Evaluating the Impact of a Proactive Policing Policy on Crime in the Absence of a Randomized Controlled Trial

Emily Owens[1]

This article will highlight general best practices and common issues more specific to police policies, in evaluating police interventions in the absence of a randomized controlled trial. In order to understand how to properly implement a non-experimental policy evaluation in a way that produces credible causal effects, it is helpful to understand exactly why a randomized controlled trial "works."

Why does an experiment produce causal estimates?

Consider a situation where all the N people in an area engage in some non-negative level of criminal activity. Many factors affect each individual person's decision about how much crime to commit, and we will define the set of factors that determine person i's offending level as Ωi. If we wanted to conduct a randomized experimental trial of a particular proactive policy, P, we would, at random, expose some subset k of the population to this policy. Now, we can rewrite the amount of crime each person will engage in as $Ci=\beta Pi + \Omega i$. Here, Pi=1 for the k people exposed to the new policy, and for the N-k other people not exposed to the policy, Pi=0. The term β represents the amount of influence that the policy has on their criminal behavior.

The random exposure to the policy P has created a wedge in the crime rates across exposed and unexposed individuals that can be used to identify the extent to which the policy Pi influences behavior in the following way: On average, the crime rate of people exposed to the policy will be $Ck = \beta + \overline{\Omega_k}$, where $\overline{\Omega_k}$ is the average general criminal propensity of the k people if the experiment had not taken place. For the N-k others, their criminal behavior, on average, will be $CN\text{-}k = \overline{\Omega_{N-k}}$, as none of them are exposed to the policy (Pi=0 for this group). The difference in criminal behavior in the two groups is simply Ck-CN-k, or $\beta + \overline{\Omega_k} - \overline{\Omega_{N-k}}$. This three term equation can be interpreted as the causal policy effect β plus the difference in average criminal propensities across the two groups of people, $\overline{\Omega_k} - \overline{\Omega_{N-k}}$. Critically, random assignment of P means that, in expectation, there should be no difference in the average factors influencing the exposed and unexposed groups, meaning that we would expect $\overline{\Omega_k} - \overline{\Omega_{N-k}}$ would be equal to zero.

If random assignment "worked," meaning that it is actually the case that $\overline{\Omega_k} - \overline{\Omega_{N-k}} = 0$, then the difference in crime rates across groups is simply β. Of course, in practice you may end up with a situation in which the exposed and unexposed groups, randomly, have different underlying propensities to engage in crime on average, and not every component of Ωi can be measured or quantified. Researchers generally demonstrate that in actuality $\overline{\Omega_k} - \overline{\Omega_{N-k}}$ is zero by showing that some quantifiable factor, like income, race, or education, is the same on average across the exposed and unexposed groups.

[1] University of California-Irvine, Department of Criminology, Law and Society, and Economics

How does a non-experimental setting change this?

Not having an experiment makes causal policy evaluation difficult because, as a matter of course, it will never be the case that $\overline{\Omega_k}$- $\overline{\Omega_{N-k}}$=0. In words, it is almost always the case that the k people exposed to a particular policy are not the same, in fundamental way, as the N-k people who are not. When this is true, and the people who are exposed to a crime control policy have a different propensity to engage in crime as those that do not, the average difference in the underlying crime groups, Ck – C N-k, will be equal to β + η, where η represents the non-zero value of that difference, $\overline{\Omega_k}$- $\overline{\Omega_{N-k}}$. As a matter of course, η could be positive, negative, very large, or very small- it simply depends on the way in which the policy was implemented. When there is uncertainty about the sign and magnitude of η, it means that non-experimental comparisons of mean crime rates across exposed and unexposed areas should not be used to evaluate the benefits or costs of a policy. A classic example of this is shown in LaLonde (1986), which compared non-experimental evaluations of a job training program to a randomized trial. An experimental evaluation suggested the program increased earnings by $886, but non-experimental estimates of the same program ranged from $1,700 increase to a $2,000 reduction (LaLonde, 1986).

In practice, some sense of the value of η can be ascertained by asking the following question: what, exactly, determined whether or not someone (or someplace) was exposed to the policy? In other words, why wasn't everyone exposed to P? This requires institutional knowledge on the part of the evaluator. Unfortunately, the less an analyst knows about how a policy was actually implemented, the more likely it tends to appear to that analyst that η is small, i.e. that exposure was close to random. This unfortunate situation is due to one of the best practice tests of randomization: comparing mean values of what we can observe across exposed and unexposed groups. If analysts simply do not know what actually determined assignment, they are more likely to test observable values that were, in fact, not used by the practitioner to determine exposure. The mean values of these irrelevant variables may, in fact, be zero, but this simply means that η is made up of variables that were untested by the analyst.

To give a tractable example of this, in the case of Ridgemont, exposure to the proactive automobile stops was determined by whether or not individuals drove through areas designated by Chief Johnsons as High Crime. By construction, we would expect more crimes in these areas than others, meaning that η >0. This does not mean, of course, that there are systematic differences in all measurable features of these neighborhoods. To give an extreme example, the analyst with access to data on pet ownership could show that you cannot reject the null hypothesis of equal rates of dog ownership in the treated and untreated areas. Of course, this would not be an insightful comparison which does not confirm that exposure to the proactive stops were an experiment—it simply means that dog ownership was not used to assign policy exposure. A researcher without information about why P varies across individuals is likely to end up with estimates of causal effects that are as credible as predictions of police activity based on preferences for dogs.

How do researchers solve this problem?

Detailed institutional knowledge of how assignment was made can allow for credible identification of η, and thus the causal impact of the policy on crime rates. Typically, conversations with the practitioners who implemented the policy are the best way to gain insight on their decision rules. In the language of causal policy analysis, the goal of these conversations is to identify a credible counterfactual for the exposed areas; how might the analyst identify a set of people who were not exposed to the policy, but other than that are similar (and on average identical) to a set of people who were exposed. Mathematically, this specifically means that, for those two groups $\overline{\Omega_k}$- $\overline{\Omega_{N-k}} = 0$? All mathematical strategies used in non-experimental causal policy analysis (regression adjustment, instrumental variables, matching, difference-in-differences analysis, regression discontinuity designs, etc.) can be characterized in this way: an attempt to identify a set of exposed and unexposed people for whom the average values of propensities to engage in crime, in the absence of the policy, are the same across groups. Whether or not a particular mathematical strategy creates a situation where this statement is true depends on whether or not the actual assignment decision is matched by the assumptions underlying the mathematical model. This means that no one econometric approach will provide causal policy estimates in all situations – a "strong" research design in one context can produce wildly inaccurate estimates in another, simply based on how the practitioner chose to implement a policy. If causal policy estimates are the goal, the correct modeling strategy must be determined on a case-by-case basis.

What are some helpful steps a research can use to find the right model to separate β and η?

When analysts are asked to evaluate the casual impact of a public policy after the fact (meaning random assignment is not possible), the first step must always be to learn as much about the way a policy was implemented. After fully understanding the policy, then the analyst must try to identify a dimension along which she can identify a sample of treated and untreated areas where policy exposure is, in practice, unrelated to crime rates.

In the case of Ridgemont, people driving through High Crime areas were subject to more frequent traffic stops. A first question would be: how were these areas determined to be High Crime? In some cases, it might be true that there are many areas, each of which were ranked based on the number of times police are sent to an area by dispatch. Any area with more than, say, 100 calls per week is designated as high crime. If this were the case, comparing crime in areas with slightly fewer than 100 calls per week to crime in areas with only slightly more than 100 calls per week may generate plausible estimates of β, since $\overline{\Omega_{101\ calls}}$- $\overline{\Omega_{99\ calls}}$ will probably be very close to zero

Alternately, it might be the case that, due to budgetary constraints or union rules, only High Crime Areas in certain jurisdictions were involved in the proactive stops. If this were true, analysis could compare crime rates in High Crime areas that, but for the non-crime related rule, would have been exposed to the policy- meaning that $\overline{\Omega_k}$- $\overline{\Omega_{Non-Crime\ Rule}}$ should be close to zero. Comparing trends in crime across High Crime and Non-High Crime areas may also be helpful, if it is the case that month-to-month fluctuations in crime are the same

(in level or percentage terms) in exposed and unexposed areas. In this case, using a fixed effects, or difference-in-differences specification may yield credible estimates of β, as $\Delta \overline{\Omega_k}$-$\Delta \overline{\Omega_{N-k}}$ would be equal to zero.

Issues Specific to Proactive Policies and Crime

The previously discussed strategies for generating credibly causal estimates of a proactive policing policy are generalizable to all policy evaluation contexts. We will now discuss issues that are more specific to crime control policies, and proactive policing in particular.

What is the proper unit of analysis?

A frequently overlooked assumption in policy evaluations is what, exactly, constitutes an exposed and unexposed area. In the case of Ridgemont, the specific policy in place was one of aggressive traffic stops in High Crime areas. Where, exactly, would we expect to see crime reductions as a result of this policy? The treated group of people, in this case, are drivers who are stopped by the police. We would expect, in theory, multiple impacts on crime, which may be revealed over different periods of time.

First, the people who would have used, sold, or consumed the drugs or weapons, are now unable to, if they were stopped. This change in resources may immediately reduce their ability to engage in criminal acts in the short run. Second, awareness of the increased police surveillance may lead people who were stopped to decide to stop carrying drugs or weapons all together, creating a longer run reduction in crime. Third, the increased rate of police activity in High Crime areas may lead drivers carrying drugs or weapons to avoid driving in these areas. For some fraction of people who carried drugs or weapons in their cars, this will simply mean they drive through other places. For others, this change in transportation pattern will make carrying drugs or weapons in their vehicle too costly a behavior to engage in. All of these are different "effects" of the policy- analysts should clearly justify the time dimension they are most interested in.

It also may not be the case that a place-based policing strategy only affects crime in the place where the policy is implemented. By changing the physical environment, place-based policies, particularly when they are targeted at very small areas, may lead people who previously would have offended in that area to commit the same crimes elsewhere. These spillover effects may be difficult to detect, particularly when the mobile criminals who change the location, but not the frequency, of their actions relocate to different areas. A statistically significant reduction of five crimes in one block may lead to an increase of one crime in five other blocks, and depending on the statistical power of the analysis, this offsetting increase in crime may not be large enough to statistically differentiate from zero.

These types of diffuse spillovers may mean that, even when policies are implemented at a geographically diffuse level, it may be relevant to think about the impact on a broader geographic scale. This is particularly true when the analyst is evaluating policies that have been fully implemented and the CBA is indented to provide guidance to other jurisdictions, rather than expanding the policy to other neighborhoods within a jurisdiction. In this case,

policy makers may want to know what the impact of geographically concentrated traffic enforcement is on crime rates in their cities as a whole, rather than the spatial distribution of crime.

Should you conduct a place-based or person-based analysis?

Geographic modelling and spatial dependence seem particularly relevant for policies that target drivers. People who are "treated" by the police are, by definition, in transit. It seems likely, in these cases, that there is little relationship between where the drivers are stopped and where they would have engaged in further crime, had they not been stopped in that particular place. To put it another way, where, exactly, should a researcher expect to see crime reductions following a policy that proactively disrupted the creation of criminal opportunities? Here, an aggregate analysis of crime in the larger jurisdiction may be warranted, particularly if it is the case that this larger jurisdiction encompasses where the disrupted crime would have occurred. A complication here is that the researcher must now think about jurisdiction-level counterfactuals, and identify a set of cities who were not exposed to the treatment of a proactive policy, but otherwise had, on average, similar determinates of crime, such that $\overline{\Omega_{Proactive\ City}} - \overline{\Omega_{Other\ Cities}} = 0$.

Alternately, the analyst may want to think about evaluating the policy at the individual level. Contingent on the records kept by the police, it may be possible to identify where the people pulled over in these aggressive stops live. Most large police jurisdictions also have access to booking records that allow researchers to determine the rate at which people living in specific places, such as census blocks, are accused of criminal activity. The frequency with which people living in those blocks were stopped by the proactive policy could be interpreted as a continuous treatment that might affect the criminal behavior (taking place anywhere in the city) of all of their neighbors.

Of course, such an analysis would also have to consider the initial challenge identified in this essay: what is determining the variation in treatment across neighborhoods? In order to credibly estimate the causal effect of increased proactive contact with the police on the criminal behavior of those living near where stopped people live, variation in police contact in the exposed and unexposed samples should be unrelated to other determinates of crime. Verifying that this is the case may involve examining both criminal activity over time, using police or court records, as well as census data on non-criminal activity like commuting patterns and vehicle ownership, which would influence the intensity with which people from different neighborhoods are exposed to the increased traffic stops.

Whether or not it makes sense to identify the frequency of crime in the location where the policy is implemented or the criminal activity of people who are exposed to the policy depends on the specific proactive policy in place. Stopping drivers seems to be a specific situation where you would want to think about the subsequent criminality of the stopped individuals and their peer group, rather than crime in the area where the stops occurred. To the extent that all proactive policies are aimed at disrupting crimes before they occur, an individual-level analysis which is less focused on evaluating crime in the places where

the policy is enacted and more focused on crimes by the people who come into contact with those policies, may be warranted more broadly.

How could such a person-level analysis of a proactive policy be completed? Local governments will, as a matter of practice, maintain records on all individuals who are booked, and using finger print supported identification numbers, it should be possible to track the criminal activity of booked individuals over time, usually along with their last known home addresses. License plate information for all people pulled over by the police should also be maintained by police departments, and may be available to researchers with specific confidentiality restrictions. However, many proactive policies create situations where police and citizen contact does not result in a booking, and also does not give police the legal authority to demand identification. A citizen who declines an invitation to identify themselves to a police officer (and, in the absence of reasonable individualized suspicion that is all that an officer is legally allowed to do) is "treated" by a proactive policy. However, there will be no administrative record generated by that encounter, and concerns about individual privacy mean that the cost of generating such records is likely quite high. The absence of a means of formally documenting who is treated by proactive policies remains a clear research challenge in modern policing.

References

LaLonde, R. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *American Economic Review, 76*(4), 604-620.