Benefit-Cost Analysis in Policing Research: Assessing Crime-Control Benefits of Proactive Enforcement Practices

Amanda Geller[1]

The police are widely recognized as an important deterrent of crime (Durlauf and Nagin, 2011), with investments in policing broadly associated with lower crime rates (Evans and Owens, 2007). Much less is known, however, about *how* investments in policing may contribute to crime reductions (Skogan and Frydl, 2004), the relative merits and risks of specific activities in which officers engage, or the broader community implications of police activity, beyond crime control. Benefit-cost analyses of specific policing practices therefore stand to contribute substantially to our understanding of the circumstances in which police can provide the greatest net social benefits.

However, even the most obvious and widely recognized benefit of policing practices, the potential for crime reduction, is difficult to quantify. This essay explores strategies for measuring this benefit, with a focus on strategies and tactics identified as "proactive" policing practices (Kubrin, Messner, Deanne, McGeever, and Stucky, 2010), and their effects on crime rates. These may include investigative stops, citations, and arrests to detect and disrupt low-level disorder or other circumstances interpreted as indicia that crime is afoot (Kelling and Coles, 1996; Kubrin et al., 2010). I begin by assessing the challenges inherent in assessing the crime-control effects of any police practices, with particular attention to those associated with the measurement and modeling of proactive enforcement. I then lay out potential strategies for addressing these challenges.

I.   Challenges inherent in assessing crime-control effects of police practices

The challenges inherent in assessing the effects of police practices on crime are well documented (Messner and Baumer, 2014; Rosenfeld and Fornango, 2014), and this essay is far from the first to attempt to overcome these challenges. I broadly classify these challenges into two categories: those related to measurement, and those related to modeling.

A. Measurement Challenges

Researchers have long understood that policing practices are the byproduct of a series of interrelated decisions (Skogan and Frydl, 2004), which I will, for now, classify as deployment decisions (i.e., how many officers to hire and where to station them), and tactical decisions (i.e., what, specifically, to ask officers to do on a given shift). When assessing the effects of a particular police practice, including, but not limited to, "proactive" practices such as investigative stops, vehicle checkpoints, or misdemeanor arrests, or activities that combine a more diverse set of tactics, researchers face persistent challenges in isolating the crime control effects of a particular tactic from the effects of other public safety activities in which officers engage, or the deterrent or surveillance effect of officer presence in the absence of particular activities. Even when opportunities exist to

---

[1] Associate Professor of Sociology, New York University. Director of the M.A. Program in Applied Quantitative Research.

perform a controlled experiment, or circumstances create a natural experiment, measuring the changes in police practices is a challenge of its own, before attempting to model the effects of these changes on crime conditions.

In addition to the aforementioned challenges inherent in measuring the quantity of police activity, measuring the quality of policing is even harder. For example, whether a particular investigative stop has a productive effect on crime reduction depends on several dimensions of the policing. First, the productive effect depends on whether the stop took place, a quantity increment. Second, the productive effect depends on whether the stop was justified by reasonable suspicion, a factor that may make it more likely to detect illegal activity and prevent subsequent activity. Third, the productive effect depends on whether it was conducted in a fair and respectful manner. Manner may enhance perceptions of police legitimacy and subsequent cooperation, not only for the individual stopped, but also for others in the community who may have witnessed the stop or heard about the individual's experience. These dimensions of proactive policing are exceedingly difficult to measure, creating challenges for assessment of their effects.

B. Modeling Challenges

Causal modeling is generally challenged by the fact that a "treatment" state must be evaluated in comparison to a counterfactual condition – what the world would have looked like in the absence of the treatment. As already noted, both the quantity and quality of policing practices are quite difficult to measure, making counterfactual conditions a challenge to model. Assessing the particular effects of police practices on crime are further complicated by treatment endogeneity: it is reasonable for cities or precincts to change policing strategies and tactics in response to a crime spike; unfortunately any changes in crime patterns observed after these endogenous changes are likely to mix any true "treatment effects" with spurious mean reversion. Randomized experiments are expensive, and true natural experiments, in which police practices are changed by exogenous shocks, are rare. Further, when such experiments do occur, they are often the result of special circumstances that may be limited in their generalizability.

II. Approaches to assessing crime-control effects of police practices

In light of these challenges, there are several ways researchers might proceed. I begin by discussing ways we might improve our measures of police practices, and what data might be required to do so, in order to assess changes in policing and their crime-control effects. Finally, I discuss ways in which we might isolate or simulate exogenous changes in police practices in order to observe changes in crime conditions that may result.

A. Measurement Approaches

As noted above, challenges persist in distinguishing measurement of the quantity and quality of policing. Especially challenging is measuring the differences between the number of officers hired and number deployed, the specific tasks that officers may perform in pursuit of crime control and how frequently they are performed, and the manner in which these tasks are performed. The most effective measurements of policing quantity will distinguish the number of officers deployed on a given day, or even in a given shift or

geographic area, from the number and type of interactions that officers have with the public. Ideally these measures would also provide a distribution across officers working in a given precinct or beat, rather than aggregates. Due to differences in assignment, it is likely that some officers will have significantly more time on the street than others, creating more opportunities for interactions with the public. Further, officers with equal time on the street may engage in different activities to fight crime, including "proactive" practices like investigative stops and arrests, but also including broader community engagement and other activities. Crime conditions are likely to be influenced by the cumulative effects of all activities in which officers engage. Understanding the crime control benefits of proactive policing therefore requires measuring the totality of police presence and police activity, as the quantity of proactive practices, distinct from other police actions.

The concept of policing quality is harder to define, let alone measure. One way in which we might think about policing quality would be in terms of fidelity to the constitutional principles that govern particular practices. Are street stops justified by reasonable suspicion of illegal activity? Are frisks justified by legitimate concerns for officer safety? Are arrests justified by probable cause?

The most concrete way to assess these dynamic measures of police practices would be through the collection of data using body-worn cameras (BWCs), though the sheer volume of data collected by BWCs itself introduces challenges. The fidelity of police practices to constitutional principles, might also be measured through officer self-reports of police actions, or the collection of data on civilian complaints, preferably by precinct or district. These data, however, may be plagued by social desirability bias because officers face incentives to characterize their behavior in a positive light or selection bias because not all individuals experiencing an encounter they feel is unjustified or excessive will know the process or choose to file a formal complaint. Conversely, some incidents that are unsatisfactory to civilians may yet be constitutionally justified. Civilian complaint data would be strengthened by a prompt and thorough investigation process that helps to quantify the extent of unjustified officer behavior.

Policing quality can also be measured in terms of its efficacy in detecting and disrupting illegal activity. One common set of measures of policing efficacy is the "hit rate" at which stops or searches yield contraband or lead to arrest (Knowles, Persico, and Todd, 2001), or the rate at which arrests lead to convictions for criminal offenses. These measures can also be expanded to measure additional effects of stops or searches, particularly if researchers are able to combine policing data with, for example, network data that identifies whether a stop or arrest of one individual has the added effect of disrupting illegal activity among others in the social network of the individual stopped/arrested.

Each of these measures can potentially be quantified in statistical models to assess the extent to which proactive policing practices might affect crime rates.

B. Modeling Strategies

In addition to the immediate disruption of criminal activity, captured to varying degrees of quality in hit rate data, proactive policing practices may also reduce local crime by deterring illegal activity through increased surveillance and an increased probability of

being actively engaged by police. As already noted, these effects cannot be effectively measured by simply comparing crime rates in a given area before and after a change in police practices, or comparing crime rates between more and less heavily policed area; police deployment is typically endogenous to crime and other conditions, complicating the disentanglement of true "policing effects" from spurious confounding factors and mean reversion.

In the absence of randomized and natural experiments and in light of their potential implementation challenges and potential limits to their external validity, researchers have several alternative approaches at their disposal. They can estimate panel models that explicitly consider the simultaneous determination of crime and police practices, or use non-parametric and semi-parametric techniques that leverage naturally occurring discontinuities, at which police practices are likely to change, but crime conditions and potential confounders such as economic processes and demographic composition are likely to be consistent in the absence of policing differences.

1. Panel Models

Regression approaches measure the association between a predictor such as police practices and an outcome such as crime conditions, net of other observable factors. In order for regression findings to be interpreted causally, however, observed associations must be non-spurious. Several modeling approaches could potentially be used to rule out potential spurious associations, including (a) fixed effects models that control for secular trends and time-invariant area characteristics, (b) lag structures that model within-area changes, (c) cross-lagged models that reflect the extent to which future police practices are determined by past and present crime conditions, and may in turn affect crime conditions going forward.

Within any of these modeling frameworks, attention must be paid to the appropriate units of measurement for observing changes, both in terms of geography (precinct level? tracts? smaller "hot spots"?) and time (years? months? weeks?). Short term changes, or changes in small geographical units, may be obscured at higher levels of aggregation. For smaller levels of geographic aggregation, models may also benefit from the inclusion of spatial lags, to model both whether policing effects spill over into neighboring areas, and whether crime declines in one area are displaced to another nearby. Parametric models must also consider potential nonlinearities in the relationships between police practices and crime conditions. A practice may be highly effective when carefully targeted, but when implemented more broadly – particularly if less discretion is exercised – may have diminishing returns.

The "correct" model for the association between proactive policing activities and crime conditions is inherently uncertain. Cohen-Cole, Durlauf, Fagan, and Nagin (2009) illustrate methods for addressing model uncertainty; they deal specifically with models of capital punishment and its deterrent effect, but similar approaches may be used to address questions related to proactive policing and crime control. Cohen-Cole and colleagues (2009) use model averaging approaches (Leamer, 1978), which treat the "true" model of a relationship as unknown, but able to be estimated based on a weighted average of "candidate models". Estimates from each candidate model are weighted based on the

4

probabilities that each model is the true representation of the relationship of interest. In so doing, researchers systematically evaluate both the findings from individual models, and the decisions inherent in model specification (Cohen- Cole et al., 2009), to avoid over-reliance on any single model.

Another set of approaches, rather than trying to model the interdependencies between police practices and crime, exploits discontinuities across which changes might be plausibly thought of as exogenous.

2. Spatial Discontinuities

One class of discontinuities that may be exploited is the spatial discontinuity between police precincts or districts. Prior analyses (Geller and Fagan, 2011) suggest that administrative (i.e., precinct) boundaries matter for police practices, holding social processes constant across locations. If we can find sub-precinct areas on either side of precinct boundaries that are socially similar, then tracking differences in police deployment and tactics in these areas over time will help us isolate differences in crime patterns associated with changes in police practices. MacDonald, Fagan, and Geller (2016) use difference-in-differences methods exploit differences between Impact Zones, high-crime regions designated for increased attention, and their surrounding areas, to find increases in both stops and arrests, and an association between this enforcement increase and a small reduction in crime. Similar methods may be employed to assess differences in police practices across other social boundaries, such as public housing projects (See, e.g., Fagan, Davies, and Carliss, 2012), and, potentially, any associated crime control benefits.

Notably, this approach will work more effectively in some precincts and neighborhoods than others, as some administrative boundaries also represent sharp social boundaries. Taking New York City as an example, the E 96th St. boundary between the 23rd (East Harlem) and the 19th (the Upper East Side) precincts provides not only an administrative boundary but also a relatively sharp demographic and socioeconomic boundary. Given such contrasts, differences in crime rates across the border may conflate policing effects with other social differences. On the other hand, boundaries such as the E 115th St. boundary between the 23rd and 25th Precincts – both in East Harlem (broadly defined) – represents a blurrier boundary between social circumstances, suggesting that crime differences might in fact reflect effects of local differences in policing practices.

A variation on this geographic model would be to construct synthetic control neighborhoods that are socially similar though not geographically contiguous. For example, Abadie and Gardeazabal (2003) assess the economic costs of conflict in the Basque country by constructing a "'Synthetic' Basque Country Without Terrorism" (p. 116) based on a composite of non-Basque, conflict-free regions, weighted to resemble the Basque region prior to the conflict on a variety of economic growth predictors. They then track the composite region into the conflict period and measure the extent to which the true Basque region diverges from its control. Similar analyses may be done on the neighborhood level within a city undergoing localized changes to policing practices. The synthetic control model may present challenges in the assessment of policing, particularly if the areas comprising the "control units" have heterogeneous approaches to policing that are difficult to aggregate across units. However, if assessing the effects of a particular policing

"treatment", then the absence of that treatment can potentially be aggregated across units in the control area.

3. Temporal Discontinuities

In addition to geographic discontinuities there may be time discontinuities - differences expected in, say, the last week of the month or quarter that may be changed in the start of a new month/quarter. Whether, for example, police officers are more proactive at investigatory stops at the end of a month, with a sharp decline at the start of a new month, is an empirical question. Such changes, if they exist, may be driven by factors endogenous to departmental culture, such as whether officers face monthly quotas or productivity goals, or otherwise feel pressured to change their behavior as the month ends. However, even if these patterns are driven by aspects of departmental climate, they are likely exogenous to the crime conditions they will potentially affect. Therefore, to the extent that policing practices change systematically across a given time horizon, any corresponding changes in crime patterns are likely to represent a causal effect of policing.

Each of the aforementioned methods exploits spatial or temporal discontinuities in police practices to identify differences in police practices that are plausibly exogenous to criminal activity and related social conditions. Differences across these discontinuities serve as a natural experiment, permitting plausible causal estimates of policing effects on crime, under the conditions measured. These estimates are unlikely to be conflated by related social factors, and therefore have relatively high internal validity. However, challenges remain in the assessment of policing effects where discontinuities in practices are less clearly defined. Furthermore, like most natural experiments, the external validity of even an ideal discontinuity-based estimate may be limited. Conclusions based on such estimates may not generalize to contexts where no sharp discontinuities exist.

III. Conclusion

Although the crime reduction associated with policing is widely recognized, the empirical strategies for assessing the crime control effects of specific police practices are inherently limited by challenges in measurement, modeling, and contextualizing any empirical findings that result. It is therefore particularly important to assess crime control associations using as many of these diverse methods as possible, and determine the robustness or sensitivity of substantive conclusions to the use of various empirical strategies.

References

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review, 93*(1): 113-132.

Cohen-Cole, Ethan, Steven N. Durlauf, Jeffrey Fagan, and Daniel Nagin. 2009. Model Uncertainty and the Deterrent Effect of Capital Punishment. *American Law and Economics Review, 11*(2): 335-369.

Durlauf, Steven N., and Daniel Nagin. 2011. "Imprisonment and Crime: Can Both Be

Reduced?" *Criminology and Public Policy, 10*(1): 13-54.

Evans, William N., and Emily Owens. 2007. "COPS and Crime." *Journal of Public Economics, 91*: 181-201. doi:10.1016/j.jpubeco.2006.05.014

Fagan, Jeffrey, Garth Davies, and Adam Carliss. 2012. "Race and Selective Enforcement in Public Housing." *Journal of Empirical Legal Studies, 9*(4): 697-728.

Geller, Amanda, and Jeffrey Fagan. 2011. *Spillover and the Demography of Urban Policing*. Paper presented at the Population Association of America, Washington, DC.

Kelling, George L. and Catherine M. Coles. 1996. *Fixing Broken Windows : Restoring Order and Reducing Crime in Our Communities*. New York, NY: Free Press.

Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy, 109*(1): 203-229.

Kubrin, Charis E., Steven F. Messner, Glenn Deanne, Kelly McGeever, and Thomas D. Stucky. 2010. "Proactive Policing and Robbery Rates Across U.S. Cities." *Criminology, 48*(1): 57-97.

Leamer, Edward E. 1978. *Specification Searches.* New York, NY: John Wiley. MacDonald, John, Jeffrey Fagan, and Amanda Geller. (2016). "The Effects of Local Police Surges on Crime and Arrests in New York City." *Plos One, 11*(6): e0157723.

Messner, Steven F., Eric P. Baumer. 2014. "Stop, Question, and Assess: Comments on Rosenfeld and Fornango." *Justice Quarterly, 31*(1): 159- 164. doi:http://dx.doi.org/10.1080/07418825.2012.712153

Rosenfeld, Richard, and Robert Fornango. 2014. "The Impact of Police Stops on Precinct Robbery and Burglary Rates in New York City, 2003-2010." *Justice Quarterly, 31*(1): 96-122. doi:http://dx.doi.org/10.1080/07418825.2012.712152

Skogan, Wesley G., and Kathryn Frydl (Eds.). 2004. *Fairness and Effectiveness in Policing: The Evidence*. Washington, DC: National Academies Press.