On the Dangers of using Randomized Experiments to Study Incentives in the Area of Crime: Comment on Radha Iyengar's "Does Arrest Deter Violence? Comparing Experimental and Non-Experimental Evidence on Mandatory Arrest Laws"

Rafael Di Tella
Harvard Business School and NBER

This interesting paper can be read as making (at least) two important points. First, it shows that mandatory arrest laws (MAL) in cases of domestic violence have been counter productive on at least one dimension, namely the frequency of serious abuse cases. Second, it shows that the method used for evaluating alternative policies against crime that has become the golden standard in criminology (and economics), namely randomized evaluation, can be seriously misleading. While I appreciate that the first contribution is important, particularly from the point of view of practice, I believe the second point has important consequences for the way we approach the policy evaluation more generally.

Let me start with the first point raised in the paper. It establishes that the moment when laws requesting the arrest of those accused of participating in domestic abuse is passed, average intimate homicides in the state (per capita) increases while that of family-member homicide falls. The effect is very large. And the difference in difference approach helps considerably in establishing the causal connection between the laws and the separation in the frequency of the two types of domestic abuse. Figure 1 illustrates.

The suggested interpretation is this: mandatory arrest laws reduce reporting by the victim (since it is more costly), so it escalates to homicide more often. Familial homicide drops because there is more reporting. The difference is presumed to occur because reporting of familial violence is made by non-victims (e.g., teachers). This puts the focus on reporting, something for which we unfortunately get no data. One alternative is to see if the response is particularly large in states with more serious punishments. There are differences in the severity of sentencing across states and over time (one example, of course is the death penalty). This is particularly important with the appearance of three-strikes-and-you-are-out laws. A similar point suggests that groups that are incarcerated at higher rates might be more inclined to reduce reporting in response to MAL. Incidentally, the similarity in response rates across racial groups is perhaps troubling for those who believe both in the results of this paper and the presence of racial bias in the legal system. So one reason this is an important paper is because it asks us to think harder about theories of reporting, which is an extremely important and understudied topic.

My second point is that the paper is making a point that far exceeds the issue of mandatory arrest laws. Indeed, I think it shows that randomized experiments are often not particularly useful in the analysis of policies against crime because we can't be sure about the quantity of information in the hands of the public. We typically want to look at incentives. But to study them, people must know about the different penalties they face. But if the criminals know, then the public will know. And they may not be keen on allowing randomization of penalties. Thus, the paper suggests that there are limits to the

use of randomized experiments in policy evaluation that we urgently need to understand better.

As the author reminds us, mandatory arrest laws were passed in the US in response to the results of the influential Minnesota Domestic Violence Experiment (MDVE). In it, the type of police intervention (arrest for at least one night, arrest and immediate release and simple warning plus reading of the rights of the victim) following the report of an incident, was randomized. The study revealed large drops in future domestic violence following arrest. Now, Iyengar's paper shows that it is wrong to extrapolate these results to justify MAL because they were obtained *conditional* on reporting. The public in general, and women in particular, were not informed of this experiment. The key point of the paper is that this difference is significant because of behavioral differences that may arise in the relationship between the battered women and their abuser.

Now the question is whether we can avoid this problem in the future by designing better studies. I am pessimistic for one simple reason: I do not think that the lack of external validity of the Minnesota experiment was the result of an avoidable mistake. It seems to me that it would have been impossible to communicate widely to women and other potential victims about the random nature of the program. And without such communication, incentives cannot be studied. I know that randomization of the treatment is standard in the scientific evaluation of the effectiveness of medicines. And that patients fully accept this. But it is also true that they voluntarily sign in to participate in such clinical trials. More importantly, we currently do not know much about the settings where the public will allow randomization of policies.

Public reaction to randomized experiments

It is of course hard to know if the public holds such heterogeneous preferences over the domain over which it is appropriate to conduct scientific evaluation of policies through randomized trials. But given that it seems important for us to know where people actually accept randomization and where they do not, I have run a small scale survey asking high school students (15-16 year olds) in Argentina the following two questions:

1. *In the US there have been two recent studies. In one, in order to find out if a certain medicine was appropriate in fighting cancer it had to be administered to only half of a group of patients (the other half receiving nothing). In order to decide which half, the researchers threw a coin. Do you find this procedure acceptable?*

   Yes
   No

2. *In the other (also in the US), in order to find out if a certain punishment was appropriate in fighting crime it had to be given to only half of a group of*

criminals (while the other half received a lower sentence). In order to decide which half, the researchers threw a coin. Do you find this procedure acceptable?

Yes
No

A sample of 18 high school students were interviewed (one on one). The results are as follows:

Question 1: Yes=14, No=4
Question 2: Yes=0, No=18

Given the cheap design, the results are obviously just suggestive. Still, they suggest that randomization is not supported in the evaluation of some policies designed to reduce crime. One possible explanation is that in the medical context it is a single agent decision problem. There is a procedure being considered and the person making the decision to engage in the trial bears all the potential costs. In the crime context there are also victims (and these need to be consulted if their victimizer's will be given lower than fair sentences).[1]

In brief, my point is that the paper shows that one of the most interesting and influential randomized experiments that we have available to inform the design of policy in the area of crime fails in some important way. And that the reason it fails is not because of a mistake that can easily be avoided in the future, but rather because experiments are not particularly useful in important areas in crime. Indeed, it is useful to remember that experiments in medicine hope to uncover the effect of a medicine on a person. In contrast, in economics, we are often interested in the effect of how affecting one person changes the incentives that other people have for engaging in specific behaviors. And it is precisely that "external" aspect that the public might refuse to study using randomized experiments.

---

[1] Incidentally, this would favor retribution vs deterrence as positive theories of punishment, as in Di Tella and Dubra (2008).