TRANSFORMING RESEARCH ON POLICING REFORM
May 18, 2021

1. The challenge

There is a crisis with policing in America. After the killing of George Floyd in the spring of 2020 millions of people around the country marched in the streets to demand change. Public opinion surveys reveal a decline in trust in police, particularly in communities of color.[1] Police kill as many people in the US in a month as England's do in 30 years.[2] The problems are not limited to abuse. There is also neglect; over the past 50 years the rate at which homicides are solved and offenders held accountable has declined from over 90% to 64.1%[3], with a large disparity in homicide clearance rates for murders involving Black versus white victims.[4] Nor is the current system serving officers themselves particularly well either. Morale among police officers themselves also appears to be down, as suggested by surveys of officers as well as news reports about record numbers of retirements[5] and difficulty recruiting new officers.[6]

One proximate cause for these problems is that far too little is known about how to improve policing. The current document is intended to help address this challenge in part by highlighting priority areas for additional research on policing. Of particular importance are efforts to understand how to improve policing in a way that is not just effective at controlling crime, but also fair. Historically most of the social science attention has gone towards understanding how to improve the crime-fighting function of police. Too little is currently known about how to reduce the harms from enforcement itself.

Beyond identifying the specific highest-priority R&D topics, the only way to solve the R&D crisis in policing is greater involvement by the federal government. We believe there is a need for the 'doing' parts of the US Department of Justice to increasingly use their grant-making in service of a 'learning' or R&D agenda about how to improve policing, so policing over time can improve.

The main challenge with policing research is the need to operate at a sufficiently large scale to be able to statistically measure impacts, alongside some form of planned variation (e.g., through randomized experimentation) that enables measurement of the independent effect of the relevant change in policy or practice. That planned variation is necessary to overcome the problem of confounding from unmeasured variable bias has been demonstrated through within-study comparisons of experimental and non-experimental studies dating back to LaLonde (1986), as well as examples in criminal justice specifically

---

[1]https://www.pbs.org/newshour/politics/two-thirds-of-black-americans-dont-trust-the-police-to-treat-them-equally-most-white-americans-do

[2] https://www.prisonpolicy.org/blog/2020/06/05/policekillings/

[3]https://www.npr.org/2015/03/30/395069137/open-cases-why-one-third-of-murders-in-america-go-unresolved

[4]https://www.npr.org/local/309/2019/10/09/768552458/chicago-s-dismal-murder-solve-rate-even-worse-when-victims-are-black

[5]https://chicago.suntimes.com/2021/1/15/22229584/police-retirements-backlash-chicago-new-york-minneapolis-john-catanzara-fop-michael-lappe

[6]https://www.forbes.com/sites/danielcassady/2020/07/29/amid-calls-for-police-reform-across-the-nation-police-struggle-with-recruiting-and-retention/?sh=726990284f94

such as with broken windows policing.[7] The need for a very large scale of operation for any demonstration project is evidenced by the fact that even in areas where randomization has been carried out in policing, such as with body worn cameras, the statistical uncertainty bands around the available estimates are so large as to often make conclusions from individual studies challenging.

The current funding mechanisms for R&D in policing make it possible, in our view, to make progress on only a subset of the most important open questions in this area. These questions are ones where the policing change could in principle be rolled out at the level of the individual officer, and so could be successfully studied within a single police department of whatever size with adequate statistical power to measure impacts. These include training initiatives of various sorts (de-escalation, implicit bias, procedural justice, etc.) and other forms of officer supports such as supplemental mental health assistance of various kinds, and perhaps alternative forms of first response to 911 calls for help.

But some of the most important policing questions involve larger organizational changes that must be made at either the district / precinct level, or even at the police department level. How do we reduce bias and improve accountability? How do we get police departments to understand and internalize public sentiment so that they can give local communities the type of police services residents want rather than focus on a potentially overly narrow objective function (reducing crimes, maximizing arrests or guns and drugs confiscated)?

Answering those questions is likely to require large-scale R&D resources of a scale historically never seen in the policing area, probably involving multiple cities at once. Most of the costs of those large-scale demonstrations will be not for the 'research' (analysis, measuring outcomes, etc.) but rather for the *operational costs* of implementing the new policies and practices themselves. This is certainly the lesson of the large policy demonstrations in other areas:
- In housing, the biggest cost of the US Department of Housing and Urban Development's Moving to Opportunity demonstration (allocated $193 million in current dollars by Congress) were thousands of voucher subsidies that have cost tens if not hundreds of millions of dollars over the life of the demonstration
- In health, the biggest cost of the $349 million RAND health insurance experiment or Oregon health insurance experiment was the health insurance subsidies
- In the Negative Income Tax experiments of the 1970s, the biggest cost driver of the $605 million spent on the different experiments was the cash payments to recipients (the same is true with current universal basic income experiments underway)

The lesson of the past 50 years of policy demonstrations, then, is that the boundary between 'operational costs' and 'research costs' is blurry in practice when learning about the effects of some new intervention requires fielding that intervention out in the world. For that reason the $86.5 million that DOJ set aside

---

[7] For example, one study of broken windows policing used natural variation in crime and arrest rates across New York City precincts to conclude that increased misdemeanor arrests led to reductions in violent crime (Kelling and Souza, https://media4.manhattan-institute.org/pdf/cr_22.pdf). The finding was later shown to be consistent with other explanations for observed drops in violent crime unrelated to broken windows tactics (Harcourt and Ludwig, https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=5364&context=uclrev).

explicitly for research activities -- for the National Institute of Justice (DOJ's research arm) and the Bureau of Justice Statistics -- is inadequate to this task.[8] Nor are the handful of millions of dollars provided each year by private foundations. Instead we need to harness the 'doing' part of the DOJ, such as by putting the $1.75 billion (with a "b") set aside for operational grants[9] to states, cities and juvenile justice agencies in service of a long-term learning agenda. Coupling federal grants with appropriately scaled evaluation studies is needed so that the reach of operational assistance extends past short-term implementation of activities, and instead helps the field learn what works to help move us towards a better future.

In what follows we provide some illustrations of priority research questions in policing, dividing them between those that could be studied under the current police research funding paradigm and those that could only be studied under an entirely new paradigm.

We understand that different researchers, funders, practitioners and the public, especially those living in communities most affected by the current policing system in America, will have different views of how to rank-order these open empirical questions. We do not intend ourselves to make any normative argument for what the right rank-order should be, but instead seek to make the larger argument that with a new funding paradigm ultimately they all can, and should, be studied (as well as many other key questions we have surely inadvertently excluded here).

2. Low-hanging fruit

The key challenge with policing research is to isolate the independent effects of the relevant policy or practice being studied on the policing outcomes of societal concern. There are two threats to accurate inference here:
- A *causal inference* challenge that arises from the fact that the people or places experiencing some new policy or practice may be different along other relevant dimensions besides exposure to the new intervention being studied.
- A *statistical inference* challenge that arises from needing to be able to measure in data impacts of the size we may expect to see from the interventions we study (for example in the survey sampling context in say a presidential race, news broadcasters often say "candidate A is leading B by X percentage points, plus or minus Y percent," that plus or minus Y percent is the sampling variability that creates the statistical inference challenge. If Y is large relative to X we cannot learn much about the world).

Most social scientists would agree that whenever possible the first-best solution to the causal inference challenge is randomized experimentation of the sort that provides gold-standard evidence in medicine.

---

[8] https://www.justice.gov/doj/page/file/1246791/download

[9] Operational grants such as those administered by the Bureau of Justice Assistance to states, local governments, and Indian Tribes serve a role in supporting innovation and development of novel programs and practices. Small-scale assistance for pilot programs (e.g., grants to individual police departments) can facilitate broader R&D efforts by contributing to operational knowledge and building out infrastructure, priming departments for larger-scale evaluation studies. But without commitment of resources towards evaluation, it is unclear which programs funded by operational grants actually succeed in improving policing outcomes.

URBANLABS
UCHICAGO
CRIME
LAB

Science in Service
of Cities.

Every social scientist would also agree that while optimizing the way a given dataset is analyzed can often help with the statistical inference challenge, at the end of the day sample sizes (number of officers, civilians, beats, districts, cities, whatever) need to be large enough to provide adequate statistical power to detect impacts on policing outcomes.

In our view there are three key topics that are currently amenable to rapid progress at research costs that we can imagine feasible under the current police-research funding paradigm:

- *Training*: Remarkably little is known about the effects of police training on policing outcomes at present, even though in other policy areas like education, we have proof-of-concept evidence that training can dramatically improve the productivity or 'output' that each employee provides (higher test scores in the education case).[10] Training can be rolled out in a way that incorporates planned variation at the officer level, almost every department will have some training capacity constraint especially for in-person training (creating the opportunity for research), solving the causal inference problem. And there are many departments across the country with enough police officers to make it feasible to solve the statistical inference challenge.

- *Officer supports*: The available data suggest that the prevalence of mental health problems among police officers exceeds (perhaps substantially) that of the general public, that diagnosis and treatment rates remain far too low, and that untreated mental health may contribute to adverse outcomes for the public at large (for example avoidable police use of force and other misconduct) and officers themselves (such as suicide). New physical and mental health supports can be feasibly studied through randomized trials because here (as with training) it is possible to roll out additional supports first to some officers before others, since most departments will again have some resource or other capacity constraints that make it difficult to start serving everyone at once.

- *Alternative first response options*: News accounts of police shootings against mentally ill people, or other requests for help that wind up in a potentially avoidable enforcement action, have increased policy interest in alternative first-response options for at least some types of 911 calls for service. Very little is currently known about the effects of these alternative services, but rapid progress could in principle be made since NIJ has already demonstrated with some initial experiments in the 1980s that it is possible to randomly assign at the call-by-call level what sort of response is deployed to a call.

We could and should learn more about each of these three policing topics, and believe it is possible to do so within current constraints and without major changes to the research funding paradigm. Those constraints mainly come down to the limited funding available. NIJ's annual spending on research, development, and evaluation is only $39-44 million per year[11], while the limited number of private foundations that also support work in this area almost surely provide even less than that in the aggregate.

---

[10] For example Joseph Allen et al. (2011) "An interaction-based approach to enhancing secondary school instruction and student achievement," *Science*, reports on the effects of a fairly light-touch coaching program for teachers called My Teaching Partner. At a cost of about $40 per student the intervention increases student test scores by 0.22 standard deviations - that is, roughly doubling the amount students learn over the course of the year (or put differently, doubling teacher effectiveness).

[11] https://www.justice.gov/doj/page/file/1246736/download

We can see the effect of this resource constraint: the list of topics that fall under the "low hanging fruit" category excludes many of the most important questions in policing.

3. Important questions that will be hard to answer well under the status quo system

Now consider what most people would consider the most important questions in policing. (These are not listed in any sort of attempted rank-order of importance, since they are all of absolutely vital importance. We also recognize that we may have omitted some equally important questions inadvertently):

- *How do we reduce racial bias in policing in America*? The current 'marketplace of ideas' has been very successful in stimulating and supporting research that documents racial bias in policing outcomes, ranging from laboratory studies of how police vs. civilians perform in use of force simulations to 'natural experiment' studies examining bias in police actions out in the field. The current research environment has been much less successful in identifying solutions. The consensus view of most experts is that implicit bias training is unlikely to substantially change bias in policing outcomes. Research on changes in hiring policies that substantially diversify who is the police have also led to disappointing results in terms of limited changes in policing outcomes.

- *How do we improve accountability for serious violence and reduce disparities in violence across communities?* Homicides are much more likely to lead to an arrest of a suspect when the victim is white compared to other races. In Chicago, homicides involving black victims are cleared at less than half the rate of homicides with white victims.[12] This pattern of neglect towards the communities most affected by violence has persisted for years and reflects a need for more equitable police response to violent crime across communities. To date, studies have relied on quasi-experimental methods to estimate the effects of specific investigative reforms on homicide clearance rates,[13] and have not explored which interventions might reduce differences in clearance rates across victim race. As homicide investigations typically involve multiple law enforcement personnel integrating diverse streams of evidence, reforms or improvements to the investigatory process would likely be difficult to roll out below the level of an entire homicide unit.

- *How do we make policing more responsive to the public*? Since the Compstat revolution in New York City during the 1990s, most large police departments now run some version of the data-driven management and accountability system. But that sort of data-driven system can only hold commanders and front-line officers accountable for outcomes that are actually measured, which historically has been 'outputs' such as crimes reported, arrests made, and guns or drugs confiscated. What is missing entirely in most cities is any measure of the harms of enforcement actions as perceived by the community, or more generally any sense of the public's satisfaction with the police services received. Absent such measures, there will inevitably be a distortion in what police try to achieve, over-emphasizing crime- and enforcement-related measures at the expense of responding to community sentiment. Some even argue that the solution could require

---

[12] https://www.wbez.org/stories/chicagos-dismal-murder-solve-rate-even-worse-when-victims-are-african-american/100a4f2f-e683-47f2-b3c5-db9e7a07d63e

[13] See Braga and Dusseault (2018), *Crime and Delinquency*.

shifting responsibility for some current policing functions out of the public sector altogether, for example to community-led NGOs.

- *How do we improve police accountability systems*? The current system for 'policing the police' seems to be the worst of all possible worlds. The general public has low confidence in the ability of the existing system to hold police accountable for misconduct, based on statistics like the share of citizen complaints against police in Chicago that wind up being sustained (only around 2%). And by many accounts police officers themselves have low confidence in the ability of the system to resolve cases fairly as well, which surely does not help in trying to overcome public concerns about a 'code of silence' where officers become reluctant to report on misconduct by other officers. Better methods are needed to improve existing data collection and facilitate both internal and external review of both individual officers and departments as a whole.

What the best candidate intervention is to even try to make progress on any of these problems is not clear at present.

But the deeper point we wish to make here is that whatever the right intervention turns out to be, there is a very reasonable chance that the 'observational unit' at which it is delivered is not the individual officer, but the police district or precinct, or even at the level of the police department itself. From decades of social science research ourselves, we are increasingly skeptical about the ability of panel-data non-experimental methods to solve the causal inference problem, given the frequent fragility of findings that result from modest changes in statistical model specification. So real progress will require randomized experiments at the district or even city level. And to solve the statistical inference challenge, that will require a scale of operation that regular current funding mechanisms are simply not designed to solve.

What is, in short, most needed to improve our R&D approach to improving policing in the US is not the right list of questions, but a re-imagining of the federal government's role in supporting policing research at a scale we have not seen to date. And the only way this could be achieved in the foreseeable future is if the operational dollars distributed by DOJ through the Office of Justice Programs become viewed as a central part of this new learning agenda.

Appendix A
Signatories

Anthony Braga
Northeastern University

Philip J. Cook
Duke University and NBER (elected member, National Academy of Medicine)

Jennifer Doleac
Texas A&M University

Ben Feigenberg
University of Illinois at Chicago

Dylan Fitzpatrick
University of Chicago

Jens Ludwig
University of Chicago and NBER (elected member, National Academy of Medicine)

Sean Malinowski
University of Chicago and Los Angeles Police Department (Retired)

Emily Owens
University of California, Irvine

Greg Ridgeway
University of Pennsylvania

Chico Tillmon
University of Chicago

Morgan Williams, Jr.
New York University