factors. *Justice Q*. 34: 1272–1309.

15 Renauer BC. 2007. Understanding variety in urban community policing: an institutional theory approach. In: Maguire ER. Duffee DE, editors. *Criminal justice theory: explaining the nature and behavior of criminal justice*. 1st ed. New York: Routledge. p. 121–150.

16 National Conference of State Legislatures, Legislative Responses for Policing—State Bill Tracking Database (accessed 2022 May 26). https://www.ncsl.org/civil-and-criminal-justice/legislative-responses-for-policing-state-bill-tracking-database.

17 Sinyangwe S. 2016. *Examining the role of use of force policies in ending police violence*. Social Science Research Network.

18 Sherman LW. 1980. Causes of police behavior: the current state of quantitative research. *J Res Crime Delinquency*. 17:69–100.

19 Sloan JJ, Paoline EA. 2021. They Need More Training! A national level analysis of police academy basic training priorities. *Police Q*. 24:486–518.

20 Flanders C, Welling J. 2015. Police use of deadly force: state statutes 30 years after garner. *St Louis Univ Public Law Rev*. 35:109–156.

21 Rushin S. 2017. Police union contracts. *Duke Law J*. 66:1191–1266.

22 Holmes MD, Painter MA, Smith BW. 2019. Race, place, and police-caused homicide in U. S. municipalities. *Justice Q*. 36:751–786.

23 Turk AT. 1969. *Criminality and legal order*. Chicago: Rand McNally.

24 Blalock HM. 1967. *Toward a theory of minority-group relations*. New York: Wiley.

25 Black D. 2010. *The behavior of law: special edition*. Bingley (UK): Emerald Group Publishing.

26 Voigt R, *et al*. 2017. Language from police body camera footage shows racial disparities in officer respect. *Proc Natl Acad Sci U S A*. 114:6521–6526.

27 Simon SJ. 2021. Training for war: academy socialization and warrior policing. *Soc Probl*. 70:spab057.

28 Schwartz GL, Jahn JL. 2020. Mapping fatal police violence across U.S. metropolitan areas: overall rates and racial/ethnic inequities, 2013–2017. *PLoS One*. 15:e0229686.

29 Campaign Zero, Mapping Police Violence (accessed on 2022 Apr 1). https://mappingpoliceviolence.org/.

30 Nix J. 2020. On the challenges associated with the study of police use of deadly force in the United States: a response to Schwartz & Jahn. *PLoS One*. 15:e0236158.

31 Kania RRE, Mackey WC. 1977. Police violence as a function of community characteristics. *Criminology*. 15:27–48.

32 Robin GD. 1963. Justifiable homicide by police officers. *J Crim Law Criminol Police Sci*. 54:225–231.

33 Fyfe JJ. 1980. Geographic correlates of police shooting: a microanalysis. *J Res Crime Delinquency*. 17:101–113.

34 Morin R, Stepler R. 2016. The racial confidence gap in police performance. *Pew Res Center's Soc Demogr Trends Proj*.

35 Circo GM. 2019. Distance to trauma centres among gunshot wound victims: identifying trauma 'deserts' and 'oases' in Detroit. *Inj Prev*. 25:i39–i43.

36 Crandall M, *et al*. 2013. Trauma deserts: distance from a trauma center, transport times, and mortality from gunshot wounds in Chicago. *Am J Public Health*. 103:1103–1109.

37 MacKenzie EJ, *et al*. 2006. A national evaluation of the effect of trauma-center care on mortality. *N Engl J Med*. 354:366–378.

38 Finch BK, *et al*. 2022. National Officer-Involved Homicide Database (NOIHD) [Data set]. Version 1. NICHD R01HD093382; ICPSR 38118. (Accessed 2021 December 23) https://fatalencounters.org/.

39 Finch BK, *et al*. 2021. Assessing data completeness, quality, and representativeness of justifiable homicides in the FBI's supplementary homicide reports: a research note. *J Quant Criminol*. 38: 267–293.

40 Feldman J. 2020. *Police killings in the US: inequalities by race/ethnicity and socioeconomic position*. People's Policy Project.

41 Nix J, Shjarback JA. 2021. Factors associated with police shooting mortality: a focus on race and a plea for more comprehensive data. *PLoS One*. 16:e0259024.

42 Siegel M, Poulson M, Sangar R, Jay J. 2021. The interaction of race and place: predictors of fatal police shootings of Black victims at the incident, census tract, city, and state levels, 2013–2018. *Race Soc Probl*. 13:245–265.

43 Johnson O, St. Vil C, Gilbert KL, Goodman M, Johnson CA. 2019. How neighborhoods matter in fatal interactions between police and men of color. *Soc Sci Med*. 220:226–235.

44 Bass S. 2001. Policing space, policing race: social control imperatives and police discretionary decisions. *Soc Justice*. 28:156–176.

45 Stuart F. 2011. Race, space, and the regulation of surplus labor: policing African Americans in Los Angeles's skid row. *Souls*. 13:197–212.

46 Goldfischer E. 2020. From encampments to hotspots: the changing policing of homelessness in New York City. *Hous Stud*. 35: 1550–1567.

47 Montiel Olea JL, O'Flaherty B, Sethi R. 2021. Empirical Bayes counterfactuals in Poisson regression with an application to police use of deadly force. https://doi.org/10.2139/ssrn.3857213 (accessed 2023 Nov 1).

48 Lichter DT, Parisi D, Taquino MC. 2015. Toward a new macro-segregation? Decomposing segregation within and between metropolitan cities and suburbs. *Am Sociol Rev*. 80:843–873.

49 Massey DS, Rothwell J, Domina T. 2009. The changing bases of segregation in the United States. *Ann Am Acad Pol Soc Sci*. 626:74–90.

50 Ba BA, Knox D, Mummolo J, Rivera R. 2021. The role of officer race and gender in police-civilian interactions in Chicago. *Science*. 371: 696–702.

51 Logan JR, Molotch HL. 1987. *Urban fortunes: the political economy of place*. University of California Press.

52 Bischoff K. 2008. School district fragmentation and racial residential segregation: how do boundaries matter? *Urban Aff Rev*. 44:182–217.

53 Reardon SF, Kalogrides D, Shores K. 2019. The geography of racial/ethnic test score gaps. *Am J Sociol*. 124:1164–1221.

54 Sampson RJ, Raudenbush SW, Earls F. 1997. Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science*. 277:918–924.

55 Raudenbush SW, Bryk AS. 2002. *Hierarchical linear models: applications and data analysis methods*. 2nd ed. Thousand Oaks (CA): SAGE Publications.

56 Azur MJ, Stuart EA, Frangakis C, Leaf PJ. 2011. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*. 20:40–49.

57 van Buuren S. 2018. *Flexible imputation of missing data*. 2nd ed. CRC Press.

58 Sierra-Arévalo M, Nix J. 2020. Gun victimization in the line of duty. *Criminol Public Policy*. 19:1041–1066.