Which Police Departments Want Reform? Barriers to Evidence-Based Policymaking*

Samantha Goerger[1], Jonathan Mummolo[2], and Sean J. Westwood[3]

[1]Independent Scholar
[2]Princeton University
[3]Dartmouth College

May 16, 2022

Abstract

Political elites increasingly express interest in evidence-based policymaking, but transparent research collaborations necessary to generate relevant evidence pose political risks, including the discovery of sub-par performance and misconduct. If aversion to collaboration is non-random, collaborations may produce evidence that fails to generalize. We assess selection into research collaborations in the critical policy arena of policing by sending requests to discuss research partnerships to roughly 3,000 law enforcement agencies in 48 states. A host of agency and jurisdiction attributes fail to predict affirmative responses to generic requests, alleviating concerns over generalizability. However, across two experiments, mentions of agency performance in our correspondence depressed affirmative responses—even among top-performing agencies—by roughly eight percentage points. Many agencies that initially indicate interest in transparent, evidence-based policymaking recoil once performance evaluations are made salient. We discuss several possible mechanisms for these dynamics, which can inhibit valuable policy experimentation in many communities.

*Samantha Goerger is an independent scholar, sgoerger@alumni.princeton.edu. Jonathan Mummolo is an Assistant Professor of Politics and Public Affairs, jmummolo@princeton.edu. Sean Westwood is an Associate Professor of Government, Sean.J.Westwood@Dartmouth.edu. This study was approved by Institutional Review Boards at Princeton University and Dartmouth College. A pre-registration plan for the national experiment appears in the Online Appendix. We thank Tori Gorton, Alexandra Koskosidis, Destiny Eisenhour, Krystal Delnoce, Grace Masback and Madeleine Marr for research assistance. Authors listed in alphabetical order. The authors declare no conflicts of interest. This research was funded by Princeton University and Dartmouth College. Samantha Goerger and Sean Westwood declare no conflicts of interest. Jonathan Mummolo works as a paid consultant for the American Civil Liberties Union, the NAACP Legal Defense Fund, and the U.S. Dept. of Justice Civil Rights Division.

The increasing availability of high-resolution data on human behavior and the development of field experimental methods in social science have made research collaborations with practitioners the gold standard in policy research (Cartwright and Hardie, 2012). These partnerships—which span substantive arenas including poverty reduction (Alatas et al., 2012), political advertising (Gerber et al., 2011), and health care (Litvack and Bodart, 1993)—offer numerous advantages, simultaneously leveraging access to otherwise restricted data, real-world settings, and rigorous experimental designs (Gerber and Green, 2012).

But like any approach to research, partnerships with practitioners have drawbacks. Chief among them is the fact that the very organizations being studied decide whether research can proceed, and there are strong reasons to suspect this decision is associated with outcomes scholars wish to understand, like agency performance. Put differently, while many political elites have recently instituted calls for "evidence-based policymaking" (Orszag and Nussle, 2017), such declarations may be cheap talk. The political risks associated with allowing outside experts to scrutinize organizational practices—e.g. the discovery of sup-par performance, or even misconduct—are substantial, especially for poorly functioning organizations (Carpenter, 2014; Levine, 2020; Moffitt, 2010). And if poorly-performing agencies are differentially likely to decline research partnerships, the body of evidence produced by one-off research collaborations could fail to generalize to organizations at large (Allcott, 2017).

In this study, we assess the determinants and generalizability of research collaborations in the important policy domain of policing. A long history of allegations of racial bias (Alexander, 2010; Lerman and Weaver, 2014; Gelman, Fagan and Kiss, 2007), a recent string of high-profile police-involved killings (Edwards, Lee and Esposito, 2019), and growing concern over the use of excessive force and militarized policing (Gunderson et al., 2019; Knox, Lowe and Mummolo, 2020) have spurred numerous collaborations between academics and law enforcement agencies to detect inequity in police procedures (e.g. Goff et al., 2016) and test the efficacy of proposed reforms (e.g. Yokum, Ravishankar and Coppock, 2019). But the highly politicized nature of policing suggests many agencies will be reluctant to partner with researchers, and that the ones who do are unrepresentative of the roughly 18,000 law enforcement agencies in the United States.

To evaluate the severity and nature of selection, we conducted two field experiments in which we sent offers to roughly 3,000 local police and sheriff's departments to discuss a potential research collaboration with scholars at two East Coast universities, and analyzed variation in responses. This design allowed us to assess both the correlates and causes of willingness to collaborate. Merging data on responses with records of jurisdiction demographics, local partisanship, department personnel, and agency performance, we first show that agencies open to discussing research collaborations are largely similar to those that declined our invitations. This finding bolsters the validity of the collaborative research approach, and suggests findings emanating from one-off research partnerships are plausibly contributing to a generalizable body of knowledge. However, across two experiments, including a pre-registered nationwide replication, a randomized mention of agency performance in our communications depressed affirmative responses by roughly eight percentage points. These negative effects hold even for top-performing agencies.

Agencies that initially show openness to research partnerships look broadly similar to those who will not consider them, but the willingness to partner with academics for policy research is not as widespread as it appears. Once discussions move from the general to the specific, and raise the prospect of performance evaluations critical to the field testing of any new policy, many agencies recoil. There may be several reasons for this reaction. Law enforcement may be averse to systematic evaluation—a sign that many agencies indicating willingness to discuss collaborations may be engaging in cheap talk. On the other hand, raising the specter of performance too early, before trust is established, may stymie an otherwise fruitful collaboration Glaser and Charbonneau (2018). Regardless of the precise mechanism at work, this dynamic reveals a barrier to research collaborations that can preclude valuable policy experimentation in many communities

Experimental Design

We began with a study in New Jersey involving 462 agencies, paired with newly released detailed data on the use of force (nj.com, 2019). During April and May of 2019, we contacted police chiefs offering to collaborate on research that "aims to make both citizens and officers safer by reducing the frequency of violence during police-citizen interactions" (see Online Appendix section B2 for full text). We relied on a custom Python script to prepare and send our messages from a dedicated institutional email. These messages contained no deception; offers to discuss collaborations were sincere. We offered to work pro-bono and cover all research costs, and added "We are not asking for a firm commitment now" but are simply asking whether the recipient is "interested in discussing a potential collaboration further." Respondents could answer (via links in email and a URL provided in print letters) yes, no, or "I am not sure, but I would like more information." Our primary outcome is a binary indicator of answering "yes," with all other responses and non-responses coded as negative responses. If we received no response after three email attempts—spaced eight days apart—we sent a posted letter one week after the final email.

Agencies in the N.J. study were randomly assigned to one of four conditions.[1] All agencies received the information above, which served as the full text for those in the control condition. Three treatment conditions included language aimed at testing how common features of research collaboration requests affect agency responses. One treatment cell included a promise of confidentiality in any publication that resulted from a research partnership, which is a common practice in such settings and which we hypothesized would increase affirmative responses. A second "ranking" condition included mention of agency performance: the agency's rank on uses of force per officer among contacted agencies. A third condition combined both the confidentiality and ranking treatments (with the order of the two treatments randomized within the text of the email across recipients).

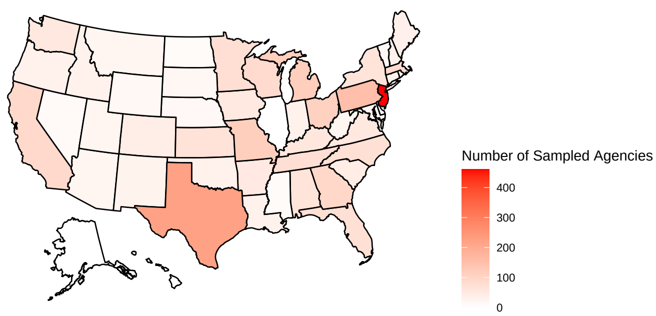
[1]See Online Appendix Section D4 for balance tests.

Following the N.J. study we deployed a second pre-registered experiment in 47 additional states during September and October of 2019, in which we attempted to contact approximately 2,500 local police and sheriff's departments, a sample size we chose based on a power analysis in order to detect a possible interaction between the performance treatment and agency rank. We randomly drew our sample from a population of roughly 7,700 agencies that consistently report crime data to the FBI and for whom we could ascertain reliable contact information (these criteria excluded of Alaska and Illinois; see Appendix section A1 for sampling details). Roughly 60% of the U.S. population reside in these agencies' jurisdictions, according to FBI data. While we would ideally wish to sample from the entire U.S., the population of agencies that remain after applying these filtering criteria are those with whom a productive research collaboration might plausibly occur. It would be difficult to form collaborations with agencies that do not regularly report basic crime data, or publicize reliable contact information. Our sample is therefore a relevant one for applied researchers.

The design of this experiment was highly similar to the N.J. study with some exceptions. Two changes were aimed at maximizing statistical power. First, we retained only the ranking treatment and control conditions. Second, we employed a matched pair design (Gerber and Green, 2012), in which agencies in the same state serving roughly the same population size were paired, and one agency was randomly assigned to treatment. Specifically, treated agencies were told how they ranked among the roughly 2,500 agencies sampled on the share of violent crimes "cleared" between 2013-2017 (crimes where an arrest was made and charge was filed)—a salient statistic for police agencies, and one on which journalists often focus (e.g. Madhani, 2018). The use of two different performance metrics across these experiments helps to ensure the robustness of any observed treatment effects.

We hypothesized that mentions of agency performance would filter out "cheap talk" and depress affirmative responses on average, since making performance evaluations salient could cause agencies to consider the political risks associated with research partnerships. However, we anticipated that this negative effect would attenuate with agency rank, since agencies informed they were performing well relative to peers may be less likely to recoil at the spectre of performance evaluations. Following both experiments, all contacted agencies were sent a debrief message informing them of the purpose of the experiment, and reinforcing that our messages contained no deception.[2]

Figure 1: **Coverage of Field Experiments**. Number of contacted agencies in each U.S. state. Over 400 agencies were contacted in the N.J. study.



Combined, contacted agencies serve jurisdictions that are home to close to 80 million people according to FBI data (see Figure 1), approximately one quarter of the U.S. population. These include large metropolitan police forces, mid-sized agencies and small rural departments employing just a handful of officers.[3]

No Evidence Collaborating Agencies Are Unrepresentative

To test whether willingness to collaborate systematically varies with agency attributes, we merged agency-level data on: crime, fatal officer-involved shootings between 2015 and 2018, personnel, and jurisdiction demographics.[4] In total, 319 agencies indicated willingness to discuss a potential research collaboration—approximately 11% of our combined sample

[2]Communication between contacted agencies could contaminate our results. This is less likely in the national experiment given its geographic spread. To the extent cross-talk occurred, it likely homogenized responses across treatment conditions and attenuated effects.

[3]Following our IRB protocol, we map agencies at the state-level to maintain anonymity.

[4]See Online Appendix Section C3 for details on data sources.

of 2,942 agencies across the two experiments—238 agencies responded negatively to our message, and 2,387 agencies did not reply at all.

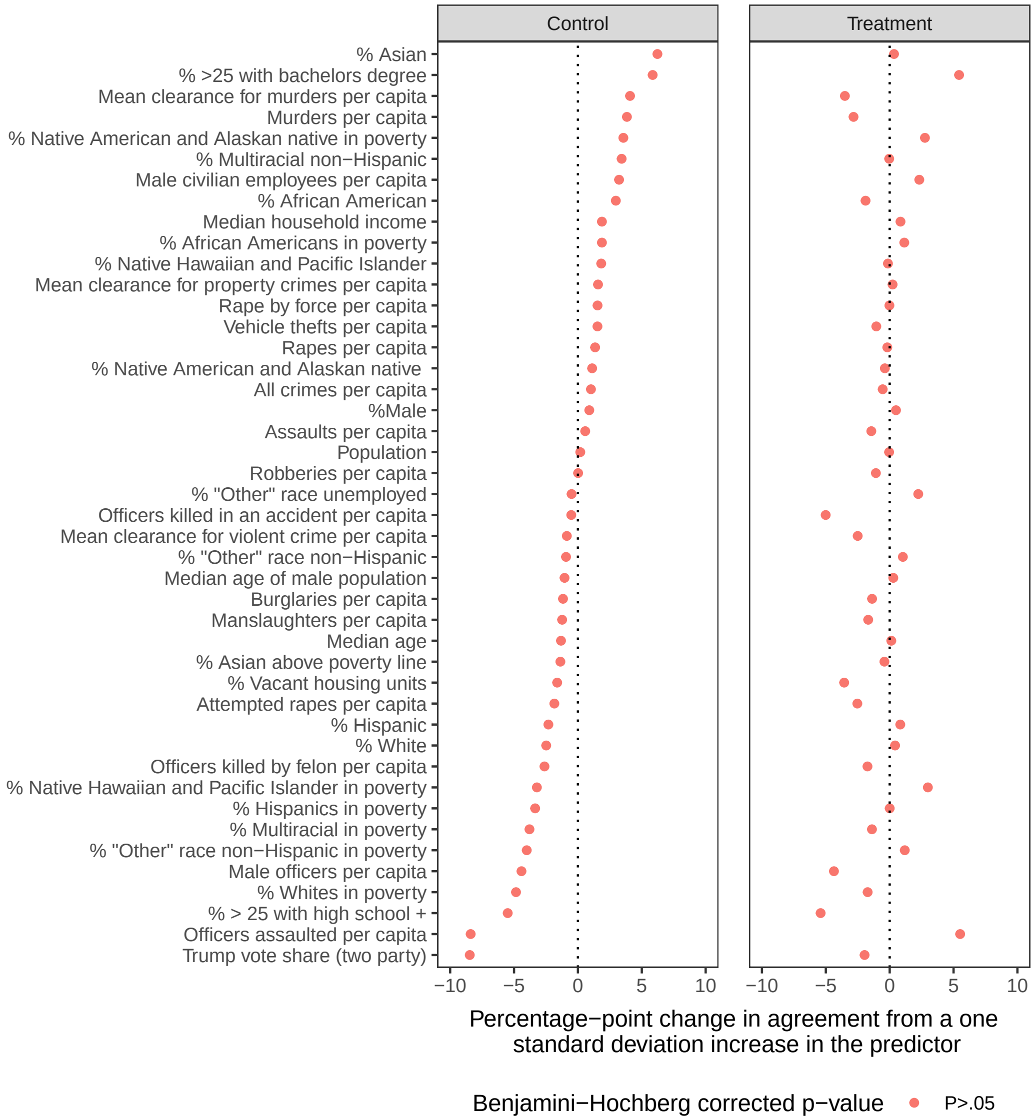
We estimated separate bivariate linear regressions predicting affirmative response as a function of each covariate, weighted by jurisdiction population. We correct resulting p-values on all regression coefficients using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995), though this adjustment makes little difference to our overall conclusions. To avoid conflating the predictive value of a regressor with the effect of our randomized interventions, we confine this analysis to the roughly 1,400 observations assigned to control in the left facet, of which 201 agencies responded affirmatively. For comparison we conduct the same analysis for agencies in the treatment group in the national study. [5]

Figure 2 displays the predicted change in the probability of an affirmative response given a one-standard-deviation increase in each predictor. Across the 44 test results in this figure, no covariate was significantly associated with responses in either the control or treatment groups. Overall, our analysis indicates selection bias—at least at the initial point of contact from researchers—poses a minimal risk in this setting.

Some may question whether we are missing meaningful associations in this analysis due to a lack of statistical power. But while additional data may allow us to detect correlations, other features of the results belie the existence of meaningful relationships. For one, several features related to agency performance generate opposing signs, e.g. assaults on officers and a host of crime measures including murders and rapes per capita. Second, only one result, officers assaulted per capita, is statistically significant when we opt not to correct for multiple testing. Third, we find no significant associations even if we include data from all experimental conditions to maximize sample size (see Figure E2 in Appendix). The largest estimated coefficient relates to Trump's share of the two-party vote in an agency's county, suggesting agencies in conservative areas may be less likely to collaborate. However,

[5] We exclude N.J. State Police in this analysis due to difficulty in accurately pairing with U.S. Census data.

Figure 2: **Agency and jurisdiction attributes do not predict affirmative responses.**

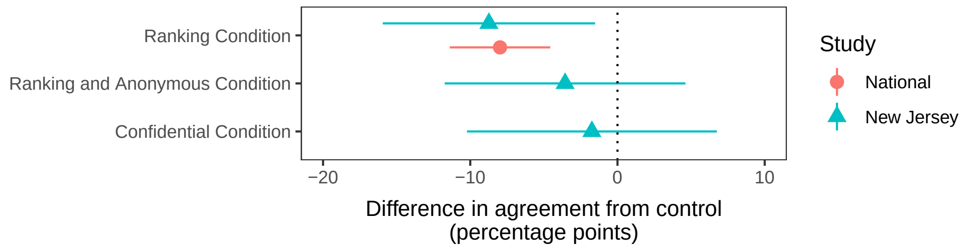


the overall pattern of results does not indicate selection related to agency performance, suggesting agencies arguably in most need of reform are not systematically declining to collaborate.

Mentions of Performance Evaluations Inhibit Collaborations

We now turn to assessing the impact of our experimental interventions. Figure 3 displays the average effect of each treatment relative to the control condition estimated via

Figure 3: **Mentioning agency performance lowers affirmative responses**



linear regression. Because we cannot guarantee all messages were reviewed, these represent Intention-to-Treat effects (ITTs), understating the effect of universally received similar messages.[6] In the national experiment, our models include indicators for all matched pairs, with standard errors clustered by matched pair.

Turning first to the N.J. experiment, we find randomized offers to keep the identity of collaborating agencies confidential, including one version where a performance cue was also supplied, had no detectable effect on response rates ($\beta = -0.02, se = 0.04, p = 0.69$ and $\beta = -0.04, se = 0.04, p > 0.40$ respectively). This was surprising, as such confidentiality offers are often made to convey a sense of security and thereby increase the likelihood of collaborations. However, because such offers still rely on academic collaborators to keep their word and effectively safeguard agency identities, this promise may ring hollow, and additional assurances may be required before agencies will consider collaborations. However, recipients told their statewide rank on mean uses of force per officer ("Ranking Condition") were roughly 9 percentage points ($se = 0.004, p = 0.02$) less likely to respond affirmatively than agencies assigned to control where about 15% of agencies agreed. Strikingly, this effect was precisely replicated in the nationwide experiment: agencies told their rank on violent crime clearance rates were about 8 percentage points ($se = 0.01, p < 0.01$) less likely to say they would discuss a potential collaboration.

Contrary to our expectations, additional tests interacting treatment assignment with agency rank show that these negative effects persist even among top performing agencies

[6]Emails and post letters sent to four agencies were returned to us due to invalid addresses. We drop these and their corresponding matched pair from all analyses since the ITT interpretation is invalid.

(see Online Appendix Figure G1). This result is consistent with police agencies having a strong aversion to outside evaluation, and suggests a powerful impediment to the formation of research partnerships. While many agencies indicate openness to collaboration, a large share recoil once the topic of agency performance is inevitably broached. This may be because agencies that performed well on a given metric in the past have no guarantee of positive results in the future, especially once outside scrutiny is allowed.

Of course, other mechanisms are also possible, and we have a limited ability to adjudicate between them with the data at hand. Prior work on collaborations with law enforcement agencies emphasizes the importance of researchers spending time with agency officials to educate themselves about the particulars of the institution and personnel (Glaser and Charbonneau, 2018; Levine, 2020). Conveying a measure of performance in an initial outreach message may have inadvertently sent the signal that the researchers had "jumped the gun," and evaluated the agency before doing proper due diligence. The specific metrics we conveyed in our interventions may also have depressed responses. Agencies may have felt other measures more accurately conveyed their level of performance, and thus inferred the researchers contacting them were ill-equipped to assist them.[7]

However, regardless of the mechanism at play, the end result is the same: mentions of performance evaluations in outreach messages inhibit collaborations. Since evaluating the efficacy of reforms on agency performance is a central goal of these collaborations, we interpret these effects as evidence of a substantial barrier to policy experimentation, but one that may be overcome with a more measured approach to solicitation that works to establish trust with police administrators before discussing performance metrics.

Discussion and Conclusion

While they offer numerous advantages over other methods of inquiry, research collaborations with outside experts also pose political risks that may preclude partnerships in ways

[7]See Appendix I9 for additional discussion of causal mechanisms.

that threaten the generalizability of results. Despite a string of recent promising collaborations with individual agencies, researchers have understandably raised concerns over external validity. If agencies willing to collaborate with academics are unrepresentative of agencies at large, then collaborative field experiments, however carefully executed, may have little value outside the agencies in which they are conducted.

In this paper, we evaluated the nature and severity of selection into research collaborations with police agencies via two field experiments. Our results, precisely replicated across studies, offer several useful insights for applied researchers. First, we find no evidence that agencies which decline to discuss research collaborations are dissimilar to those that respond affirmatively across a range of agency and jurisdiction attributes. We also find that the vast majority of agencies we contacted did not respond at all. This low response rate underscores the difficulty of initiating academic collaborations with practitioners, and suggests the need to develop institutions that can faciliate such connections moving forward. However, our experimental results imply that many agencies who profess an openness to evidence-based policymaking may be engaging in cheap talk, as a mere mention of agency performance substantially depresses affirmative responses. Our analysis is confined to the initial stage of contacting agencies to develop research partnerships. As this process unfolds and the possibility of negative publicity that sometimes results from transparent research is made more apparent, it is possible that even more agencies would be unwilling to collaborate on evidence-based policy research. However, we also recognize that alternative mechanisms may be at play. Specifically, mentions of agency performance in an initial outreach message may have decreased trust in the research team. This reaction need not be limited to policing collaborations: given these results, researchers seeking collaborations with schools, legislatures, and a host of other institutions may face similar hurdles if performance evaluation is mentioned too hastily. Our results suggest that a more cautious approach to solicitation of these partnerships that seeks to build a relationship over several interactions before discussing the details of performance evaluations may be more fruitful, though such an approach would add costs to the already burdensome process of establishing research partnerships with practitioners. Future experiments could be deployed to disentangle these competing theories, and to test their validity in other policy domains.

Increasing openness to evidence-based policymaking offers a valuable opportunity to generate effective reforms in a range of social institutions. However, we have little systematic evidence on the demand for such collaborations by practitioners (but see Levine (2020)). This paper supplies such evidence, and provides a replicable template for future work in a range of policy domains. Accumulating additional scientific knowledge on the scale and determinants of the willingness of practitioners to collaborate with academics can serve to streamline and accelerate the process of policy experimentation.

References

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A. Olken and Julia Tobias. 2012. "Targeting the poor: evidence from a field experiment in Indonesia." *American Economic Review* 102(4):1206–40.

Alexander, Michelle. 2010. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. The New Press.

Allcott, Hunt. 2017. "Site selection bias in program evaluation." *The Quarterly Journal of Economics* 130(3):1117–1165.

Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57(1):289–300.

Carpenter, Daniel. 2014. *Reputation and power: organizational image and pharmaceutical regulation at the FDA*. Princeton University Press.

Cartwright, Nancy and Jeremy Hardie. 2012. *Evidence-based policy: A practical guide to doing it better*. Oxford University Press.

Edwards, Frank, Hedwig Lee and Michael Esposito. 2019. "Risk of being killed by police use of force in the United States by age, race–ethnicity, and sex." *Proceedings of the National*

Academy of Sciences .

Gelman, Andrew, Jeffrey Fagan and Alex Kiss. 2007. "An Analysis of the New York City Police Department's "Stop-and-Frisk'" Policy in the Context of Claims of Racial Bias." *Journal of the American Statistical Association* 102(429):813–823.

Gerber, Alan S. and Donald P. Green. 2012. *Field experiments: Design, analysis, and interpretation*. WW Norton.

Gerber, Alan S., James G. Gimpel, Donald P. Green and Daron R. Shaw. 2011. "How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment." *American Political Science Review* 105(1):135–150.

Glaser, J. and A. Charbonneau. 2018. *Making research matter: A psychologist's guide to public engagement*. American Psychological Association chapter Law enforcement: Finding common purpose.

Goff, Phillip Atiba, Dean Obermark, Nancy La Vigne, , Jennifer Yahner and Amanda Geller. 2016. The Science of Policing Equity: Measuring Fairness in the Austin Police Department. Technical report Center for Policing Equity & John Jay College of Criminal Justice.

Gunderson, A., E. Cohen, K. Jackson, T.S. Clark, A. Glynn and M.L. Owens. 2019. "Does Military Aid to Police Decrease Crime? Counterevidence from the Federal 1033 Program and Local Police Jurisdictions in the United States." *Working Paper* . `https://tinyurl.com/y66j6s8g`.

Knox, Dean, Will Lowe and Jonathan Mummolo. 2020. "Administrative Records Mask Racially Biased Policing." *American Political Science Review* 114:619–637.

Lerman, Amy and Vesla Weaver. 2014. *Arresting Citizenship: The Democratic Consequences of American Crime Control*. University of Chicago Press.

Levine, Adam Seth. 2020. "Why do practitioners want to connect with researchers? Evidence from a field experiment." *PS: Political Science & Politics* 53(4):712–717.

Litvack, Jennie I. and Claude Bodart. 1993. "User fees plus quality equals improved access to health care: results of a field experiment in Cameroon." *Social Science & Medicine* 37(3):369–383.

Madhani, Aamer. 2018. "Chicago police solved fewer than one in six homicides in the first half of 2018." *USA Today* . `https://bit.ly/2Q3yt1u`.

Moffitt, Susan L. 2010. "Promoting agency reputation through public advice: Advisory committee use in the FDA." *The Journal of Politics* 72(3):880–893.

Mummolo, Jonathan, Erik Peterson and Sean Westwood. 2019. "The Limits of Partisan Loyalty." *Political Behavior* pp. 1–24.

nj.com. 2019. "The Force Report.". `https://force.nj.com/`.

Orszag, Peter and Jim Nussle. 2017. "Policymaking commission offers a glimmer of hope in hyper-partisan Washington." *The Hill* . `https://tinyurl.com/ry8hvax`.

Westwood, Sean and Erik Peterson. 2019. "Compound political identity: How partisan and racial identities overlap and reinforce." *Available at SSRN 3417476* .

Westwood, Sean J, Erik Peterson and Yphtach Lelkes. 2019. "Are There Still Limits on Partisan Prejudice?" *Public Opinion Quarterly* 83(3):584–597.

Yokum, David, Anita Ravishankar and Alexander Coppock. 2019. "A randomized control trial evaluating the effects of police body-worn cameras." *Proceedings of the National Academy of Sciences* 116(21):10329–10332.

Which Police Departments Want Reform? Barriers to Evidence-Based Policymaking

Online Appendix

Contents

A1 Sampling Procedure

A1.1 Criteria for the N.J. and National Samples

The agencies contacted in the N.J. experiment were drawn from the list of agencies for which use-of-force data was tabulated by nj.com (nj.com, 2019). Excluding the N.J. State Police, which we also contacted, these 460 agencies represent close to 90% of local police and sheriff's agencies in the state.

The research team manually matched the agencies with contact information purchased from the National Public Safety Information Bureau (`https://www.safetysource.com/index.cfm?`). Roughly 150 agencies that did not list contact information were searched online or contacted by phone or email through the county clerk or agency itself, allowing us to contact all agencies. [1]

The agencies contacted in the national experiment were drawn from data on agencies which report crime data to the FBI through the Uniform Crime Reporting (UCR) program, roughly 23,000 agencies between 2013 and 2017. Before assigning treatment, we filtered these agencies according to several criteria. We excluded agencies: from New Jersey; with zero population in their jurisdiction according to the FBI; which were not local police and sheriff's departments; which report crime data through another nearby "parent" agency; which reported a mean five-year (2013-2017) violence crime clearance rate that fell outside the interval [0,1] (those outside the interval likely contained serious data errors); and which cleared zero crimes of any type in a given year, as these agencies are likely not reporting data. We also screened out the Long Beach, CA Police Department from the national sample prior to random assignment due to the aforementioned mistake in the pilot study. After applying these filters, we were left with about 9,800 agencies from which to sample. We then set about the task of pairing these 9,800 agencies to a list of contact information for police chiefs and sheriffs purchased from Power Almanac (`https://www.poweralmanac.com/`). In hopes of minimizing undeliverable emails, we changed our data source because Power Almanac updates the contact information on a continuing basis, with all agencies receiving updates twice annually. In the event a direct email was not available for a chief or sheriff, we used general agency emails or contact information for a lower level employee. Power Almanac is a private firm that continuously updates contact information for municipal government agencies. If agency contact information is not included in the Power Almanac database it is very likely that the agency does not publicize its email address. Because of the

[1] After validating the contact data we dropped Long Beach Township, N.J. Police Department from the sample because we mistakenly contacted Long Beach, CA Police Department. We manually verified this was the only such mistake.

scale of the data, we employed the probabilistic matching algorithm outlined in Enamorado, Fifield and Imai (2017) using the R package `fastLink` to merge agency data with contact information. We employed this algorithm separately for police and sheriff's agencies, and then matched based on code, agency name and street address listed in the UCR data and Power Almanac data. Following the match, we discarded agencies which did not have at least a 95% probability of being a correct match according to the model developed in Enamorado, Fifield and Imai (2017), as well as agencies for whom contact information could not be independently obtained by the research team via web searches. Following these steps, 7,702 agencies remained in sample, which oversee jurisdictions that are home to roughly 60% of the U.S. population according to FBI data. We drew 2,500 agencies from these 7,702 and assigned treatment based on the matched pair procedure we describe below.

A1.2 Matched Pair Design

To maximize statistical power, we implemented a matched pair design in the national experiment. The random assignment procedure went as follows. We first divided our sample of roughly 7,700 agencies into quartiles based on the population sizes of their jurisdictions. We then random drew half of the desired sample (2,500 agencies/2) from this population and assigned them to treatment. For each treated case, we identified a set of possible control cases that belonged to the same population quartile and U.S. same state. If such a set existed, we then randomly sampled one case from the possible control set, paired it with the treated case, and assigned it to the control condition. The intent of the conditional random assignment in this design, (as opposed to complete random assignment across the pooled sample) is that state and population size are prognostic of the dependent variable (response to our emails). By including indicators for these matched pairs during estimation, we thus sought to shrink the standard errors in our regressions by explaining variation in the outcome Gerber and Green (2012).

A1.3 Protocol in the event of returned emails and letters

Emails were sent to the list of purchased and manually collected NJ email addresses. Many emails were bounced back as undeliverable due to retirements, promotions or other organizational changes. In these cases, the research team made contact with either county clerks or police agencies again or searched for addresses online to obtain correct addresses. If our messages were blocked by SPAM filters, we relied on postal mail to deliver our treatments. In New Jersey there were no agencies who did not receive either our email or postal letter.

For the national study we sent an initial round of emails to the set of purchased email addresses. When we received notice that an email address had changed or that a agency head had retired we updated our contact information and sent out a new round of emails. As with New Jersey, some emails were blocked by SPAM filters. There were three agencies who we were unable to contact (by email or postal letter). These three agencies and their corresponding matched agency were dropped from our analysis.

B2 Experimental Protocol and Treatments

1. Agencies were emailed on a Monday. Responses were recorded either directly through a reply to our email message or by monitoring visits to our project website from specially constructed links in each email.

2. Agencies were emailed on a Monday. Responses were recorded either directly through a reply to our email message or by monitoring visits to our project website from specially constructed links in each email.

3. Any agency that did not respond to the first email was sent a second email eight days after the first email (the following Tuesday).

4. Any agency that did not respond to the second email was sent a third email eight days after the second email (the following Wednesday).

5. A paper letter with content similar to the emails (direct links were replaced with a URL) was then sent to be mailed eight days after the third email to any agency from whom we had not received a response.

6. A debrief email was sent the week of 11/25/2019 to all contacted agencies. A mailed version of the debrief was also sent.

Agencies could respond in three primary ways:

- **Yes**: Measured by clicking the yes link our emails or the yes button on our project website. Or by agreeing by direct email.
- **No**: Either explicitly measured through a no response on by link or on our website, or implicitly measured from a lack of a response to our messages. Or by declining by direct email.
- **"Learn more"**: Measured by link or on our website. This page displayed additional information about the project and offered an additional chance to say "yes" or "no" to our request. Or by asking for more information by direct email

We are currently in the process of following up with agencies who expressed interest in a possible collaboration and are in negotiations with several over the details of a joint project. Data collection for the national study ended the last week of November 2019.

B2.1 New Jersey Treatments

The following values were populated with data we collected and/or computed for each agency:

- **$name**: the name of the lead law enforcement officer
- **$events**: use of force incidents per officer between the years 2012 and 2016
- **$rank**: computed agency rank

In the control condition the ranking and confidentiality treatments were omitted. In the confidentiality condition the ranking treatment was omitted. In the ranking condition the confidentiality treatment was omitted. In the ranking and confidential condition both treatments were displayed in a random order.

Dear $name:

We are writing to invite you to collaborate on a research project conducted by researchers at <anonymized for review> and <anonymized for review> that aims to make both citizens and officers safer by reducing the frequency of violence during police-citizen interactions.

In exchange for your participation in a research project **we are offering to provide free consulting and data analysis services** that we believe could benefit your agency. We would also be willing to help your agency conduct an opinion survey of citizens in your jurisdiction to help identify areas where your agency could improve its performance. We are not seeking funds and will work pro bono.

[Ranking treatment: According to publicly available data (more information) your agency had an average of **$events** use of force incidents per officer between the years 2012 and 2016. By this metric, your agency ranked **$rank** out of 464 law enforcement agencies in New Jersey during this time (where a rank of 1 indicates the highest rate of using force).]

[Confidentiality treatment: If you agree to collaborate with us on a research project, the names of your agency and personnel will be kept confidential in any published research that we produce.]

We are not asking for a firm commitment now, but if you are interested in discussing a potential collaboration further, please click the appropriate link below and we will get in touch to continue the conversation.

- Yes, I am interested in discussing a research collaboration.
- I am not sure, but I would like more information
- No, I am not interested in discussing a research collaboration.

By way of background, we are non-partisan political scientists at <anonymized for review> and <anonymized for review> who received our PhDs from <anonymized for review>. We are trained in statistics, data analysis, experimental design and policy evaluation. We are seeking to partner with law enforcement agencies to help develop effective ways to improve police-citizen interactions, and are writing to gauge interest in conducting a collaborative research project with the aim of producing published, peer-reviewed research that could benefit the law enforcement community at large.

Please feel free to email us with any questions. Thank you for your consideration.

Sincerely,
<anonymized for review>

B2.2 New Jersey "Know more" website content

Thank you for your interest

By way of further introduction, we are faculty members in the Political Science/Government departments at <anonymized for review> and <anonymized for review>. We specialize in quantitative research on public agency performance and of public opinion. We would be interested in collaborating with your agency to conduct any of the following types of research projects, with the aim of publishing peer-reviewed research that can benefit your agency as well as the law enforcement community at large.

A public opinion survey assessing public perceptions of your agency's performance. This survey may serve to help you identify areas in which citizens perceive your agency can improve service.

A program evaluation of a new tactic. For example, we are interested in evaluating the effects of:

- Police-worn body cameras on police use of force
- The use of tactical teams on crime and agency reputation

Response items:

Yes, I am interested in discussing a research collaboration.
No, I am not interested in discussing a research collaboration.

B2.3 New Jersey Debrief

Dear $name:

We recently contacted your agency with an offer to discuss a possible research collaboration with faculty members at <anonymized for review> and <anonymized for review>. We are writing to supply some additional information. Depending on the version of the email you saw, we may or may not have included information on your agency's past performance (possibly relative to other agencies), or an offer of anonymity for your agency in any published findings. The purpose of this information was to gauge how various features of our message affected response rates. However, we note that our emails contained no false information and your responses will be kept confidential. If you have any questions or concerns, please feel free to contact the principal investigators <anonymized for review>. You can also contact <anonymized for review> Institutional Review Board at <anonymized for review>.
Sincerely,

<anonymized for review>

B2.4 National Sample Treatments

The following values were populated with data we collected and/or computed for each agency:

- **$title**: the title of the lead law enforcement officer
- **$name**: the name of the lead law enforcement officer
- **$cleared**: total cleared violent crimes occurring between 2013 and 2017
- **$total**: number of violent crimes between 2013 and 2017
- **$rank**: computed agency rank

In the control condition the ranking treatment was omitted. The full letter was shared in the treatment condition.

Dear $title $name:

We are writing to invite you to collaborate on a research project conducted by researchers at <anonymized for review> and <anonymized for review> that aims to study the efficacy of policing strategies designed to serve the interests of both officers and citizens.

In exchange for your participation in a research project **we are offering to provide free consulting and data analysis services** that we believe could benefit your agency. We would also be willing to help your agency conduct an opinion survey of citizens in your jurisdiction to help identify areas where your agency could improve its performance. We are not seeking funds and will work pro bono.

[Ranking treatment: According to publicly available data (more information) your agency cleared **$cleared** out of **$total** violent crimes between 2013 and 2017. By this metric, your agency ranked **$rank** out of approximately 2,500 law enforcement agencies we analyzed during this time, where a rank of 1 indicates the largest ("best") proportion of cleared violent crimes.]

We are not asking for a firm commitment now, but if you are interested in discussing a potential collaboration further, please click the appropriate link below and we will get in touch to continue the conversation.

- Yes, I am interested in discussing a research collaboration.
- I am not sure, but I would like more information
- No, I am not interested in discussing a research collaboration.

By way of background, we are non-partisan political scientists at <anonymized for review> and <anonymized for review> who received our PhDs from <anonymized for review>. We are trained in statistics, data analysis, experimental design and policy evaluation. We are seeking to partner with law enforcement agencies to help develop effective ways to improve police-citizen interactions, and are writing to gauge interest in conducting a collaborative research project with the aim of producing published, peer-reviewed research that could benefit the law enforcement community at large.

Please feel free to email us with any questions. Thank you for your consideration.

Sincerely,

<anonymized for review>

B2.5 National "Know more" website content

Thank you for your interest

By way of further introduction, we are faculty members in the Political Science/Government departments at <anonymized for review> and <anonymized for review>. We specialize in quantitative research on public agency performance and of public opinion. We would be interested in collaborating with your agency to conduct any of the following types of research projects, with the aim of publishing peer-reviewed research that can benefit your agency as well as the law enforcement community at large.

A public opinion survey assessing public perceptions of your agency's performance. This survey may serve to help you identify areas in which citizens perceive your agency can improve service.

A program evaluation of a new tactic. For example, we are interested in evaluating the effects of:

- Police-worn body cameras on police use of force
- The use of tactical teams on crime and agency reputation

Response options:

Yes, I am interested in discussing a research collaboration.
No, I am not interested in discussing a research collaboration.

B2.6 National Study Jersey Debrief

Dear $title $name:

We recently contacted your agency with an offer to discuss a possible research collaboration with faculty members at <anonymized for review> and <anonymized for review>.We are writing to supply some additional information. Depending on the version of the email you saw, we may or may not have included information on your agency's past performance (possibly relative to other agencies). The purpose of this information was to gauge how various features of our message affected response rates. However, we note that our emails contained no false information and your responses will be kept confidential. If you have any questions or concerns, please feel free to contact the principal investigators <anonymized for review>. You can also contact <anonymized for review> Institutional Review Board at <anonymized for review>.

Sincerely,

<anonymized for review>

C3 Data on Police Agencies

We merged data on contacted agencies with data on crime (Kaplan, 2019), fatal officer-involved shootings (Sullivan et al., 2018), agency personnel (DOJ, 2017), county-level election results (Data and Lab, 2018), and local U.S. Census records (Census, 2017).

C3.1 U.S. Census Data

We rely on U.S. Census data from the 2017 American Community Survey (Census, 2017). We sought to match agencies with demographic data measured at the geographic level of their jurisdictions. County agencies were paired with county census data, including parishes in Louisiana. Cities, towns, boroughs and other sub-county agencies were paired with either Census Designated Place (CDP) data, or county subdivision Census data, depending on how the Census classified localities. Five regional agencies which oversee multiple localities were paired with the sum (for count variables) or the means (for measures such as median household income) of Census data from those localities.

The analysis in Figure 2 is generated by separate bivariate regressions. The number of observations varies due to partial missing data across covariates. With the exception of census measures of percent of a jurisdiction living in poverty among the Asian American, "other race", American Indian and native Hawaiian categories, all variables display less than 10% missing data.

Table C1: Types of Census Data Paired with Agencies

	county	county subdivision	place	regional pd
No. of Agencies	616	449	1874	5

C3.2 Coding of agency covariates

Table C2: Coding Rules and Data Sources for Agency Covariates in Figure 2

variable	coding/source
Community Covariates:	
% "Other" race non-Hispanic	census data
% "Other" race non-Hispanic in poverty	census data
% "Other" race unemployed	census data
% African American	census data
% African Americans in poverty	census data
% Asian	census data
% Asian above poverty line	census data
% Hispanic	census data
% Hispanics in poverty	census data
%Male	census data
Median household income	census data
% Multiracial in poverty	census data
% Multiracial non-Hispanic	census data
% Native American and Alaskan native	census data
% Native American and Alaskan native in poverty	census data
% Native Hawaiian and Pacific Islander	census data
% Native Hawaiian and Pacific Islander in poverty	census data
% without health insurance	census data
Trump vote share (two party)	MIT Election Data \& Science Lab
Population	census data
% < 18 in poverty	census data
% > 25 with high school +	census data
% >25 with bachelors degree	census data
Population density	census data
% foreign born	census data
% non-citizen foreign born	census data
% Vacant housing units	census data
% White	census data
% Whites in poverty	census data
Median age	census data
Median age of male population	census data
Agency Covariates:	
Police employees per capita	total employees 2014 according to Law Enforcement Agency Roster (2016) count / census population for jurisdiction Male civilian employees per capita
employees in 2014 according to Law Enforcement Agency Roster (2016) count / total employees	
Shootings per capita	fatal officer-involved shootings according to Washington Post (2018) / census population for jurisdiction
Male officers per capita	male sworn officers in 2014 according to Law Enforcement Agency Roster (2016) count / total employees
Mean clearance for property crimes per capita	mean cleared index property crimes 2013-2017 (FBI) / census population for jurisdiction
Mean clearance for violent crime per capita	mean cleared index violent crimes 2013-2017 (FBI) / census population for jurisdiction
Mean clearance for murders per capita	mean cleared murders 2013-2017 (FBI) / census population for jurisdiction
Mean police shootings per capita	mean fatal officer-involved shootings according to Washington Post (2015-2018) / census population for jurisdiction
Officers killed in an accident per capita	mean officers accidentally killed 2013-2017 (FBI) / census population for jurisdiction
Officers assaulted per capita	mean officers assaulted 2013-2017 (FBI) / census population for jurisdiction
Officers killed by felon per capita	mean officers feloniously killed 2013-2017 (FBI) / census population for jurisdiction
Rapes per capita	mean rapes 2013-2017 (FBI) / census population for jurisdiction
Crime Covariates:	
Assaults per capita	mean assaults 2013-2017 (FBI) / census population for jurisdiction
Attempted rapes per capita	mean attempted rapes 2013-2017 (FBI) / census population for jurisdiction
Burglaries per capita	mean burglaries 2013-2017 (FBI) / census population for jurisdiction
Rape by force per capita	mean forcible rape 2013-2017 (FBI) / census population for jurisdiction
Manslaughters per capita	mean manslaughter 2013-2017 (FBI) / census population for jurisdiction
Vehicle thefts per capita	mean motor vehicle thefts 2013-2017 (FBI) / census population for jurisdiction
Murders per capita	mean murders 2013-2017 (FBI) / census population for jurisdiction
Robberies per capita	mean robberies 2013-2017 (FBI) / census population for jurisdiction
All crimes per capita	mean total crimes 2013-2017 (FBI) / census population for jurisdiction

D4 Random assignment balance tests

Table D1: New Jersey Balance

	Dependent variable:		
	Confidentiality Treatment	Ranking Treatment	Ranking and Confidentiality Treatment
	(1)	(2)	(3)
Intercept	0.14	0.28	0.32
	(0.22)	(0.23)	(0.23)
Population	−0.0001	−0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
Median Age	0.004	−0.002	−0.0001
	(0.004)	(0.004)	(0.004)
Male population	0.0001*	0.0000	−0.0001
	(0.0001)	(0.0001)	(0.0000)
Foreign born population	−0.0000	0.0000	−0.0000
	(0.0000)	(0.0000)	(0.0000)
	(0.0000)	(0.0000)	(0.0000)
Total crimes	0.0001	−0.0000	0.0000
	(0.0001)	(0.0001)	(0.0001)
Officers assaulted	−0.0002	−0.002	0.005
	(0.004)	(0.003)	(0.003)
White population	0.0000	0.0000	−0.0000
	(0.0000)	(0.0000)	(0.0000)
Native American/Alaskan Native population	−7.97	−7.96	8.22
	(8.69)	(8.62)	(12.20)
Asian Population	0.31	−0.20	0.09
	(0.31)	(0.24)	(0.30)
African American population	0.13	0.37	−0.26
	(0.27)	(0.28)	(0.20)
Hawaiian and Pacific Islander population	15.29	−2.49	1.53
	(8.85)	(7.34)	(11.98)
Hispanic Population	0.20	−0.14	0.10
	(0.21)	(0.19)	(0.21)
Murder clearance rate	−0.07	0.05	−0.03
	(0.04)	(0.04)	(0.03)
Violent crime clearance rate	0.004**	−0.001	−0.001
	(0.001)	(0.001)	(0.001)
Property crime clearance rate	−0.001	0.0004	−0.0002
	(0.0004)	(0.0005)	(0.0003)
Trump vote share (two party)	−0.26	0.11	−0.11
	(0.19)	(0.21)	(0.19)
Observations	460	460	460
R^2	0.05	0.01	0.02
F Statistic (df = 17; 442)	1.43	0.38	0.58
Note:			*p<0.05; **p<0.01; ***p<0.001

Table D2: National Study Balance

	Dependent variable:
	Assigned to Ranking Treatment
Intercept	1.50***
	(0.08)
Population	−0.0000
	(0.0000)
Median Age	0.0003
	(0.001)
Male population	0.0000
	(0.0000)
Foreign born population	0.0000
	(0.0000)
	(0.0000)
Total crimes	0.00
	(0.0000)
Officers assaulted	−0.001
	(0.0005)
White population	−0.0000
	(0.0000)
Native American/Alaskan Native population	−0.06
	(0.28)
Asian Population	−0.03
	(0.30)
African American population	−0.04
	(0.07)
Hawaiian and Pacific Islander population	−0.34
	(2.84)
Hispanic Population	0.04
	(0.07)
Murder clearance rate	0.005
	(0.01)
Violent crime clearance rate	−0.0000
	(0.0001)
Property crime clearance rate	−0.0000
	(0.0001)
Trump vote share (two party)	−0.0002
	(0.001)
Observations	2,482
R^2	0.002
F Statistic	0.27 (df = 17; 2464)
Note:	*p<0.05; **p<0.01; ***p<0.001

E5 Tabular Results

E5.1 Geographic Distribution of Sample Agencies

Table E1: Figure 1: Sampled agencies by state

State	Number of sampled agencies
AL	66
AR	52
AZ	24
CA	92
CO	42
CT	34
DE	4
FL	74
GA	94
HI	2
IA	54
ID	32
IN	32
KS	70
KY	58
LA	32
MA	88
MD	24
ME	36
MI	118
MN	78
MO	116
MS	10
MT	22
NE	16
NC	74
ND	10
NH	32
NJ	460
NM	30
NV	12
NY	68
OH	110
OK	34
OR	32
PA	154
RI	12
SC	44
SD	20
TN	78
TX	222
UT	24
VA	56
VT	18
WA	56
WI	86
WV	24
WY	18
AK	0
IL	0

E5.2 Regression Estimates

Table E2: Estimates from Figure 2: Correlates of responding

Variable	Estimate	Uncorrected p-value	Benjamini-Hochberg corrected p-value
Officers assaulted per capita	-8.39	0.01	0.39
Population	0.19	0.91	0.94
%Male	0.89	0.77	0.94
% White	-2.48	0.54	0.94
% African American	2.97	0.37	0.94
% Native American and Alaskan native	1.11	0.70	0.94
% Asian	6.23	0.31	0.94
% Native Hawaiian and Pacific Islander	1.83	0.85	0.94
% "Other" race non-Hispanic	-0.93	0.33	0.94
% Multiracial non-Hispanic	3.42	0.48	0.94
% Hispanic	-2.31	0.29	0.94
% > 25 with high school +	-5.50	0.15	0.94
% >25 with bachelors degree	5.85	0.20	0.94
% "Other" race unemployed	-0.50	0.87	0.94
% Whites in poverty	-4.85	0.13	0.94
% African Americans in poverty	1.88	0.57	0.94
% Native American and Alaskan native in poverty	3.56	0.30	0.94
% Asian above poverty line	-1.37	0.68	0.94
% Native Hawaiian and Pacific Islander in poverty	-3.21	0.36	0.94
% "Other" race non-Hispanic in poverty	-4.01	0.29	0.94
% Multiracial in poverty	-3.79	0.38	0.94
% Hispanics in poverty	-3.35	0.48	0.94
% Vacant housing units	-1.62	0.71	0.94
Male officers per capita	-4.42	0.36	0.94
Male civilian employees per capita	3.22	0.41	0.94
Officers killed by felon per capita	-2.62	0.31	0.94
Officers killed in an accident per capita	-0.52	0.53	0.94
Murders per capita	3.84	0.51	0.94
Manslaughters per capita	-1.23	0.73	0.94
Rapes per capita	1.35	0.79	0.94
Rape by force per capita	1.54	0.76	0.94
Attempted rapes per capita	-1.84	0.73	0.94
Robberies per capita	0.02	0.99	0.99
Assaults per capita	0.57	0.91	0.94
Burglaries per capita	-1.17	0.81	0.94
Vehicle thefts per capita	1.53	0.67	0.94
All crimes per capita	1.02	0.83	0.94
Mean clearance for murders per capita	4.08	0.62	0.94
Mean clearance for violent crime per capita	-0.87	0.87	0.94
Mean clearance for property crimes per capita	1.57	0.72	0.94
Median age	-1.32	0.64	0.94
Median age of male population	-1.05	0.72	0.94
Median household income	1.88	0.73	0.94
Trump vote share (two party)	-8.47	0.13	0.94

Table E3: Estimates from Figure 3: New Jersey Regression Results

	Dependent variable:		
	Agreed to Discuss Collaboration		
	(1)	(2)	(3)
Intercept	0.13***	0.21**	0.19**
	(0.03)	(0.07)	(0.06)
Confidential Condition	−0.02	−0.14	−0.09
	(0.04)	(0.09)	(0.08)
Ranking Condition	−0.09*	−0.17*	−0.16*
	(0.04)	(0.08)	(0.06)
Ranking and Anonymous Condition	−0.04	−0.04	−0.02
	(0.04)	(0.10)	(0.09)
Numeric Rank		−0.0003	
		(0.0003)	
Confidential Condition X Numeric Rank		0.001	
		(0.0003)	
Ranking Condition X Numeric Rank		0.0004	
		(0.0003)	
Ranking and Anonymous Condition X Numeric Rank		0.0000	
		(0.0003)	
Middle Rank Tercile			−0.08
			(0.08)
Top Rank Tercile			−0.10
			(0.08)
Confidential Condition X Middle Rank Tercile			0.08
			(0.11)
Ranking Condition X Middle Rank Tercile			0.11
			(0.09)
Ranking and Anonymous Condition X Middle Rank Tercile			−0.02
			(0.11)
Confidential Condition X Top Rank Tercile			0.14
			(0.11)
Ranking Condition X Top Rank Tercile			0.13
			(0.09)
Ranking and Anonymous Condition X Top Rank Tercile			−0.01
			(0.11)
Observations	462	462	462
R^2	0.01	0.02	0.03
Note:	*p<0.05; **p<0.01; ***p<0.001		

Robust standard errors reported

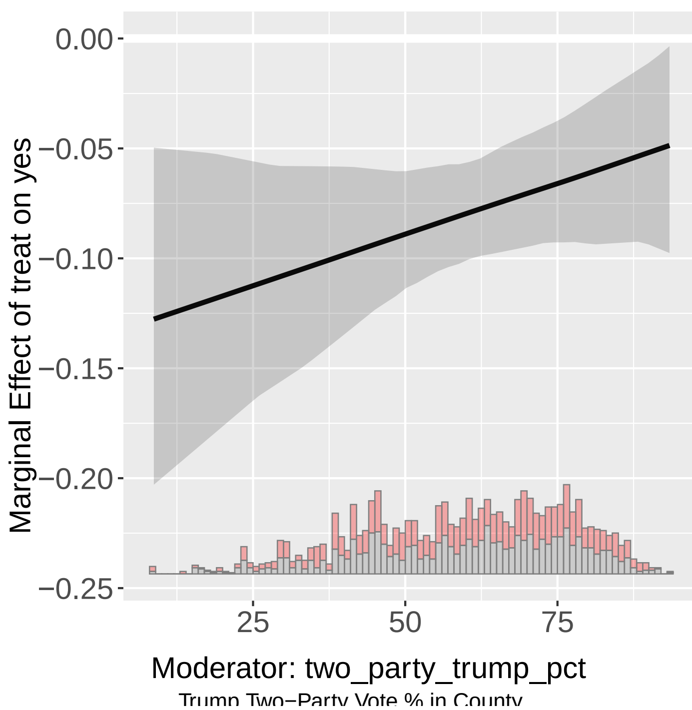
Table E4: Estimates from Figure 3: National Study Regression Results

	Dependent variable:			
	Agreed to Discuss Collaboration			
	(1)	(2)	(3)	(4)
Intercept	0.15***	0.04***	0.01	0.01
	(0.01)	(0.01)	(0.04)	(0.03)
Ranking Condition	−0.08***	−0.08***	−0.07	−0.05
	(0.01)	(0.02)	(0.04)	(0.04)
Numeric Rank			0.0000	
			(0.0000)	
Middle Rank Tercile				0.06
				(0.04)
Top Rank Tercile				0.04
				(0.04)
Ranking Condition X Numeric Rank			−0.0000	
			(0.0000)	
Ranking Condition X Middle Rank Tercile				−0.07
				(0.06)
Ranking Condition X Top Rank Tercile				−0.03
				(0.06)
Matched Pair FE	No	Yes	Yes	Yes
Observations	2,482	2,482	2,482	2,482
R^2	0.02	0.52	0.52	0.53
Note:	*p<0.05; **p<0.01; ***p<0.001			

Table E5: Trump Two Party Vote Share and Agreement

	Dependent variable:
	Agreed to Discuss Collaboration
Intercept	0.08
	(0.04)
Ranking Condition	−0.09*
	(0.04)
Middle Trump Vote Share Tercile	0.01
	(0.05)
Top Trump Vote Share Tercile	−0.07
	(0.05)
Ranking Condition X Middle Trump Vote Share Tercile	−0.03
	(0.06)
Ranking Condition X Top Trump Vote Share Tercile	0.05
	(0.05)
Matched Pair FE	Yes
Observations	2,482
R^2	0.53
Note:	*p<0.05; **p<0.01; ***p<0.001

Figure E1: **County Trump Vote Share Does Not Condition Response to Treatment.** The figure shows the marginal effect of the agency performance cue in the national experiment conditional on the two party Trump vote share in each county. To relax functional form assumptions, the marginal effect was computed across the range of the moderator using the flexible kernel estimator recommended in Hainmueller, Mummolo and Xu (2019). The stacked histogram along the bottom of the figure shows the distribution of treatment assignment at various levels of the moderator. The effect of the performance cue among counties with low, medium and high levels of Trump support are statistically indistinguishable.



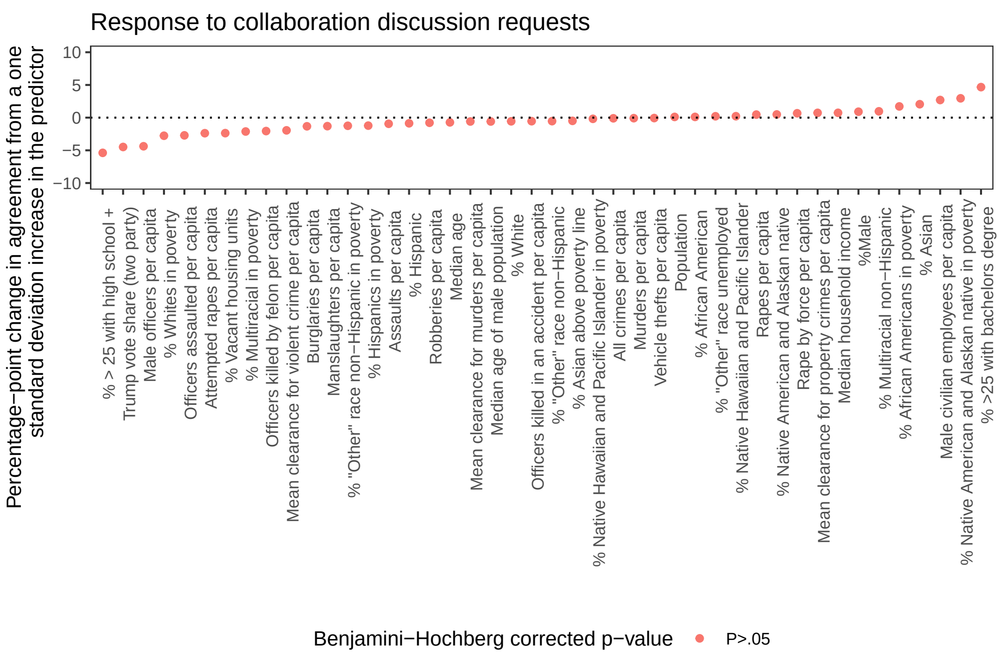


Figure E2: **Alternative Figure 2: Using Data from All Agencies**

F6 Pre-registration plan for national experiment

Note: In our pre-analysis plan, we also planned on examining an alternate coding of the dependent variable in which agencies who indicated they were already collaborating with another research team were coded as responding affirmatively. We omit this specification because these responses only occurred twice.

EGAP Registry Form Schema

Note from EGAP: while the standard workflow is down, this form replaces the registration form on egap.org. For this alternate workflow, the time/date that your email is sent will become the timestamp for your registration. It may still take up to three business days to review, upload, and post your submission, but the timestamp will be locked in as described.

B1 Title of Study – Determinants of Academic Collaborations with Law Enforcement Agencies

B2 Authors

<anonymized for review>
B3 Acknowledgements –

B4 Is one of the study authors a university faculty member? – multiple choice (SELECT ONE)
Yes

B5 Is this Registration Prospective or Retrospective? – multiple choice (SELECT ONE)

Registration prior to realization of outcomes

B6 Is this an experimental study? – multiple choice (SELECT ONE)
Yes

B7 Date of start of study – 09/16/2019

B8 Gate date – 09/16/2020

B9 Was this design presented at an EGAP meeting? – multiple choice (SELECT ONE)
No

B10 Is there a pre-analysis plan associated with this registration? – multiple choice (SELECT ONE)
Yes

For the next three fields (C1-C3), the response box is a long answer plain text box. Please try to limit your response to 300 words at most, and use your pre-analysis plan to elaborate further if necessary. Also, the plain text field limits formatting, so please do not include bullet point lists with multiple indentations, footnotes, tables, images, or other complicated formatting.

C1 Background and explanation of rationale – Police agencies are increasingly partnering with academic researchers to scientifically evaluate the effects of various police tactics and reforms in order to lower crime rates, improve the quality of police-citizen interactions and reduce the rate of police brutality, among other goals. But researchers have long suspected that the results of such collaborations suffer from selection bias: agencies which are in the most need of reform may be less likely to collaborate with researchers, while agencies that are performing well may be eager to advertise their performance via academic collaborations. This study will use large-scale field experiments to evaluate whether such selection bias exists and if so, what causes it and how it can be mitigated.

C2 What are the hypotheses to be tested/quantities of interest to be estimated? –

We predict that there will be a relationship between an agency's response and the agency's performance on measures of crime. We seek to measure the size of this bias. This is an extension of a successful pilot study (conducted with roughly 450 agencies New Jersey). We extend the design in two ways. First, we seek to replicate our findings with a different measure of performance (violent crime clearance instead of use of force). Second, we seek to replicate our results on a national sample.

This study will make use of an original email list of police administrators/agencies in the United States. The list is an augmented version of the 2016 Law Enforcement Agency Roster (LEAR), a federal survey of every law enforcement agency in the United States. The list contains basic information on over 15,000 state and local agencies, including mailing addresses, but does not include the names of administrators (police chiefs and sheriffs) or email addresses. We have linked these data with the FBI's Uniform Crime Report. This allows us to rank agencies on their violent crime clearance rate (the ratio of closed violent crime cases to total violent crimes).

In our pre-test we identified an effect of our main treatment relative to our control, which alerted agencies to their statewide ranking on the measure of uses of force per officer. This treatment lowered response rates by approximately 8 percentage points. We anticipate another negative effect by alerting agencies to violent crime clearance rates.

To maximize power we will generate a national sample of matched pairs of agencies, one of which will be randomly assigned to receive the control version of our invitation and one which will receive the treatment version. These pairs will be matched on state and jurisdiction size (population, binned in quartiles). The universe of agencies will be local police and sheriff's agencies (excluding those in U.S. territories) who recorded a violent crime clearance rate between 0 and 1 inclusive, and those which recorded non-missing data on needed variables. Based on a power analysis, we will send messages to a national sample of approximately 2,500 agencies that will exclude NJ.

Following our pretest we will then send emails and (if needed) posted letters to the sampled agencies. The content of these emails/letters are attached.

We will also pair each agency with publicly available data on crime, police-involved shootings and administrative data such as agency budgets and staff sizes in order to see how responses to our email vary with these metrics for a purely descriptive analysis.

Experimental conditions:

1. Control message:

See attached.

2. Performance Treatment:

This condition add texts on the performance of the agency closing violent crime cases relative to other American police departments. See attached.

Mode and protocol for contact:

We will consider two modes of contact: email and post.

We will send three emails spaced eight days apart. The first email will be sent on a Monday, then a Tuesday and finally on a Wednesday. To mitigate concerns that these emails will be trapped by SPAM filters we will send the messages from a <anonymized for review> SMTP server and will space each message by a randomly drawn period [500 milliseconds - 20,000 milliseconds].

If an agency does not respond to these emails we will send the content of the assigned email by mail. The posted letter will be sent on <anonymized for review> letterhead.

Dependent measures:

There will be three primary dependent measures:

1. Each email/letter will include links indicating a willingness to participate in the project.
2. After clicking on the "yes" or "more information" links respondents will be taken to a webpage with additional information on the project. They will be able to send us text on what they would like to gain from a collaboration.
3. Respondents can also directly respond to our emails/letters. We will record the text of these responses.

C3 How will these hypotheses be tested? –

1. We will compute the average treatment effect of our experimental intervention by comparing the rate of affirmative responses among those in the treatment arm to the same rate in the control arm via OLS, conditional on a set of dummy variables for each matched pair. Affirmative responses will be coded in two ways: 1) Those indicating "yes" and 2) Those either indicating "yes" or those who decline but indicate the reason is that they are already collaborating with another institution on research.

2. We will estimate an interaction model which adds to the specification in 1. a variable

measuring how each agency ranks nationally on violent crime clearance rates as well as a multiplicative term interacting this rank variable and the treatment indicator. Our expectation is that poorly ranked agencies will respond more negatively to treatment than well-ranked agencies, since the treatment will inspire greater concern over reputational damage among those low-ranked group. We will use the procedures described in Hainmueller, Mummolo and Xu (2019) to estimate this continuous interaction.

C4 Country – U.S.A.

C5 Sample Size (of Units) – approximately 2,500

C6 Was a power analysis conducted prior to data collection? – multiple choice (SELECT ONE)
Yes

C7 Has this research received Institutional Review Board (IRB) or ethics committee approval? – multiple choice (SELECT ONE)
Yes

C8 IRB Number – <anonymized for review>

C9 Date of IRB Approval – 09/13/2019

C10 Will the intervention be implemented by the researcher or a third party? If a third party, please provide the name. – multiple choice (SELECT AS MANY AS APPLICABLE)
Researchers

C11 Did any of the research team receive remuneration from the implementing agency for taking part in this research? – multiple choice (SELECT ONE)
N/A

C12 If relevant, is there an advance agreement with the implementation group that all results can be published? – multiple choice (SELECT ONE)
N/A

C13 JEL classification(s) – short answer; please provide alphanumeric code(s)

Methodology – select all that apply
Experimental Design

Field Experiments
Statistics
Survey Methodology

Policy – select all that apply
Conflict and Violence
Governance

Certification – indicate agreement
By submitting this form and accompanying documents with EGAP, I confirm that I have rights to put this information in the public domain and I understand that this information will remain on the EGAP registry in perpetuity, regardless of whether the research is subsequently implemented or not.

We agree to these terms.

Confirmation – indicate agreement
You should receive a confirmation of your registration within three business days. Your registration is considered complete only when confirmation is received. If you do not receive confirmation within three business days please contact paps@egap.org. Hitting SAVE at the bottom of this page will submit the registration. Please only do so when you are ready to submit. ONCE YOU HAVE HIT SAVE AT THE BOTTOM OF THIS PAGE PLEASE DO NOT HIT THE BACK BUTTON. Doing so creates multiple registrations, and we will delete all but the most recent. If you accidentally created multiple registrations, please contact paps@egap.org

We agree to these terms.

Note from EGAP: while the standard workflow is down, this form replaces the registration form on egap.org. For this alternate workflow, the time/date that your email is sent will become the timestamp for your registration. It may still take up to three business days to review, upload, and post your submission, but the timestamp will be locked in as described.

Additional Documentation – please attach your pre-analysis plan, survey instrument, or any other files associated with the registration (files must be under 5MB)

G7 Interaction Model Results

We estimated models that interacted treatment assignment in the national experiment with agency rank on violent crime clearance rates. We exclude the N.J. sample from this analysis. Because ranks for the N.J. experiment were computed within the state, they are not interchangeable with ranks in the nationwide sample. The interaction model would also be severely underpowered if we estimated it using only the N.J. sample.

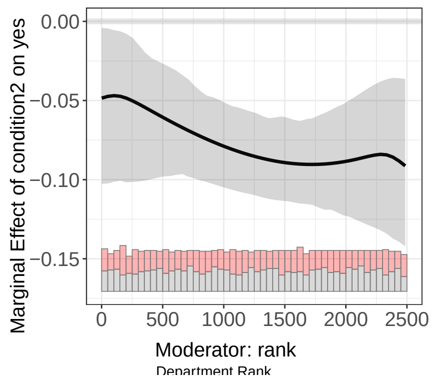
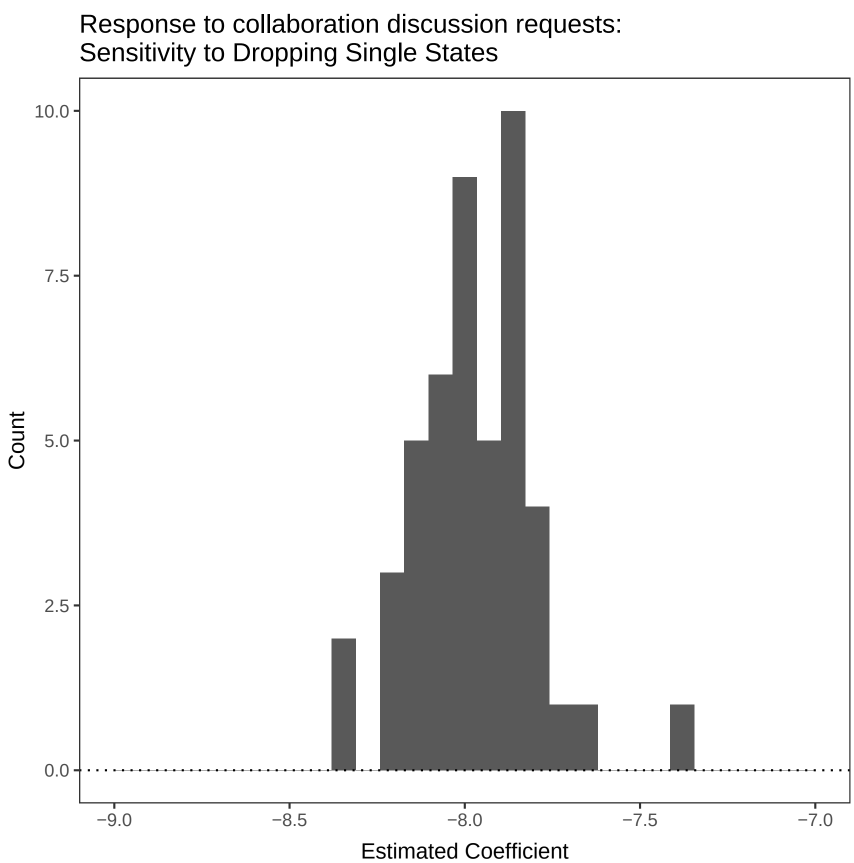


Figure G1: **Agency Performance Does Not Condition Response to Treatment.** The figure shows the marginal effect of the agency performance cue in the national experiment conditional on each agency's rank on violent crime clearance rates among the $\approx$ 2,500 agencies contacted (Hainmueller, Mummolo and Xu, 2019). The stacked histogram shows the distribution of treatment across levels of the moderator. The treatment effect among low, medium and high performing agencies are statistically indistinguishable.

H8 Sensitivity Analysis

Figure H1: **No Single State Drives Our Results.** This figures shows estimates derived from a dataset that iteratively drops one of the states included in the sample. Results are consistent across the 47 models.



I9 Alternative Mechanisms

One drawback of our research design is that we are unable to observe how agencies would have responded had they been contacted by researchers from institutions with different reputations. Our research team is drawn from multiple institutions, alleviating concerns that one particular university's reputation is driving results. But in the absence of more familiarity with the research team, our invitations, and our performance interventions in particular, could have been viewed with high levels of suspicion, prompting concerns that we lacked objectivity.

This concern was reinforced when we received an email from an agency in a very conservative state declining our invitation, which stated:

> "I have spoken to my administrative staff and all agree we should not participate in your study. Some had reservations on both colleges being 'liberal' schools. Thank you."

For this agency, at least, the partisan reputation of our universities (or perhaps, academia in general) precluded collaboration. If this interpretation were widespread, we would interpret our results quite differently.

This is a reasonable concern, but we do not think the perceived partisanship of the research team is producing these results. If it were, we expect agencies from more conservative areas would decline our invitations at higher rates. But as Figure 2 shows, county-level electoral support for Donald Trump does not predict response ($\beta = -2.20, p = 0.85$). In a separate analysis, we interact treatment in the national experiment with Trump's county vote share, and find no evidence of heterogeneous effects (see Online Appendix Figure E1 and Table E5).[2]

These results suggest our findings are not simply due to the perceived partisanship of the research team. The fact that even agencies told they were performing superbly—a cue unlikely to engender mistrust—reacted negatively, also suggests performance evaluation suppresses responses in general. Still, we acknowledge there exist multiple ways in which our messages may have triggered reputational concerns that we cannot distinguish. We discuss the implications of these various mechanisms in the following section.

[2]This analysis was not pre-registered, but we conduct it anyway in response to aforementioned email and feedback we received from colleagues concerning alternative mechanisms.

J10 Alternative Coding

In the national study a single agency responded by email in a way that could either be interpreted as a "yes" or a request for additional information. Out of an abundance of caution, we coded this agency as a "no" in our primary analysis. Below we show that coding this agency as a "yes" has no meaningful effect on our results.

Table J1: Alternate Coding for National Study

	Dependent variable:			
	Agreed to Discuss Collaboration			
	(1)	(2)	(3)	(4)
Intercept	0.15***	0.04***	0.01	0.005
	(0.01)	(0.01)	(0.04)	(0.03)
Ranking Condition	−0.08***	−0.08***	−0.06	−0.04
	(0.01)	(0.02)	(0.04)	(0.04)
Numeric Rank			0.0000	
			(0.0000)	
Middle Rank Tercile				0.07
				(0.04)
Top Rank Tercile				0.04
				(0.04)
Ranking Condition X Numeric Rank			−0.0000	
			(0.0000)	
Ranking Condition X Middle Rank Tercile				−0.07
				(0.06)
Ranking Condition X Top Rank Tercile				−0.03
				(0.06)
Matched Pair FE	No	Yes	Yes	Yes
Observations	2,482	2,482	2,482	2,482
R^2	0.02	0.52	0.52	0.52
Note:	*p<0.05; **p<0.01; ***p<0.001			

K11 Detail on Covariates

Community demographics:

- % > 25 with high school +
- % Whites in poverty
- Trump vote share (two party)
- % Ötherr̈ace non-Hispanic in poverty
- % Multiracial in poverty
- % Hispanics in poverty
- % Native Hawaiian and Pacific Islander in poverty
- %Male
- % Native American and Alaskan native
- % White
- % Hispanic
- % African American
- % Native American and Alaskan native in poverty
- % >25 with bachelors degree
- % Asian
- % Native Hawaiian and Pacific Islander
- % Multiracial non-Hispanic
- % Ötherr̈ace non-Hispanic
- % African Americans in poverty
- Median household income
- % Vacant housing units
- % Asian above poverty line
- Median age
- Median age of male population
- Population
- % Ötherr̈ace unemployed

Agency statistics:

- Officers assaulted per capita
- Male officers per capita
- Officers killed by felon per capita
- Mean clearance for violent crime per capita
- Officers killed in an accident per capita

- Mean clearance for property crimes per capita
- Male civilian employees per capita
- Mean clearance for murders per capita

Crime statistics:

- Attempted rapes per capita
- Manslaughters per capita
- Burglaries per capita
- Robberies per capita
- Assaults per capita
- All crimes per capita
- Rapes per capita
- Rape by force per capita
- Vehicle thefts per capita
- Murders per capita

L12 Differences between Police Departments and Sheriff Departments

References

Census, U.S. 2017. Selected characteristics, 2012-2017 5-Year American Community Survey. Technical report U.S. Census Bureau.

Data, MIT Election and Science Lab. 2018. "County Presidential Election Returns 2000-2016.".

DOJ. 2017. Law Enforcement Agency Roster (LEAR), 2016. Technical report United States Department of Justice, Office of Justice Programs, Bureau of Justice Statistics Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]: . `https://doi.org/10.3886/ICPSR36697.v1`.

Enamorado, Ted, Benjamin Fifield and Kosuke Imai. 2017. "Using a Probabilistic Model to Assist Merging of Large-scale Administrative Records." *Working Paper* .

Gerber, Alan S. and Donald P. Green. 2012. *Field experiments: Design, analysis, and interpretation*. WW Norton.

Table L1: Comparison of Treatment Effect Between Police Departments and Sheriff Departments. (Note: these tests were conducted at the request of an anonymous reviewer and were not included in our pre-analysis plan.)

	Dependent variable:	
	Agreed to Discuss Collaboration	
	(1)	(2)
Intercept	0.15***	0.01
	(0.02)	(0.04)
Ranking Condition	−0.09***	−0.11*
	(0.02)	(0.04)
Numeric Rank	0.005	0.03
	(0.02)	(0.04)
Sheriff Department	0.02	0.04
	(0.03)	(0.05)
Matched Pair FE	No	Yes
Observations	2,471	2,471
R^2	0.02	0.54
Note:	*p<0.05; **p<0.01; ***p<0.001	

Hainmueller, Jens, Jonathan Mummolo and Yiqing Xu. 2019. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice." *Political Analysis* 27(2):163–192.

Kaplan, Jacob. 2019. Jacob KaplanâĂŹs Concatenated Files: Uniform Crime Reporting Program Data: Offenses Known and Clearances by Arrest, 1960-2017. Technical report Inter-university Consortium for Political and Social Research [distributor] Ann Arbor, MI: .

nj.com. 2019. "The Force Report.". `https://force.nj.com/`.

Sullivan, John, Zane Anthony, Julie Tate and Jennifer Jenkins. 2018. "Nationwide, police shot and killed nearly 1,000 people in 2017." *The Washington Post* . `https://wapo.st/2Q5AAC7`.