J. Benefit Cost Anal. 2017; 8(3):348–359
doi:10.1017/bca.2017.26
© Society for Benefit-Cost Analysis, 2017

Rebecca Goldstein

Evaluating Proactive Police Units: A Case Study of Retrospective Benefit-Cost Analysis with Nonexperimental Data

Abstract: In recent decades, advocates for police reform on the political left and right have proposed numerous changes to how street-level policing operates. One such proposed reform, which has been adopted in jurisdictions nationwide, is "proactive policing," that is, policing strategies based on the notion that by proactively regulating minor offenses, the police can reduce both serious crime and fear of crime in the community. Yet, as with many proposed police reforms, researchers have not undertaken a through benefit-cost analysis of proactive policing. This article lays out strategies for estimating the impacts of proactive policing, including direct, indirect, and distributional impacts. First, I describe *quasi-experimental approaches*, which entail partnerships between researchers and police departments and would be particularly useful when a municipality is considering a move to proactive policing in the first instance, expansion of small-scale proactive policing to a larger area, or the introduction of particular new tactics. Second, I describe *non-experimental retrospective approaches*, including conventional regression analysis, which can also allow researchers to estimate the effects of proactive policing. I discuss potential threats to validity for both strategies. I close by describing the data that researchers wishing to engage in benefit-cost analysis of proactive policing would need in order to do so.

Keywords: law enforcement; local; police; regional; research methods; state; urban.

JEL classifications: C; K.

1 Introduction

Although there are no reliable data on how many U.S. police departments have proactive patrol units, the strategy has been widespread in various forms since the 1990s. Proactive policing, broadly defined, is any policing strategy based on the notion that "by proactively regulating minor offenses, the police can reduce

Rebecca Goldstein: Harvard University, Government, Cambridge, Massachusetts 02138, USA, e-mail: goldsteinr@post.harvard.edu

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press

both serious crime and fear of crime among the community by sending a message that crime will not be tolerated" (Nix & Rojek, 2014). Proactive patrol units engage in policing activities aligned with this idea. These tactics can include, but are not limited to, the aggressive use of traffic tickets to search vehicles for contraband and firearms, frequent use of in-person stops to search persons for contraband and firearms, and generally increasing staff levels in high-crime areas in order to "increase the number of citations and/or arrests to deter more serious crime" (Nix & Rojek, 2014).

A strategy of proactive policing, like nearly all policy interventions, comes with both benefits and costs, both direct and indirect. The most direct benefits of proactive policing are that it allows the police to arrest individuals who pose a danger to the community, as well as to confiscate illegal drugs and firearms. Making these arrests and confiscating contraband directly benefit community safety. The benefits of proactive policing would be all the greater if it also deterred more serious crimes, reduced fear of crime among community members, or led to greater public or private investment in particular neighborhoods.

With respect to costs, one direct cost of proactive policing is that it requires more police officers performing different tasks, as compared with reactive policing, and thus imposes costs on police departments. Because proactive policing demands significant investment of officer time, implementation will often require either shifting of existing officers into proactive policing (and thus away from other priorities), or, alternatively, maintenance of current staffing levels for other priorities while adding proactive policing patrols (thus imposing labor costs on departments). In addition, departments will bear the costs of training officers for engagement in proactive policing and ensuring compliance with a department's internal rules and applicable legal rules governing policing tactics. Proactive policing will also impose direct costs on individuals who are stopped by policing as part of the proactive policing but would not have been stopped otherwise. Proactive policing can impose several types of costs on individuals who are stopped: costs in time, money, and psychological harm. Indirect costs of proactive policing include the costs borne by the public, such as erosion of police–community relations and increased arrests and incarcerations of community members.

The costs of proactive policing will not fall on all individuals or communities equally. Instead, these costs will fall predominately on individuals and communities who tend to have more frequent involuntary encounters with police. Studies of past proactive policing efforts provide strong evidence that the costs of proactive policing policies fall disproportionately on some demographic groups. A seminal study of New York City's "stop-and-frisk" policy, for example, found that "persons of African and Hispanic descent were stopped more frequently than Whites, even after controlling for precinct variability and race-specific estimates of crime

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press

participation" (Gelman, Fagan & Kiss, 2007, p. 1). Proactive policing – at least as it was practiced in New York City – also results in stops of substantially more men than women and stops of substantially more young people than middle-aged and elderly people (New York Civil Liberties Union, 2017). For nearly any proactive policing policy that a department considers adopting, the policy's costs will fall predominately on those who are young, male, or non-White (or those who fall into several of those demographic groups). In this issue, Cohen (2017) sets out the impacts that racially targeted police encounters have on the individuals targeted, on other members of the targeted individuals' demographic group, and on society as a whole.[1]

The full effects of introducing proactive policing, including these benefits and costs, cannot be estimated with certainty without a fully randomized control trial (RCT), in which some neighborhoods are randomly assigned to proactive patrols and other similar neighborhoods are randomly assigned to standard patrols. While RCTs are the gold standard in social science research (Jones & Podolsky, 2015), they can be extremely challenging to conduct in the context of police reform. The key research design challenge of RCTs evaluating policing policy is that they require the police to treat similarly situated high-crime neighborhoods differently for no other reason than those neighborhoods' randomization into either the treatment or control group. While typically possible in theory, such a randomization can create serious challenges in practice. Communities may object to a policy that allocates benefits and burdens (or perceived benefits and burdens) at random (Boruch, 2015); policymakers may be hesitant to implement policies through a randomization protocol; and, relatedly, policymakers confident in the likely effects of a possible reform and feeling pressure to take action may not wish to take the more incremental steps necessary for a valid RCT. For these reasons, an RCT will often not be possible in practice. Researchers may thus wish to evaluate proactive enforcement strategies in the absence of an RCT.

This article seeks to inform researchers of police practices, both within police departments and in academia, of the most accurate ways to fully evaluate the benefits and costs of various police reforms in the absence of an RCT, without falling victim to the many pitfalls of simple regression-based or before-and-after studies. Because police departments may be hesitant to treat similarly situated high-crime neighborhoods differently, studies that use existing data will often be researchers' best options for evaluation of police policy reforms, including various forms of proactive policing.

1 Social scientists have developed a number of ways to account for distributional effects in benefit-cost analysis. One tool is "distributional weighting," which some have endorsed as a means of showing policymakers how the benefits and costs of a policy fall on different populations (Weisbach, 2015; Adler, 2016).

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press

There are two basic frameworks for approaching this evaluation challenge: a quasi-experimental approach and a nonexperimental retrospective approach. In this article, I discuss each approach in turn, highlighting advantages and disadvantages, along with what each would entail in terms of police cooperation and data collection. I also discuss threats to validity for each approach and how those threats could be addressed.

2 Quasi-experimental approaches

The key virtue of an RCT is that, on average, there are no differences between treatment and control units on observed or unobserved background covariates (Imbens & Rubin, 2015, p. 40). While this covariate balance is impossible without active randomization, a quasi-experiment can emulate this setup in at least two ways for evaluation of proactive policing.

First, some neighborhoods could be assigned to proactive enforcement while others are assigned to standard patrol during the course of a staggered rollout of proactive enforcement across a city. Such a strategy may not be feasible for a municipality that already has a municipality-wide practice of proactive policing, but it could be effective in several other cases. Some examples include a municipality considering a move to proactive policing in the first instance, a municipality considering expansion of small-scale proactive policing to a larger geographic area, or a municipality considering expansion of proactive policing through the introduction of particular new tactics. Staggered rollout of proactive enforcement would have the most important advantage of an RCT (simultaneous comparison between treated and nontreated neighborhoods) without the most daunting disadvantage (the forced assignment, by researchers, of some neighborhoods to treatment and others simultaneously to control). Moreover, given that new policies are often rolled out on a staggered basis for reasons having nothing to do with research design, policymakers are more likely to be willing to cooperate with researchers on a staggered rollout than to allow for a conventional RCT.

Such a research design, if implemented, lends itself most naturally to a difference-in-differences (DID) analysis. A DID analysis is appropriate

> "in problems with multiple subpopulations – some subject to a policy intervention or treatment and others not – and outcomes that are measured in each group before and after the policy intervention. To account for changes over time unrelated to the intervention, the change experienced by the group subject to the intervention (referred to as the treatment group) is adjusted by the change experienced by the group not subject to treatment (the control group).

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press

> The underlying assumption is that the time trend in the control group is an adequate proxy for the time trend that would have occurred in the treatment group in the absence of the policy intervention" (Athey & Imbens, 2006).

Probably the best-known example of a DID analysis is an evaluation of the effect of a minimum wage increase on employment levels when the minimum wage increased in New Jersey but not in neighboring Pennsylvania (Card & Krueger, 1994). That study leveraged an increase in the minimum wage in New Jersey in 1992, and compared fast-food restaurant employment levels before and after the minimum wage increase in New Jersey and in Pennsylvania. In the crime context, the best-known DID study is that of Di Tella and Schargrodsky (2004), where the authors were able to compare an increase in police patrols in specific areas of Buenos Aires with other areas that had simultaneously not received the same increase, finding that certain types of crime decreased in more heavily policed neighborhoods. In the proactive policing context, DID analysis can allow scholars to compare crime rates – and perhaps also other variables of interest, such as public attitudes toward the police – in neighborhoods that were subject to proactive policing with otherwise similarly situated neighborhoods that were not.

A quasi-experimental design could also allow researchers to discern a policy's effects by using slightly different proactive enforcement approaches for different neighborhoods in the city. For example, in two similar neighborhoods in different parts of the city, the police could assign one to a "high-staff" proactive patrol and one to a "low-staff" proactive patrol, with a greater density of proactive enforcement officers assigned to the former neighborhood compared with the latter. Alternatively, police could focus on one type of proactive enforcement in one neighborhood and another type in a different neighborhood. Doing so could enable one to discern the differential effects of, for example, increased foot patrols compared with increased traffic enforcement. Even though assignment would not necessarily be random, this would allow for a dose–response comparison between multiple neighborhoods (Farrelly, Davis, Haviland, Messeri, & Healton, 2005); that is, if the average effect of proactive patrol is indeed crime reduction, it should be that the higher-staff patrol has a greater crime reduction effect than a lower-staff patrol. Moreover, if particular strategies are more effective at crime reduction than others, a comparison between them would similarly allow researchers to observe those differences.

The most important threat to the validity of these quasi-experimental designs for evaluation of the effects of proactive policing is the risk of spillover effects. Spillover effects occur when the experimental or quasi-experimental treatment affects the outcomes of control units as well as treatment units. In this case, if proactive enforcement is by design confined to certain areas or neighborhoods

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press

of the city, but it affects crime outcomes in other areas or neighborhoods of the city as well, the control unit outcomes are being affected by the treatment, and a spillover is taking place. In colloquial terms, spillovers are a problem because they muddy the difference between treatment and control units so that a comparison between them is no longer a clean one. In technical terms, they are a problem for causal inference because they represent violations of the stable unit treatment value assumption (SUTVA), which requires that, for causal statements to be made, "the potential outcomes for any unit do not vary with the treatments assigned to other units" (Imbens & Rubin, 2015, p. 10).[2]

For proactive policing, the risk of the police's efforts spilling over into neighborhoods adjacent to those assigned to the new patrol can be substantial. Most obviously, patrol officers are usually expected to respond to any calls that take place nearby, so an increased police presence in one neighborhood may affect the level of policing in nearby neighborhoods. In addition, at the level of police resource allocation, allocation of more or better officers to some neighborhoods may have spillover effects on other neighborhoods, depending on how the police department goes about staffing. Moreover, at the level of criminal activity, there are perhaps some criminals who would, upon witnessing increasing police presence in some neighborhoods, shift their criminal activities to neighborhoods with a lower police presence. There are thus several respects in which a quasi-experimental design poses risks of spillover effects.

There are at least two ways to minimize spillover effects, although it is impossible to eliminate them entirely. First, if researchers are involved in the initial rollout of proactive enforcement, they can (1) be clear that officers should remain in their assigned patrol areas and (2) choose, if faced with choosing between beginning proactive enforcement in otherwise similar neighborhoods, to implement proactive enforcement first in high-crime neighborhoods that border low-crime neighborhoods (the logic being that police are not likely to be looking for illegal guns or drugs in a low-crime neighborhood, so proactive enforcement in a nearby high-crime neighborhood is unlikely to affect police activity there). If researchers are not involved in the initial design, they can use arrest and confiscation data to try to estimate the extent to which proactive enforcement has spread beyond the assigned neighborhoods, and accordingly deflate their estimated treatment effect (Chen, Mu, & Ravallion, 2009), and they can observe the extent to which police were likely to have been using proactive enforcement strategies in nearby neighborhoods based on historical crime rates in those neighborhoods.

2 This is only the first part of the SUTVA; the second part requires that "for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes."

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press

Ultimately, a quasi-experimental approach is very similar to an RCT. Unless researchers are fortunate enough to discover that a staggered rollout took place before they became involved in an evaluation, or that proactive enforcement was taking place in some high-crime areas but not in others, a quasi-experiment requires prospective planning around where and how to assign officers to proactive enforcement, very similar to what would take place in the RCT context. Police department requests for or researcher initiatives to conduct fully nonexperimental approaches are likely to be more common, and so I discuss evaluations of those sorts in the next section.

3 Nonexperimental retrospective approaches

In a nonexperimental retrospective research design, evaluators are faced with simply assessing the change in crime rates before and after the introduction of a proactive enforcement policy, and trying as best as possible to take account of all of the factors that could bias the results in any direction. In practice, this means using regression of the outcome variable or variables of interest (such as measures of the crime rate) on the treatment indicator (treated or untreated by proactive enforcement) and all covariates of interest that could potentially confound the analysis. This sort of analysis is plagued by a number of threats to validity. Most generally, the researcher must be confident that all possible confounders have been taken into account and ignorability thus achieved, even though "[i]n general, one can never prove that the treatment assignment process in an observational study is ignorable" (Gelman & Hill, 2006, p. 184). Beyond this general threat to validity in all observational regression analysis, there are four additional threats to validity for the specific case of assessing crime rates before and after the introduction of proactive enforcement. I discuss each in turn below.

3.1 Simultaneous policy changes

If increases in proactive enforcement occurred simultaneously with other policing policy changes, it would be extremely difficult to know whether observed changes in the crime rate were due to proactive enforcement or to these other changes. Many cities considering proactive enforcement might also, for the same reasons that they are considering the introduction of proactive enforcement, be considering the introduction of other anti-crime measures, such as the introduction of new technologies or the hiring of additional staff. To obviate the threat that simultaneous policy changes would pose to an evaluation of the effectiveness of proactive enforcement, researchers can, first and foremost, explain to police departments that the introduction of a raft of reforms all at once will make it nearly impossible to evaluate

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press

independent effects of the different reforms. At the very least, departments must be advised to keep close track of the new reforms – specifically, where they have been implemented and how aggressively they have been implemented – so as to at least be able to estimate different effects of specific reforms. Alternatively, researchers could alter the question of interest to reflect the joint effect of all reforms, thereby defining away this particular problem.

3.2 Overall temporal and seasonal trends in the city or regional crime rate

Researchers need to have accurate estimates of ex ante crime rate trends in the city and region so as to know how much of an observed ex post change in crime rates might be due to these broader trends. There are well-documented seasonal fluctuations in the crime rate (Andresen & Malleson, 2013, p. 43), for example, and there are also significant long-term crime trends which transcend geography and are not well understood (Blumstein & Wallman, 2006). Without thorough knowledge of these trends, it would be impossible to have confidence in any estimate of the effect of a policy change.

3.3 Hawthorne effects

Hawthorne effects occur when a population knows that it is the subject of a research study, and thus alters its behavior – in particular, alters its behavior in a way that is prosocial vis-à-vis its perception of the goals of the study (McCarney et al., 2007, p. 1). In this case, police officers who know that the proactive enforcement policy is going to be the subject of researchers' scrutiny could take a number of steps to alter measured outcomes in ways that they might not have had they not known about the research, including downgrading crimes from felonies to misdemeanors, electing not to make arrests when not absolutely necessary, making more arrests than may be strictly necessary, or being especially vigilant about the possibility of citizen complaints (Greenberg, 2014). A fully retrospective observational study might avoid this problem if police did not know that a research evaluation would take place at the time reforms were introduced.

3.4 Consequential under- or over-reporting of (some types of) crime

In some cases, aggressive or violent police actions have been shown to decrease the rate of citizen reporting for some types of crimes. The only case in which this has been definitively documented, however, was an extreme case of police violence

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press

(Desmond, Papachristos & Kirk, 2016). A corollary issue is that it is not uncommon for individuals who are chronically in contact with law enforcement to use law enforcement as an instrument to act out their personal agendas, such as by calling the police on friends or family members with whom they have disputes (Venkatesh, 2009). In sum, proactive enforcement might lead to under-reporting of crimes when individuals are concerned that the police might take aggressive action against friends, neighbors, or relatives. Conversely, proactive enforcement might lead to over-reporting of crimes where individuals see an increased police presence in their neighborhoods. If individuals perceive that officers are eager to make arrests, individuals might be more likely to use the police as instruments in interpersonal disputes. These possible effects on citizen behavior would complicate efforts to measure the effects of proactive enforcement alone, rather than the effects of proactive enforcement coupled with corresponding changes in citizen behavior.

There are strategies, however, for researchers to account for possible changes in citizen reporting behavior brought about by the implementation of proactive policing practices. To measure possible under-reporting, researchers can take advantage of 911/call-for-service data for at least one year before the policy change in order to estimate any changes in reporting that may have resulted from the introduction of proactive enforcement. Qualitative research can also be used to interview residents of high-crime neighborhoods and ask whether they have or have seen others change what they might report to the police due to the new policy (Brunson, 2007). To measure possible over-reporting, researchers could take advantage of a 911 dispatcher's coding of various incidents. Dispatchers code incidents by assigning a category based on the type of incident, in light of what is told to the dispatcher on the 911 call. This coding could allow researchers to keep track of incidents that are first coded as family fights or domestic disputes, and which are resolved with administrative codes for preserving the peace, civil matter, or arrests for previously outstanding warrants. These types of commencement/resolution combinations should not increase as a result of proactive enforcement, so any increases that do take place are evidence that citizen behavior has changed in light of proactive enforcement.

4 Relevant nonexperimental data

Fortunately, neither the quasi-experimental nor the nonexperimental retrospective research designs research requires data that are not regularly collected by police agencies, although some additional data would be useful. I have mentioned various data sources in the above discussion, but I list them in more detail below. With

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press

one exception (noted below), these data require a strong partnership with a police department that is willing to share detailed current and historical crime data.

- *Crime data*. Data on stops, arrests, confiscations, and fines, including detailed location information, are absolutely crucial to either the quasi-experimental or the nonexperimental retrospective research designs. Additionally, crime report data that reflect what types of incidents officers responded to and how those incidents were resolved reflect the "dosage" of proactive enforcement (how much proactive enforcement behavior is in fact taking place) and whether proactive enforcement has shifted the types of calls or incidents to which the police respond.
- *Officer assignment data*. Data on where officers were on patrol and when, including which incidents they responded to, are important for assessing whether any complaints about proactive enforcement are coming from the conduct of a small number of officers who could potentially be reassigned (Terrill & Ingram, 2016).
- *Historical and regional crime data*. These data are necessary for the nonexperimental retrospective evaluation, because researchers need to be able to assess whether observed changes are due to historical or regional trends rather than policy changes. The most appropriate data source for this purpose would be the FBI's Uniform Crime Reporting program.
- *911/call-for-service data*. These data are needed to estimate whether over- or under-reporting is taking place relative to before the introduction of the proactive enforcement policy.
- *Interview data*. These data are not strictly necessary, but interviews with the residents of affected neighborhoods would be useful for an understanding of how the proactive enforcement policy is being perceived, and whether it might be leading to changes in crime or reporting behavior.

5 Conclusion

Researchers seeking to evaluate a proactive enforcement policy without an RCT have two options: a quasi-experimental evaluation and a nonexperimental retrospective evaluation. Ultimately, a quasi-experimental evaluation may be nearly as difficult and costly as an RCT, unless researchers find that a natural experiment took place before they arrived. A nonexperimental retrospective evaluation will almost always be easier, but will require the use of a wider range of data to eliminate the many threats to validity faced by such an analysis. The key threats to validity for an observational study of proactive policing are simultaneous policy change, regional

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press

and historical crime trends that may wash out policy effects in initial before-and-after summary statistics, Hawthorne effects, and under- or over-reporting of certain types of crimes as a result of the new policy. Researchers may estimate the size of the cross-cutting effects generated by these outside factors by using crime data, police patrol assignment data, historical and regional crime data, 911/call-for-service data, and interview data. Moreover, even a study design that overcomes these threats will need to account for the distributional effects of proactive policing, given that its costs and benefits fall on different individuals.

Proactive enforcement, like all police practices, should be subject to rigorous analysis to determine its benefits and costs. While the full range of benefits and costs may be difficult to measure, the effect of proactive enforcement on crime rates is, happily, more readily measured than other questions of interest in police policy. The ability to clearly identify both the intervention (proactive enforcement in particular areas) and one major effect (crime rates in those areas) makes benefit-cost analysis a potentially viable possibility. Through either a quasi-experimental or a purely randomized design, scholars can estimate proactive enforcement's effects on crime rates, and incorporate that estimate into a broader analysis of whether the benefits of proactive enforcement outweigh the costs, given the nature of those benefits and costs, and the populations on whom they fall.

Acknowledgments: I am grateful to the editors and referees of the Journal of Benefit-Cost Analysis for helpful comments on this paper, and to Barry Friedman and the Policing Project at NYU Law School for hosting the conference for which this paper was initially written.

References

Adler, Matthew D. (2016). Benefit-Cost Analysis and Distributional Weights: An Overview. *Review of Environmental Economics and Policy*, *10*(2), 264.

Andresen, Martin A. & Malleson, Nicolas (2013). Crime Seasonality and its Variations Across Space. *Applied Geography*, *43*(Mar.), 25–35.

Athey, Susan & Imbens, Guido W. (2006). Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica*, *74*(2), 431–497.

Blumstein, Alfred & Wallman, Joel (Eds.) (2006). In *The Crime Drop in America*. New York: Cambridge University Press.

Boruch, Robert. (2015). Street Walking: Randomized Controlled Trials in Criminology, Education, and Elsewhere. *Journal of Experimental Criminology*, *11*(4), 485–499.

Brunson, Rod K. (2007). Police Don't Like Black People: African-American Young Men's Accumulated Police Experiences. *Criminology & Public Policy*, *6*(1), 71–101.

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press

Card, David & Krueger, Alan B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, *84*(4), 772–793.

Chen, Shaohua, Mu, Ren & Ravallion, Martin (2009). Are there Lasting Impacts of Aid to Poor Areas? *Journal of Public Economics*, *93*(3), 512–528.

Cohen, Mark A. (2017). The Social Cost of a Racially Targeted Police Encounter. *Journal of Benefit-Cost Analysis*, *8*(3), 369–384.

Desmond, Matthew, Papachristos, Andrew V. & Kirk, David S. (2016). Police Violence and Citizen Crime Reporting in the Black Community. *American Sociological Review*, *81*(5), 857–876.

Di Tella, Rafael & Schargrodsky, Ernesto (2004). Do Police Reduce Crime? Estimates Using the Allocation of Police Forces after a Terrorist Attack. *The American Economic Review*, *94*(1), 115–133.

Farrelly, Matthew C., Davis, Kevin C., Haviland, M. Lyndon, Messeri, Peter & Healton, Cheryl G. (2005). Evidence of a Dose–Response Relationship Between Truth Anti-smoking Ads and Youth Smoking Prevalence. *American Journal of Public Health*, *95*(3), 425–431.

Gelman, Andrew, Fagan, Jeffrey & Kiss, Alex (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*(479), 813–823.

Gelman, Andrew & Hill, Jennifer (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Greenberg, David F. (2014). Studying New York City's Crime Decline: Methodological Issues. *Justice Quarterly*, *31*(1), 154–188.

Imbens, Guido W. & Rubin, Donald B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. New York: Cambridge University Press.

Jones, David S. & Podolsky, Scott H. (2015). The History and Fate of the Gold Standard. *Lancet*, *385*(Apr.), 1502–1503.

McCarney, Rob, Warner, James, Iliffe, Steve, van Haselen, Robbert, Griffin, Mark & Fisher, Peter (2007). The Hawthorne Effect: A Randomised, Controlled Trial. *BMC Medical Research Methodology*, *7*(1).

New York Civil Liberties Union. *Stop-and-Frisk Data*. Available at: https://www.nyclu.org/en/stop-and-frisk-data (Accessed September 15, 2017).

Nix, Justin & Rojek, Jeff (2014). Proactive Policing. In *Criminology – Oxford Bibliographies*. Available at: http://www.oxfordbibliographies.com.ezp-prod1.hul.harvard.edu/view/document/obo-9780195396607/obo-9780195396607-0183.xml (Accessed September 15, 2017).

Terrill, William & Ingram, Jason R. (2016). Citizen Complaints Against the Police: An Eight City Examination. *Police Quarterly*, *19*(2), 150–179.

Venkatesh, Sudhir Alladi (2009). *American Project: The Rise and Fall of a Modern Ghetto*. Cambridge: Harvard University Press.

Weisbach, David. A. (2015). Distributionally Weighted Cost–Benefit Analysis: Welfare Economics Meets Organizational Design. *Journal of Legal Analysis*, *7*(1), 151–182.

https://doi.org/10.1017/bca.2017.26 Published online by Cambridge University Press