Pre-Analysis Plan
Policing in Historical and Cultural Context
Anita Ravishankar, Kevin H. Wilson, and Ryan T. Moore

February 27, 2020

Abstract

Police officers and residents often have different prior beliefs and expectations about policing that inform how they interact with one another. The Metropolitan Police Department (MPD) in Washington, DC acknowledges this difference, and its potentially persistent role in shaping police-community relations. Beginning in 2017, MPD sought to equip officers with detailed knowledge of their local community and historical context of policing in DC. Working with history professors from the University of the District of Columbia, MPD developed a novel training program—including a tour of the National Museum of African American History and Culture—to be delivered to all personnel. MPD’s objective is to provide officers with a better understanding of the community’s perspective. By teaching the history of the policing profession, particularly with respect to African Americans in the U.S. and specifically here in DC, MPD hopes to empower its officers to engage more effectively with residents. To test whether the training program is having the desired effect, MPD partnered with The Lab @ DC. We will measure the effect of the training on police-resident interactions using random assignment methods.

1 Contact: The Lab @ DC, Office of the City Administrator, Executive Office of the Mayor, District of Columbia Government, 1350 Pennsylvania Ave NW, Suite 533; Email: thelab@dc.gov.

We would like to express our gratitude to our partners at the Metropolitan Police Department, Chief Peter Newsham, Ben Haiman, Assistant Chief Robert Contee, Commander Ralph Ennis, Matthew Bromeland, and Heidi Fieselmann for helping to shepherd this research project from inception to implementation. Special thanks to Don Braman and Bill Egar from The Lab @ DC; Bill Axinn, Christian Davenport, and Nicholas Valentino from the University of Michigan; Hakeem Jefferson at Stanford University; Alex Coppock at Yale University; and Derek Hyra at American University for sharing invaluable insights and feedback, as well as David Yokum and Sam Quinney for general guidance and feedback.
Contents

Purpose of a Pre-Analysis Plan 3
Program & Project Context 3
Objectives 4
  Current State 4
  Meaningful Effects 5
Insights & Interventions 5
  Research Insights 5
  Intervention Overview 8
Methods 9
  Design. 9
  Randomization. 10
  Sample. 12
  Power Calculations. 12
Measurement & Analysis Plan 13
  Dependent variables - Measuring Change in Officer Behavior 13
  Primary Hypotheses 15
  Statistical Analysis. 16
  Secondary Analyses/Exploratory Analyses. 17
Discussion 21
  Limitations. 21
  Timeline. 22
A. Purpose of a Pre-Analysis Plan

The objective of this plan is to promote scientific research integrity by reducing researcher discretion after experimental outcomes have been realized. This plan is modeled on best practices for conducting and pre-registering field experiments.² We pre-commit to reporting all the analyses specified below. If we report any analyses in addition to those described here, we will indicate that the analysis was not pre-registered.

B. Program & Project Context

The Metropolitan Police Department of the District of Columbia (MPD) worked with professors from the University of the District of Columbia Community College (UDC-CC) to develop a new training program designed to give MPD officers a better understanding of the historical and cultural context in which they police. The course includes a lecture and a guided tour of the National Museum of African American History and Culture (NMAAHC) as well as a component specifically about the history of policing in DC. In addition, the instructors encourage officers to view a documentary, 13th,³ in advance of training.

MPD officers participating in the NMAAHC program are expected to learn more about the department’s role in shaping DC’s history and get a better sense of why and how police-community relations vary across the city. By learning the history of the policing profession, particularly with respect to African Americans in the U.S. and in DC, MPD wants officers to come away with a better understanding of the community’s perspective. This understanding should empower them to be more effective officers as a result. To determine the effectiveness of this novel approach to improving community engagement and building public trust in the police, The Lab @ DC is leading a randomized evaluation of this training program. Our findings will inform MPD’s future implementation of this training, as well as provide valuable insights for law enforcement agencies considering similar efforts in other jurisdictions.

The program was designed in late 2017, with implementation beginning in the Metropolitan Police Academy for all recruits. The program was part of the 2018 annual professional


development training provided to all sworn members, and all civilian staff will also have completed the training by January 2020.

Key partners on this project include Dr. Bernard Demczuk and Dr. Sharita Thompson from UDC-CC, who are leading the training. In addition, Dr. Nicholas Valentino, Dr. Bill Axinn, and Dr. Christian Davenport from the University of Michigan are consulting on the research design.

This project is exempt from IRB review as it falls under the program evaluation exemption indicated in Title 45 of the Code of Federal Regulations, §46.101(b).

C. Objectives

1. Current State

MPD conducts annual professional development training (PDT) for all sworn members to ensure officers are trained in the latest best practices and to improve the overall performance of its police force. One overarching objective is to build and maintain trust with the community MPD serves, with nearly all trainings having some connection to this long-standing goal. Whether officers are receiving training on procedural justice and implicit bias, or learning the nuances of search and seizure permitted under the 4th Amendment, all trainings are ultimately intended to equip MPD officers with the knowledge and skills to perform their duties to protect and serve DC’s residents more effectively.

All MPD members learn about the history of the police department as part of the curriculum delivered to recruits in the Metropolitan Police Academy (MPA). More nuanced lessons on the historical relationship between the police and diverse communities in Washington, DC, or background on the experiences of District residents, have not previously been a focus of past trainings provided to all sworn members.

A primary objective of this training is to improve police-community engagement. Leveraging administrative data, we measure outcomes including police use of force and complaints filed against MPD officers to understand the effects of this training. In 2017, prior to the

---

4 Police departments are generally divided into “sworn members,” typically whom we think of as police, and “non-sworn” members, often called the “civilian workforce.” In DC, the sworn staff includes officers on foot patrol, detectives, the police chief, and anyone else who can make an arrest. Non-sworn staff, such as author AR, typically perform administrative, policy, and research functions.

5 As this project will also be part of Anita Ravishankar’s doctoral dissertation, however, an application for determination of exemption was submitted to the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board and was granted on April 17, 2018.
implementation of the new training evaluated in this study, MPD reported 1,035 use of force incidents\(^6\) and 206 complaints filed against MPD officers (of which 59 were sustained).\(^7\)

2. Meaningful Effects

MPD is concerned primarily with the direction of the effect of the training, meaning that participation in the training reduces complaints and overall use of force, as measured using administrative data. If participation in the training is associated with a negative shift in officers’ behaviors, this would be considered a decidedly counterproductive outcome, and would lead MPD leadership to revamp the training (e.g., change the content and/or delivery mechanism). If our evaluation indicates that participating in the training appears to have no discernible effect (e.g., a null result), MPD leadership would consider extending the evaluation (e.g., through interviews/focus groups of participants) to help pinpoint opportunities to either a) improve the effectiveness of this training and update it accordingly, or b) develop alternate approaches to improving police-community relations.\(^8\)

D. Insights & Interventions

1. Research Insights

Police officers and residents often have different prior beliefs and expectations about policing that inform how they interact with one another. These differences can stem from the unique background, training, culture, and overall set of experiences that each party possesses—factors that influence the lens through which police officer and resident individually process a given interaction.

For residents specifically, variation among their perceptions of interactions with the police—especially along racial lines—is well documented.\(^9\) On the policing side, literature in


\(^8\) A null result may be due to a variety of reasons—it is possible the training has an effect, but we do not measure it during the study period; we do not capture the effect of the training in the outcomes we study; or, simply, the training has no effect. Qualitative research can assist in determining why a null result is observed, if this is the outcome of the study.

\(^9\) For example, see Hutchings 2015; Hutchings 2009; Kinder and Sanders 1996; Peffley and Hurwitz 2010; Peffley and Hurwitz 2007; see also Ghandnoosh 2014 for a review of this literature.
Criminology, psychology, and sociology makes note of the unique culture of policing, engendered and reinforced by the stressful nature of the work as well as the hierarchical, paramilitary structure of law enforcement agencies. Historically, this culture is characterized by the coping mechanisms officers use to manage their occupational environment: “officers cope with danger and coercive authority by being suspicious” and by “distancing themselves from the chief source of danger—citizens.” Though police culture is not necessarily monolithic, the existing literature suggests police culture and identity is distinct from that of the residents with whom they interact.

Acknowledging these differences, and their potentially persistent role in shaping overall police-community relations, MPD sought to equip officers with detailed knowledge of the community and the historical context in which they work. By providing officers with a greater understanding of how different members of the community view police and the historical foundations of those views—by updating officers’ beliefs and expectations—the training aims to promote more effective engagement and relationship building between officers and residents.

To our knowledge, no analogous interventions have been implemented and rigorously evaluated in law enforcement contexts, with the exception of a 2018 training on “principled policing” administered to police executives and trainers at 28 departments in California. Researchers evaluated this training—which included modules on expectations and legitimacy, procedural justice, the historical and general effects of policing, and implicit bias—using pre/post surveys of participants. They found participants reported substantial increases in their understanding of key concepts and sympathy toward community concerns. Their findings suggest that this type of training content may be effective, though additional research is necessary to measure the causal effect of such training on officer attitudes and behavior.

Evaluations of cultural awareness and sensitivity training delivered in other settings, to teachers and public sector contractors, have had mixed results, and with small sample sizes, raise some concerns as to generalizability. To date, much of the research on critical race theory has focused more on the potential utility of applying this paradigm in teaching and training contexts,

---

13 For example, Paoline III (2004) documents different occupational attitudes.
14 As of January 2020, approximately 17% of MPD’s sworn members report residency in DC; 60% in Maryland, and 22% in Virginia, and 1% elsewhere.
without much known regarding the effectiveness of this frame in improving understanding and/or relationships between racial groups.

Returning to the policing context, a 2004 National Research Council-sponsored report found “scarcely more than a handful of studies” on the effects of police training. \(^{18}\) Though there are over 18,000 law enforcement agencies across the United States, each of which provides a variety of trainings to its members, the empirical record on police training—of any type—remains thin to this day. \(^{19}\) The literature has not grown substantially in the intervening years, save for more recent work on procedural justice training \(^{20}\) and a growing interest in evaluating implicit bias training programs. \(^{21,22}\)

Police departments have developed and implemented a variety of innovative trainings in recent years, such as Holocaust Museum-based law enforcement trainings conducted in DC \(^{23}\) and Illinois, \(^{24}\) and an “Images of America” training on the historical relationship between police and African Americans administered in New Haven, CT. \(^{25}\) Many agencies have also implemented


\(^{22}\) There is a significant body of research on implicit bias interventions (related to both race and gender) outside the context of law enforcement. For example, see (1) Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of experimental social psychology*, 48(6), 1267-1278; (2) Lai, C. K., Hoffman, K. M., & Nosek, B. A. (2013). Reducing implicit prejudice. *Social and Personality Psychology Compass*, 7(5), 315-330; (3) Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J. E. L., Joy-Gaba, J. A., ... & Frazier, R. S. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General*, 143(4), 1765. This literature finds that there may be some intervention formats that are more effective in breaking prejudice habits or sustaining a reduction in bias as measured using an implicit association test or evaluation in an experimental/lab setting. Additional research on behavioral effects of these interventions is needed. We also underscore that, while the MPD training evaluated here is related to implicit bias, it is not implicit bias training.


\(^{25}\) Ibid.
trainings that include units on implicit bias and cultural awareness, though formal evaluations of these efforts have not been completed to date. Decision-makers have relied primarily on anecdotal evidence to assess the effectiveness of these efforts. Absent a systematic evaluation, it is difficult to state with certainty whether these trainings have the intended effect, as well as whether any effects are sustained over time.

2. Intervention Overview

The intervention we test is the administration of the NMAAHC training program to all MPD patrol sergeants and officers as part of MPD’s 2018 annual professional development training (PDT). Officers were informed of this training requirement through the teletype notification system used to communicate within the department, with participation coordinated by designated administrative staff within each police district.

The day-long training consists of a three-hour lecture by historians from UDC-CC, followed by a guided tour of the NMAAHC and a debriefing and discussion session. The lecture covers the history of African Americans in the US from the perspective of critical race theory (from the 1600s through to the present), with a specific emphasis on both the role of police through this period (e.g., in enforcing discriminatory laws) and local DC history. The guided tour of the museum highlights exhibits that connect back to the lecture materials, and also includes a lunch and free period to explore the pop culture, arts, and sports exhibits at the museum. The discussion session at the conclusion of the training allows participants to reflect on the material learned during the training and engage in an open dialogue on how the training might inform their daily work.

The objective of this training is to equip MPD members with a nuanced understanding of the historical and cultural context in which they police, and better understand the experiences of District residents. While the particular content and focus of this training are novel, the notion of updating one’s information environment in order to shift attitudes and behaviors is well established in the psychology literature.

---

26 See, for example: Anderson, J. “Stockton police chief helps lead statewide initiative.” TCA Regional News, November 19, 2015; Ferrain, E. “Suburban police take on challenge of implicit bias.” Daily Herald. September 5, 2017; Klecker, M. “Woodbury police use history exhibit in implicit bias training.” Star Tribune. February 9, 2019; training offered by the National Training Institute on Race and Equity and Fair and Impartial Policing, LLC.


28 For example, see Wilson, T. (2011). Redirect: The surprising new science of psychological change. Penguin UK.
E. Methods

1. Design.

This evaluation employs a stepped wedge randomized controlled trial to measure the effect of training on officer behavior, as well as pre/post surveys to measure change in officer attitudes. This pre-analysis plan focuses on the randomized evaluation; a separate pre-analysis plan will detail the survey methodology applied.

While all 4,500 sworn and civilian MPD personnel will receive this training, the evaluation focuses on measuring the effects of the NMAAHC training program on the frontline police force—full-duty, sworn patrol officers in public-facing, non-administrative positions of rank sergeant and below. The eligible population of 2,270 officers was randomly assigned to attend the training over the course of calendar year 2018 in what is called a stepped wedge design. Each month, a group of officers is randomly assigned to treatment — to attend the NMAAHC training. Eligible officers who have not yet received the training based on the randomly chosen order form the comparison, or control, group. For all months prior to the month in which that group of officers, or cohort, receives training, that cohort is in the control group; once treated, the cohort crosses over into treatment and remains in the treated condition for the remainder of the study, as depicted in the example shown in Figure 1 below. We measure outcomes for all participants at monthly intervals, beginning in January 2018 (t=0) and continuing monthly through December 2019 (t=23), at which point all cohorts had received the training.

---


30 In the MPD, the rank of sergeant is the first rank at which a sworn officer performs supervisory duties. Officers must have at least 5 years of patrol experience before being eligible for promotion.

31 At the time the study was designed and random assignments generated, it was the expectation of all parties (MPD and The Lab) that the study population would complete training during calendar year 2018. However, due to a variety of factors, all members of the study population did not complete training until December 2019. Reasons for delays in training ranged from rescheduling due to operational reasons (e.g., the officer was required to deploy) to the federal government shutdown in late 2018/early 2019, which closed the NMAAHC for several weeks, among other factors. We do not have any reason to believe that delays in training are due to the training itself; all participants knew that completion of the training was mandatory, regardless of when they were able to attend.
Figure 1. Visual depiction of a stepped wedge randomization design. At t = 0, no cluster is treated. At t = 1, Cluster 1 crosses over from control to treatment; at t = 2, Cluster 1 remains in the treatment group and Cluster 2 crosses over into treatment. Every month, a new cluster crosses over from control to treatment. At t = 5, all clusters are treated; outcomes are tracked for some time afterward to assess how sustained the effects (if any) are over time.

To measure the effects of the training on officer behavior, we compare outcomes for treatment and control officers. In particular, we measure police use of force, complaints filed against MPD members, and discretionary arrests using administrative data; the full list of outcomes is provided in Section F below.

2. Randomization.

Officers were assigned to treatment or control using a block randomized assignment procedure. This approach uses pre-treatment information to group officers into matched pairs, and then randomly assigns one member of the pair to treatment. We used the blockTools package for R to pair officers who are most similar to one another across the following characteristics: district, assigned unit, sex, race, tenure, past use of force, and past complaints. One member of each pair was assigned uniformly at random to receive treatment in the first half of the training sessions. The other would receive treatment approximately six months later. This approach ensures that the treatment and control groups are balanced across these covariates.

---


Random assignments were provided to district training coordinators, who are responsible for scheduling the members in their district for training. The first training session was scheduled for February 6, 2018. At the time of assignment, approximately 50% of the eligible sample was expected to have completed the training by mid-July 2018, with the entire eligible sample assigned to complete training by mid-December 2018.

Compliance with random assignment is monitored in two ways. First, the district training coordinators submit scheduling requests to the Metropolitan Police Academy (MPA). Staff at the Academy then must approve the request in order to finalize the date on which any member participates in the NMAAHC training. MPA staff were provided with the randomized assignments as well, and thus serve as a second check in the process to ensure members are scheduled to attend training for the assigned dates. MPA also tracks the date on which all members participate in training, allowing us to calculate the precise rate of compliance with randomly assigned order of training attendance.

While participation in the NMAAHC training is mandatory, given the operational environment in which this evaluation is being conducted, we expected some degree of two-sided non-compliance (e.g., due to new assignments, retirements, leave, etc.). We will conduct all primary analyses according to the original random assignment, thus estimating the intention-to-treat effect (ITT).

---

34 While delivery of the finalized version of training began on February 6, 2018 (with classes randomly assigned), several MPD members (generally command staff, some civilian groups, individuals in the academy) attended earlier pilot sessions of the training. Refinements to the training content were made based on this feedback, and our evaluation thus commences with the implementation of the “final” version of the training to randomly assigned MPD members.

35 The individual assigned to treatment in each pair was randomly assigned to complete training between February and mid-July 2018. The other member of the pair — the individual assigned to control — was assigned to receive the training approximately 5.5 months after his or her “treated partner” completes the training.

36 See Gerber and Green 2012, Chapter 6. As noted in Section E.1. above, we did encounter two-sided non-compliance in the sense that not all officers attended the training in their assigned month, though most officers in the study population (with the exception of those who left MPD over the course of the study) did eventually receive the treatment, e.g., complete the training. Specifically, of 2,270 individuals in the study population, 1,304 MPD members attended training in the assigned month; 709 went to a later training date, 55 went to an earlier training date, and 236 did not complete the training as they separated from MPD. Again, we have no reason to believe that the pre- and postponements of training were related to the training itself; rather, operational needs, scheduled leave, illness, scheduling conflicts with the NMAAHC, the government shutdown, etc. all affected the original training schedule. We report the ITT and also conduct robustness checks on compliance, reporting the complier average causal effect (CACE) as specified in Section F.

37 We will also estimate and report the Complier Average Causal Effect (CACE), which essentially involves rescaling the ITT by the proportion of compliers in the sample. As statistical inference is conducted on the ITT estimates, however, the choice of estimator does not affect hypothesis testing. See Gerber and Green 2012, Chapters 5-6.
3. Sample.

The study sample consists of all 2,270 MPD sergeants and officers assigned to non-administrative roles in the Patrol Services North and Patrol Services South Bureaus, the two bureaus with primary responsibility for patrol operations. While all MPD sworn members (as well as recruits and civilians) will be provided with the training, 1) patrol officers and sergeants are guaranteed (and required) to complete the training as part of the 2018 PDT; and 2) they engage most frequently with members of the public (and make up the majority of the police force), and so they are the focus of this evaluation.


Computations were performed with the **steppedwedge** command in the **st0341** Stata package. Throughout, we make conservative assumptions in calculating the minimum detectable effect this study is powered to measure. The sample size is set at 2,270, which is the population of patrol sergeants and officers as of January 2018, when the training program was finalized and randomization conducted. The evaluation is designed as a stepped wedge randomized controlled trial (RCT). In such a study, participants receive the treatment in waves, or clusters, at randomly assigned points in time.

The primary outcomes of interest for the behavioral change analysis are police use of force, complaints, and arrests. We calculate the minimum detectable difference in means to be approximately 0.05 uses of force based on power = 0.9, average cluster size = 190 officers (e.g., officers trained per month), intracluster correlation = 0.25, \( \alpha = 0.05 \), and control mean of 1.11 uses of force per person per year. We note that given our cluster assignment is random, we expect the intracluster correlation to be approximately 0; however, we use 0.25 out of an abundance of caution.

The experiment is well-powered to detect small differences between officers who receive the training and those who have not.

---

38 We exclude specialized units assigned to other MPD bureaus, such as Homeland Security and Investigative Services, as the nature of their responsibilities/operations differs from those sworn members assigned to patrol duties.


40 We calculate the control mean based on 2017 use of force records for officers in the control group.
F. Measurement & Analysis Plan

1. Dependent variables - Measuring Change in Officer Behavior

The randomized component of the evaluation is designed to assess the effect of the NMAAHC training on officer behavior. This analysis leverages MPD’s administrative data, measuring each outcome of interest at the month-officer level over the course of the study period, which runs from January 2018 (t = 0) through December 2019 (t = 23). We study the following seven families of outcomes:

- Use of force (all uses of force, serious uses of force)
- Complaints (all complaints filed, complaints sustained/not sustained)
- Two families of outcomes measuring officer discretion:
  - Traffic tickets and warnings
  - Discretionary arrests
- Three families of outcomes measuring officer performance:
  - Awards and commendations
  - Self-initiated calls for service
  - Sick leave taken

Each set of outcomes is described in detail below.

**Use of Force Incidents.**

Per MPD policy, a use of force incident is a self-reported use of force, and MPD requires its members to submit reports documenting all uses of force, as defined in General Order RAR-901.07. Uses of force can be subdivided, per MPD policy, into “serious uses of force” and “use of force.”

**Use of force (serious).** This includes:

- Firearm discharges
- Officer involved shootings
- Use of force resulting in a broken bone or an injury requiring hospitalization
- Use of baton
- All head strikes with an impact weapon
- Use of force resulting in loss of consciousness
- Use of force creating a substantial risk of death, serious disfigurement, disability or impairment of the functioning of any body part or organ
- MPD canine bites
● Use of force involving the use of neck restraints or techniques intended to restrict a subject’s ability to breathe
● Other use of force resulting in death

Use of force (other). This includes all uses of force not categorized as a serious use of force (e.g., hand controls, tactical takedowns, OC spray).

Complaints.
Complaints are reported to and investigated by the Office of Police Complaints (OPC). Information about each complaint is provided to MPD and entered into the Personnel Performance Management System (PPMS). In addition to measuring the effect of the training on the number of complaints filed, we also look at whether a complaint is sustained or not sustained, as defined below:

Complaint Sustained. A complaint is sustained when the allegation is deemed to be “supported by sufficient evidence to determine that the incident occurred, and the actions of the member were improper.”

Complaint Not Sustained Complaints that are not sustained have a disposition other than “sustained” (e.g., insufficient facts, exonerated, unfounded, pending).

Officer Discretion.
We examine the effect of training on officer discretion by measuring the following outcomes of interest:

Traffic Tickets and Warnings issued. The tickets and warnings included in this data were issued in personal, face-to-face interactions between MPD officers and members of the public (e.g., no parking tickets or red-light camera tickets are included).

Discretionary Arrests. Though all MPD officers enforce the same set of laws in Washington, DC and are required to follow the same policies and procedures set forth by the Department, officers have some leeway — some discretion — around when to make an arrest or issue a ticket (vs. a warning) for a traffic violation. Where we would expect officers to have less discretion in cases involving more serious offenses (e.g., violent offenses), officers have more leeway to determine when to make an arrest on more minor offenses; as such, we focus on this subset of arrests.

Per consultation with MPD officials, officers can exercise greater discretion to make arrests on charges in the following subset of offense categories:

● Disorderly Conduct
● Simple Assault
• Traffic Violations

We hypothesize that completing this training will reduce arrests in the three categories listed above, as the contextual information learned in the training may factor into officer decision-making in resolving an incident.

We compare the control and treatment groups on the number of arrests officers in each group made on charges in these three categories as a measure of the training’s effects on officer discretion.

**Officer Performance**

To evaluate possible pro-social effects of the training, we measure a variety of outcomes related to officer performance:

**Awards and Commendations.** Sworn members may be recognized for their service through commendations submitted by members of the public as well as through awards presented by MPD acknowledging a member’s achievements. These awards are tracked by MPD’s Human Resources Management Division. We account for both nominations and award receipt. We record awards and commendations as a binary variable (e.g., 1 if received in that month; 0 if no award or commendation received in that month).

**Engagement.** We examine the number of self-initiated calls for service as a window into officer proactivity and engagement.42

**Sick Leave.** We examine the number of hours of sick leave taken as a window into job satisfaction. We hypothesize the number of sick leave hours taken will decrease if the training successfully equips officers to be more effective in engaging the community, and thus, more satisfied with their work.43

2. Primary Hypotheses

We will test the following hypotheses:

- Officers in the treatment group will document fewer uses of force and receive fewer

---

41 A self-initiated call for service is one in which a patrol officer initiates some police action based on their own observations. For instance, if an officer witnessed a robbery, they do not need to wait for the person being robbed to call 911 before taking action. Self-initiated calls are an example of officer discretion, where different understandings of communities may lead to different interpretations of events on the ground which leads to differential choices of when to initiate a call.

42 We follow the Office of Unified Communications (OUC) business rule to identify self-initiated calls — those calls for which the call to queue time is 0 for field events.

complaints than officers in the control group.

- Officers in the treatment group will make fewer discretionary arrests than officers in the control group.

The directional hypotheses above reflect our expectations regarding the intervention’s effects. Because we are interested in evidence of positive or negative effects of the training, however, we use two-sided tests — e.g., we also specify the null hypothesis that the treatment will have no effect on average for any of the outcome measures described above. In total, we conduct nine tests.


A note on the population / sample distinction in our study. In a rarity for RCTs, our study is implemented, in some sense, on the entire population of interest, namely, MPD sergeants and officers, with a randomly assigned treatment. Thus, some of the traditional reasons for performing certain corrections below, e.g., implementing controls in a regression, may not apply.

We choose to implement controls and other corrections to give this study greater potential external validity. Consider for instance that MPD is not the only police force operating in Washington, DC, much less the entire United States, thus, knowledge of coefficients of control variables can help other departments adjust our results to their local context.

On the other hand, we are randomly assigning officers to treatment, we expect that the treatment assignment will be independent of any control variables. Since we are working with linear models, our extra controls can serve only to reduce the variance of our estimated treatment effect.

Note on categorical variables. In our regressions, we employ several categorical regressors that have more than two categories. In particular, we employ race, assigned district, and length of service at MPD. Each metric is measured in January of 2018. We will transform these categories with \( c \) values into \( c - 1 \) binary indicators.

- In the case of race, we consider it to have five values: white, black, hispanic, unspecified, and other.\(^{44}\)
- In the case of assigned district, there are seven districts in MPD.
- In the case of length of service at MPD, we will block this continuous variable into quartiles.

\(^{44}\) At the time of writing, some of our sources do not separate race and ethnicity and simply have an “Hispanic” race category. Thus, we collapse race and ethnicity into these categories.

\(^{45}\) “Other” will include, in particular, all multiracial categories.
The main specification in our analyses will be a regression of each outcome on treatment, with controls for the pre-treatment value of the outcome and officer demographics. Equation 1, below, provides the exact specification:

$$Y_{it} = \alpha + \theta Z_{it} + \beta_1 Y_{i0} + \beta_2 X_i + \beta_3 + \epsilon_{it} \tag{1}$$

where $Y_{it}$ is the value of the outcome under study, $Z_{it}$ is the treatment indicator (making $\theta$ the treatment effect of interest), $Y_{i0}$ is the pretreatment value of the outcome under study, $X_i$ is a vector of pre-treatment covariates that includes indicators for race, sex, assigned district, and length of service at MPD, $\beta_3$ is the month fixed effect, and $\epsilon_{it}$ is the error term. Note that we include a month fixed effect because the prevalence of particular outcomes is known to change month over month. For instance, overall, more crimes occur during summer instead of winter, and thus there is likely more chances for an officer and a civilian to interact. Individuals are indexed by $i = 1, \ldots, n$ and time periods are indexed by $t = 1, 2, \ldots, T$ (where $t$ is the month in which an individual is assigned to treatment). We take $Z_{it} = 1$ for the treatment month. This setup makes $\theta$ the true intent to treat effect.

We will estimate Equation 1 using weighted least squares (WLS) regression and estimate HC2 robust standard errors.

When computing $p$-values, we will employ randomization inference. The procedure is as follows:

1. Estimate $\hat{\theta}$ on the observed data using Equation 1.
2. Redraw $Z_{ik}$ 10,000 times according to the randomization scheme described above. In each case, estimate $\hat{\theta}_k$ according to Equation 1.
3. Observe how frequently the simulated $\hat{\theta}_k$ are as large or larger in absolute value than is the observed $\hat{\theta}$.
4. If this frequency (p-value) is smaller than 5%, we will reject the null hypothesis (of no difference in outcomes between Control and Treatment groups).

To guard against drawing false inferences due to multiple comparisons, we will report (in addition to raw $p$-values), $p$-values that are corrected by the Benjamini-Hochberg procedure, within “families” of outcome variables (“families” are described above in Section F-1).  


This section describes secondary and exploratory analyses that we plan to engage in if time permits and appropriate data are available. We are not committing to conduct them, but describe them here as a transparent record of analyses or intentions at the time of

---

pre-registration. If we do conduct any of the analyses below, however, we are pre-committing to report them and highlight that they were exploratory. If we conduct analyses not described here in the final report, we will specify that they were not pre-registered.

**Compliance**

Since the museum training is mandatory for MPD officers and sergeants, we anticipate very little non-compliance in the sense that people whom we assign to treatment do not eventually receive it. However, we did encounter a healthy amount of non-compliance in the sense that people did not attend treatment in the original month assigned. For instance, during the course of this study, the federal government (and hence the National Museum of African American History and Culture) was shut down due to a lack of federal appropriations.47 This caused a sizable number of officers’ trainings to be rescheduled. Moreover, the vicissitudes of life—illness, family obligations, operational needs—are likely to have affected on-time attendance. Of 2,270 members in the sample population:

- 1,304 members completed training in the assigned month;
- 55 members attended an earlier training date;
- 709 members attended a later training date; and
- 236 members did not complete the training.

Thus, we find that most non-compliance pushes the training forward into time. That is, an officer scheduled to attend training in April of 2018 is much more likely to actually attend training in May 2018 than March 2018.

While our primary analysis will remain an intent to treat estimate as described in Equation (1), we will provide two additional analyses that examine the effects of compliance.

First, we will compute the Local Average Treatment Effect (LATE), sometimes called the Complier Average Causal Effect (CACE). In general, this is equal to the intent to treat effect size \( \hat{\theta}_{ITT} \), divided by the compliance rate:

\[
\hat{\theta}_{LATE} = \frac{\hat{\theta}_{ITT}}{Compliance\ Rate} \quad (2)
\]

As explained above, our notion of compliance is a bit more difficult than a study that measures participants at a single point in time. For the purpose of computing (2), we will define compliance as follows: For each participant \( i \) and each time \( t \), let \( T_{it} \) be 1 if participant \( i \) had attended training at time \( t \), and 0 otherwise.

Then we define $C_{it} = T_{it} + Z_{it} - \max\{ T_{it}, Z_{it} \}$. Note that $C_{it} = 1$ exactly when $T_{it} = Z_{it}$. Then our compliance rate is

$$
\text{Compliance Rate} = \frac{\sum_{i,t} C_{it}}{\sum_{i,t} 1}
$$

As a second method of testing compliance, we would like to simply replace $Z_{it}$ in Equation (1) with $T_{it}$. However, given that we did not control the underlying distribution of $T_{it}$, we have no guarantees that $T_{it}$ is unconfounded. To wit, we apply the techniques of observational studies to estimate the treatment effect. Specifically, we follow Chernozhukov et al.\(^48\) and apply a double machine learning technique.

Specifically, we will start by dividing our data set in half, randomly assigning half of officers to a set $R$ and the other half to a set $S$. Using $R$ and five-fold cross validation, we will train a random forest to predict $Y_{it}$ given all the terms in Equations (1) and (2) except $Z_{it}$. Call the resulting model $g_{\theta R}(X_i, Y_{it}, t)$. Similarly, using $R$ and five-fold cross validation, we will train a random forest to predict $T_{it}$ given all the terms in Equation (1) except $Z_{it}$. Call the resulting model $m_{\theta R}(X_i, Y_{it}, t)$.

Similarly, we construct $g_S$ and $m_S$ using the other set.

We then form the estimate

$$
\hat{\theta}_{DML} = \frac{1}{2} \left( \frac{1}{|R|} \sum_{i,t \in R} (T_{it} - m_{R}(X_i, Y_{it}, t)T_{it}) \right)^{-1} \left( \frac{1}{|R|} \sum_{i,t \in R} (T_{it} - m_{R}(X_i, Y_{it}, t)T_{it}) + \frac{1}{|S|} \sum_{i,t \in S} (T_{it} - m_{S}(X_i, Y_{it}, t)T_{it}) \right)^{-1} \left( \frac{1}{|R|} \sum_{i,t \in R} (T_{it} - m_{R}(X_i, Y_{it}, t)T_{it}) - \frac{1}{|R|} \sum_{i,t \in R} (T_{it} - m_{S}(X_i, Y_{it}, t)T_{it}) \right)
$$

We will report this “double machine learning” estimate of the treatment effect.

**Rare Events**

While we employ varieties of OLS as our principal estimation tool, several of the outcomes of interest in this study, including use of force and complaints, are quite rare events. While estimation error due to the shape of measurement noise is eventually overcome by the central limit theorem, it is possible that models other than OLS may provide lower variance estimates in our regime. As such, for binary outcomes $Y_{it}$ appearing in equation (1), we will perform a negative binomial regression whenever the percent of $Y_{it} = 0$ is less than 5%. That is, if $\Sigma_{it} Y_{it} / \Sigma_{it}$

1 < 0.05, then we will use the same linear predictors indicated in equation (1) but drop \( \varepsilon_{it} \) and instead use a negative binomial link function.

If these robustness checks indicate a result different from the OLS model, we will report these results in the main text of our paper, urging caution for accepting that particular outcome. As robustness checks, we will provide analyses with several alternative generalized linear models, including at least a negative binomial model.

**Attrition Effects.**
As this training was administered over an extended period of time, we assume there will be some non-trivial degree of attrition as officers retire, separate from MPD, or change roles within the department. As such, we also conduct the behavioral change analyses described above on the study population still employed at MPD as of February 1, 2020 and report these results in the appendix.

**Spillover Effects.**
The primary treatment in this study is participation in the lecture and NMAAHC guided tour described above. Yet, we note there may be a spillover effect that increases over time as the proportion of the study population that has completed the training increases, thereby exposing the untreated population to some degree of spillover. This interference may diminish or amplify the primary treatment effect, as the lessons learned from the training program diffuse into the study sample through indirect means. To account for this time-varying, indirect exposure to treatment, we run Equation (1) but include an additional control, \( M_{it} \), for the log proportion of the sample treated at the midpoint of the month in which an officer was assigned to receive treatment, e.g., Equation (1) becomes:

\[
Y_{it} = \alpha + \theta Z_{it} + \beta_1 Y_{i0} + \beta_2 X_i + \beta_t + \beta_3 M_{it} + \varepsilon_{it}
\]

(4)

**Heterogeneous Effects.**
If we have sufficient statistical power, we will explore whether there are heterogeneous treatment effects based on officer age group, tenure on the force, sex, race, and assigned district. We will assess these effects through analyses that include interaction terms between covariates and treatment status. Though the study employs a large sample size, many of the outcome variables in the administrative data are relatively rare (e.g., use of force), and we anticipate a small effect size on outcomes that do occur more frequently (such as traffic stops or arrests). Thus, we consider these estimates exploratory.

---

49 For example, as more officers are trained, there may be a critical mass at which widespread behavior change occurs. This would diminish the treatment effect. On the other hand, there is some possibility that an "incomplete training" that we might contemplate from a spillover results in negative outcomes where a "complete training" results in a positive outcome.
G. Discussion

1. Limitations.

The intervention in this study is the first of its kind to be delivered broadly to a large, diverse police department in a metropolitan setting. As such, there are several limitations to the research design presented here, and we caution against overgeneralizing the findings, whatever they may be.

First, our quantitative analyses are based on the available administrative data, collected over the time frame allotted for training implementation. It may be the case that we do not detect behavioral change based on the measures presented here, but that alternate measures that were not available to us (e.g., public sentiment, perceptions of police legitimacy) would have detected the effect of the program on officer attitudes and/or behaviors. To address this concern, we pair the randomized evaluation with surveys and exploratory qualitative work (e.g., focus groups) to shed light on officer attitudes and directions for future research more broadly, as well as to help identify areas for improvements to the program content and delivery. For example, do certain individuals react more positively to the program than others? Is there a particular point in one’s career at which the training is particularly effective? Though we do not anticipate sufficient statistical power to conduct quantitative analyses to answer these types of questions, surveys, focus groups and interviews can help guide our next steps, and should be interpreted as such.

We also note that our findings will be generalizable to the extent that policing contexts elsewhere are comparable to MPD in 2018-2019, and training content is similarly structured and implemented. It may be the case that changes to the program design and/or content lead to different effects.

As noted above, spillover also presents a limitation—treated officers are not isolated from those who have not yet completed training, and spillover is likely to increase over the course of the study as more and more officers receive the training, and cross over into the treatment group. We include an additional control in our model to account for this concern.

2. Timeline.

50 The survey and qualitative research methodology will be detailed in a separate pre-analysis plan and posted to the Open Science Framework prior to conducting those analyses.
<table>
<thead>
<tr>
<th>Time Frame</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2017</td>
<td>● Design and pilot training program curriculum</td>
</tr>
<tr>
<td>Dec 2017</td>
<td>● Refine training program curriculum</td>
</tr>
<tr>
<td>January 2018</td>
<td>● Finalize training program curriculum</td>
</tr>
<tr>
<td></td>
<td>● Conduct random assignment</td>
</tr>
<tr>
<td>February-March 2018</td>
<td>● Training program begins</td>
</tr>
<tr>
<td>Fall 2019</td>
<td>● Conduct community engagement sessions</td>
</tr>
<tr>
<td>December 2019</td>
<td>● Complete training delivery</td>
</tr>
<tr>
<td>Early 2020</td>
<td>● Publish Pre-analysis Plan #1 (for Randomized Evaluation) on Open Science Framework</td>
</tr>
<tr>
<td></td>
<td>● Initiate analyses, and draft report</td>
</tr>
<tr>
<td>Spring 2020</td>
<td>● Publish Pre-analysis Plan #2 (for survey analysis) on Open Science Framework</td>
</tr>
<tr>
<td></td>
<td>● Initiate analyses of survey data and draft report</td>
</tr>
<tr>
<td>Summer 2020</td>
<td>● Deliver final report</td>
</tr>
<tr>
<td></td>
<td>● Conduct community engagement sessions on study findings and implications</td>
</tr>
</tbody>
</table>