The Effect of Field Training Officers on Police Use of Force

Chandon Adger, Matthew Ross, and CarlyWill Sloan *

January 6, 2022

Abstract

Over the past decade, police use of force has become an increasingly charged political issue with growing calls for reform. In this paper, we consider the link between a field training officer and a recruit's subsequent enforcement behavior. We leverage a unique setting where recruits are as good as randomly assigned to field training officers and where we have detailed information on calls for service. We document meaningful differences across field training officers in terms of their propensity to use force prior to being paired with a recruit. Further, we find that a one standard deviation increase in a field training officer's propensity to use force is associated with a 12 percent increase in their recruit's subsequent propensity to use force. The effect of having a more forceful field training officer persists for as much as three years after the recruit completes training. We find no evidence of an effect on other enforcement behaviors like arrests.

*Adger: Texas A&M University; Ross: Claremont Graduate University Sloan: Claremont Graduate University, carlywill.sloan@cgu.edu

Acknowledgments: We are grateful for useful comments from Steve Ross, the Dallas Police Department, and participants at the 2021 Southern Economic Association Conference. Sloan and Ross also thank the Russell Sage Foundation for providing generous support for this work.

1 Introduction

Over the past decade, police use of force has become an increasingly charged political issue with growing calls for reform. As of 2020, 65 percent of Americans believe that police officers are using an inappropriate amount of force (Pew Research Center, 2020). Concerns about appropriate use of force, combined with high-profile killings of unarmed Black individuals by police, has led most Americans to support the need for reform (Gallup Panel, 2020). However, there is substantial disagreement amongst policymakers and the public about how best to implement policing reform. For example, the well-known "defund the police" movement is supported by only 31 percent of Americans (Rakich, 2020) while policies like eliminating the enforcement of nonviolent crimes are only supported by 50 percent (Gallup Panel, 2020). Policymakers and the policing community frequently cite police training as a key approach to reducing police violence and racial disparities. In fact, the need for more and better police training is one of the few areas where the public, advocates, and the policing community can agree on a potential policy solution.

Despite the perceived importance of training, there has been little empirical research that considers the impact of police training on enforcement outcomes such as use of force. The few studies in this area have focused on conventional classroom-based training occurring at the police Academy or as part of continuing education requirements (McLean et al., 2020; Johnson et al., 2021; Owens et al., 2018). However, there is a significant gap in the research on the experiential phase of training broadly referred to as "field training". Apprenticeship-style models of field training are nearly universal among policing agencies in the United States and consist of an inexperienced recruit, who has graduated the police Academy, partnering with a more experienced officer on patrol for about six months. Although the few studies that have focused on classroom style training have found little to no long-term effect on subsequent enforcement behavior, there are two key reasons to believe that field training might have a larger and more persistent impact. First, the law enforcement community generally characterizes their field training as "the most important stage in the process of

becoming an independent police officer" (Warners, Ronald, 2020). Second, recent work by West (2019) suggests that an officer's on-the-job-experiences are a significant factor in their subsequent enforcement behavior.

In this paper, we ask whether recruits who are assigned to more aggressive (as measured by propensity to use force) field training officers (FTOs) are subsequently more likely to use force in the years following field training. To address this question, we leverage administrative data on calls for service (i.e. 911 calls) from the Dallas Police Department covering a period from 2013 to 2019. In this particular institutional setting, there is as-good-as-random assignment of recruits to training officers over the sample period. We believe that this quasi-random variation closely mirrors the ideal experiment and allows us to identify the causal effect of being assigned to an FTO who is more likely to use force on a recruit's subsequent enforcement behavior.[1] We characterize aggressive FTOs by constructing a measure of propensity to use force in the period prior to being assigned a recruit. In constructing FTO force, we account for the fact that some FTOs may be assigned to more dangerous calls or locations by considering factors such as neighborhood, characteristics of the call, and aspects related to date and time. Our measure allows us to assess whether a recruit is assigned to a particular FTO who has historically responded to similar incidents more aggressively than their peers.

In our preferred specification, we estimate that a one standard deviation increase in a FTO's prior propensity to use of force is associated with a 12 percent increase in a recruit's subsequent likelihood to use force after training. We also find that being assigned to a more aggressive FTO does not affect the likelihood of a recruit to make an arrest of any type in the years after field training. Further, we provide compelling evidence that suggests these results are not driven by treated officers choosing to work in neighborhoods with more severe

[1]We use the calls for service data to approximate a given recruit's FTO by calculating the senior officer that they appear with most frequently during each of the four field training phases. To account for the fact that calls with fewer officers are more informative about which officer is likely a recruit's FTO, we weight calls by the inverse number of officers on the scene. Although we have validated our call-based FTO definitions against additional data that explicitly identifies FTOs, we feel that our primary definition better represents exposure to a senior police officer even if that particular officer isn't always their official FTO.

calls for service. We interpret our results as indicating that those recruits who were assigned to a more aggressive training officers are not necessarily engaged in more active forms of policing. Rather, our results are consistent with the idea that these officer's threshold for applying force on any given call for service is simply lower than their peers who were paired with a less aggressive FTO. We also document that the effect of field training officers is long-lasting. Namely, we show field training officer effects persists for three years after the completion of training.

Our finding, that a recruit's subsequent enforcement behavior is shaped by their FTO, contributes to a broader criminology literature on the impact of culture in policing agencies. Drawing on an established literature in criminology (e.g., Skolnick, 1966; Westley, 1970; Paoline III, 2003), the President's Task Force on 21st Century Policing (2015) recommends that policing agencies transition from a "warrior" to "guardian" culture and states that "Field Training Officers impart the organizational culture to the newest members." Paoline III (2003) identifies the field as one of the most important settings for the transfer of policing culture and Paoline III and Gau (2018) states culture is key to reducing aggressive policing. Our work also contributes to literature in both criminology and economics that emphasizes the characteristics of individual police officers, rather than a monolithic culture, in shaping police use of force. Criminologists were among the first to investigate heterogeneity within policing agencies (Crank, 2004; Chan, 1997; Paoline III, 2006; Bruinsma and Weisburd, 2014; Woody, 2005). Most of the recent work in economics has focused on the role that an officer's race has on their propensity to use force and has reached a range of conclusions (Ba et al., 2021, Weisburst, 2019, Fryer Jr, 2019, Johnson et al., 2021, Hoekstra and Sloan, 2020). Our findings represent rigorous empirical evidence supporting policy initiatives aimed at reducing force by changing police culture particularly through field training interventions. Our findings also suggest that characteristics like aggression or prior enforcement behavior is a key predictor of aggregate rates of force.

We also add to the broader literature on the influence of social interactions in the work-

place. This literature documents that across many different settings, social influence by peers and superiors shape decision making. Studies have shown the significance of peers in a variety of settings such as school (e.g., Sacerdote, 2001; Whitmore, 2005; Carrell et al., 2013; Anelli and Peri, 2019, Bifulco et al., 2011), work (Mas and Moretti, 2009), neighborhoods (Glaeser et al., 1996, Billings et al., 2019, Billings and Schnepel, 2020), and the military (Murphy, 2019) are important for shaping future decision making. Related literature also documents the importance of teachers and managers (e.g., Bertrand and Schoar, 2003; Bloom and Van Reenen, 2007; Lazear et al., 2015; Giorcelli, 2019; Fenizia, 2021) in altering the subsequent behavior of students and employees. Our work contributes to both of these fields of research by considering the role of a particularly influential peer/superior in an extremely high-stakes setting, i.e. policing and use of force. Our finding also emphasize the importance of policy interventions that alter the composition of one's peers or their supervisor, particularly during on-the-job training.

The two previous papers closest to our work are Holz et al., 2020 and Getty et al., 2014. In Holz et al. (2020), the authors examine the impact of an officer's peer being injured on the job and find that other officers are more likely to use force, receive complaints, and injure suspects in the week following their peer's injury. We differ from Holz et al., 2020 by considering a fundamentally different determinant of force (i.e. FTO force propensity) and our findings have a much longer-term impact on subsequent policing behavior. Namely, we document higher use of force for officers trained by more aggressive officers for three years after the completion of their training. In Getty et al., 2014, the authors also use data from the Dallas Police Department to study the impact of FTOs on subsequent allegations of misconduct. Using a selection on observables approach and hierarchical modeling, the authors find that 26.5% of the variation in recruit complaints are attributable to their FTO. Relative to Getty et al., 2014, we implement a more rigorous empirical design which allows us to obtain plausibly causal estimates to address a different empirical question, i.e. the relationship between an FTO and a recruit's propensity to use force.

The findings of this paper have several important policy implications. First, by demonstrating that FTOs are an important determinant to a recruit's subsequent propensity to use force, we have identified a particularly promising target for policy interventions aimed at reducing force. To our knowledge, there has never been any targeted interventions aimed at substituting high for low-force FTOs in an effort to reduce aggregate rates of police force. Given our findings and the fact that short-term procedural justice interventions (McLean et al., 2020; Owens et al., 2018; Wheller et al., 2013) have lead to large but short-lived effects on enforcement behavior, there is reason to believe that reforms to field training might have larger and longer-lasting impacts on aggregate force. [2] Second, our findings are conservative in that we do not attempt to explore the dynamic effects of changing the composition of FTOs. In particular, our findings suggest that selecting less-forceful FTOs in any given year would lead to less forceful recruits who will then subsequently become FTOs in the future. Thus, the dynamic effects of altering the composition of FTOs in any given year are potentially large and last very far into the future. Third, changing the composition of FTOs is much more cost-effective (even potentially costless) relative to other interventions involving formalized police training. Finally, reforms aimed at changing the composition of FTOs are much more practical and politically palatable than large-scale initiatives to defund or dramatically alter enforcement or the composition of the police force.

2 Police Officer Training and Institutional Background

The training received by recruits in Dallas is representative of how most agencies across the country train new police officers. In particular, training is divided into two distinct phases

[2]Procedural justice refers to the idea of fairness in the processes that resolve disputes and allocate resources. It is a concept that, when embraced, promotes positive organizational change and bolsters better relationships. Procedural justice speaks to four principles, often referred to as the four pillars: fairness in the processes, transparency in action, opportunities for voice, impartiality in decision making (COPS, 2020). There is also body of literature focuses on the impact of short procedural justice training sessions on officer attitudes (e.g., McLean et al., 2020; Rosenbaum and Lawrence, 2017; Schaefer and Hughes, 2016; Skogan et al., 2015. We differ from this literature because of our focus on officer actions.

before recruits become a full-fledged police officer, i.e. Academy training (phase 1) and field training (phase 2).[3] Although our paper focuses exclusively on the field training component of a recruit's preparation for becoming a police officer we provide a brief but comprehensive discussion of the full training process in this section.

First, recruits must complete basic training at the Dallas Police Academy. Training at the Academy lasts 36 weeks and consists of 1,431 hours of instruction. At the beginning of their time at the Academy, recruits undergo mental and physical training aimed at preparing recruits for the demands of a career in policing. Next, the recruits complete legislatively mandated classroom and scenario-based training as well as a number of additional courses required by Dallas PD. The legislatively mandated courses are developed by the Texas Commission on Law Enforcement which is the regulatory agency governing the licensure of all peace officers in Texas. This agency also regulates subsequent in-service training requirements which are necessary to maintain a peace officer license in Texas. In most states, there is a similar governing agency (known colloquially as "Post", i.e. Police Officer Standards & Training) which sets both Academy and in-service training requirements. Although there is variation in the specific training required in different states, a national organization (the International Association of Directors of Law Enforcement Standards and Training) issues a core set of recommendations which have been broadly implemented across the country and are consistent with how Dallas trains police officers.

After completing basic training at the Academy, recruits enter a second phase of training referred to as "field training". As noted above, this second phase of training is the focus of our study and has largely been overlooked by the existing empirical literature on policing. During field training, recruits ride with more experienced officers (FTOs) in an apprenticeship style model where they are gradually afforded more autonomy. FTOs have a dual responsibility of providing service in their assigned sectors and providing on-the-job training for recruits. At the end of field training, recruits are evaluated on their performance and graduate to

[3]Police Officer is the lowest rank in the Dallas Police Department

becoming full-fledged police officers. In our setting, nearly all recruits who successfully graduate from the Academy also complete field training and virtually all of those officers are initially assigned to patrol, i.e. answering calls for service. This apprenticeship style model of on-the-job training was first developed by the San Jose Police Department in the 1960s and has since become a near universal standard in law enforcement agencies across the country.

It is important to note that recruits in Dallas have no discretion in choosing their FTO or their initial assignment. In particular, recruits are assigned to one of seven divisions in Dallas to complete their field training. This assignment is based on the staffing needs of the division rather than the skills or performance of recruits. In interviews with Dallas police officers, they noted that recruits are assigned to divisions based largely on retirements or promotions. Command staff in Dallas also indicated that the decision regarding which recruits are assigned to which divisions is unrelated to a recruit's characteristics or performance in the Academy. Within a given division, recruits are then assigned to FTOs and their associated patrol sectors/beats. As before, command staff at Dallas PD indicated that these decisions are unrelated to recruit characteristics or performance at the Academy. In a subsequent section, we provide empirical evidence supporting the claim that the initial assignments provide as-good-as random variation in the pairing of recruits to FTOs.

In Dallas, the field training process takes a total of six months to complete and consists of four phases. In the first and fourth phase, a recruit is paired with the same FTO. In the second and third phase, the FTO is different. The first three training phases of field training are seven weeks long. The final evaluation phase is conducted by the initial (i.e. phase one) FTO and lasts three weeks. When field training begins, recruits are instructed to take on a more observational role. As training progresses, they are given more autonomy and become an active participant in responding to calls for service. For example, in the early weeks of field training, a recruit may simply watch a FTO respond to a call for service. In later phases, the recruit may lead the response under the guidance and observation of their FTO. FTOs also

conduct frequent, often daily, evaluations of recruits. According to command staff in Dallas, these evaluations are largely used to provide the recruit with extensive feedback on their performance. After field training is complete, recruits then spend another year on probation where they are required to stay in their initial division assignment and associated sector. During the first six months of probation, commonly called "little t" by Dallas command staff, recruits are required to choose a more experienced officer to ride with as their partner. Finally, one year after completing the Academy, recruits are taken off probation and advance to the position of Police Officer.

This paper focuses on the impact that the first FTO has on the recruit's subsequent enforcement behavior. We made this decision for two reasons. First, in our conversations with Dallas police officers, they communicated that field training shapes officers' policing "style" much more than their training at the Academy. Command staff in Dallas also emphasized that this phase of training is the most critical part of a recruit's development and that all peace officers remember the lessons learned during field training for the rest of their career. Second, Dallas command staff also noted that the first phase of field training is the most significant because it's a recruit's first exposure to providing service. Further, recruits often return to their initial FTO for their final training and evaluation phase. We also focus on FTOs rather than the officer a recruit chooses to work with during "little t" because FTOs are conditionally randomly assigned. In contrast, a recruit may select their partner during "little t."

In our study, we document FTO and recruit behavior using 911 calls for service. When a civilian calls 911 in Dallas, they are first connected to a 911 operator. The operator will then record essential characteristics of the call such as location, description of events, and time in the Computer Aided Dispatch System (CAD). The operator will also place the call into a standardized category, such as "domestic disturbance." Finally, the operator also records their perception of the urgency and severity of a call. This is referred to as the priority of the call and is assigned values from 1 to 4, with 1 being the highest priority. The information

recorded in CAD system is then provided to police dispatchers who assign calls to police officers. Dispatchers look at the queue of active calls and assign available officers to active incidents. Dispatchers can locate officers on a map and identify which officers are available based on whether the officers are involved with any other incidents at the time of the call. Officers are assigned primarily based on location and availability. If there are many more active calls than available officers, lower-priority calls are postponed until higher priority calls are resolved. Dispatchers also decide the number of officers to assign to a call based on the type of call. For example, more serious incidents (such as shootings and mental health calls) may involve the dispatch of multiple officers. Once an officer responds to a call, the officer is afforded a significant amount of discretion in how they handle an incident in terms of their decision to make an arrest or use force.

To measure officers use of force, we link 911 calls to force reports.[4] After any force incident, officers are required to make a Response to Resistance entry in BlueTeam, the Dallas Police Department's incident tracking system.[5] All incidents are reviewed by a supervisor (Dallas Police Department General Orders, 2015). According to the Dallas Police "The physical control techniques used may range from the use of handcuffs in an arrest, strikes with an impact weapon, or the use of a firearm" (Dallas Police Department, 2019). We also link 911 calls to arrest reports. Here we observe the type of arrest made (felony, misdemeanor, or n-class) and well as demographics of the arrestee (race, gender, age).[6]

[4]Dallas refers to force as a response to resistance.

[5]Any Response to Resistance that is Soft Empty Hand Control or above on the Response Continuum, with the exception of "Compliant Handcuffing" only. This will include, but not be limited to the following: 1. All take-downs, pressure points, joint locks 2. Any use of Oleoresin Capsicum Chemical Spray. 3. Any deployment of the Pepperball System. 4. Personal weapons such as hands and feet. 5. Any use of the baton or any other type of instrument that is used as an impact weapon. 6. Any use of an Electronic Control Weapon (Taser). This includes accidental discharges of the Taser. 7. The deployment of a firearm which is pointed directly at any individual.

[6]Most "n-class" arrests are made for warrants.

3 Data and Summary Statistics

Our data comes from the Dallas Police Department and includes all calls for service from July 2014 to July 2019. We link this data to Dallas force reports, arrest records, and officer characteristics.[7] To our knowledge, the Dallas Police Department does not keep a list of recruit-FTO pairings for each of the four field training phases. However, we have been provided detailed information on the dates of specific assignments for each officer in our sample as well as Academy graduation dates. Thus, we are able to construct recruit-FTO pairings for each field training phase using these dates as well as the likelihood a recruit arrives to a call with a senior police officer.[8] In particular, we construct a set of dates for each recruit which are associated with each phase of field training. We then identify the senior officer that a recruit is most likely to arrive to a call with during each phase and assume that this is their FTO during that phase. To account for the fact that many officers are assigned to more severe calls, we apply a set of weights equal to the inverse number of senior officers on a given call. Thus, calls where there is only one other senior officer at the scene with a recruit receive more weight than those with many other officers on the scene. In our sample, we have a total of 411 recruits and we identify a total of 232 distinct phase 1 FTOs.

The Dallas police department typically requires that FTOs achieve at least the rank of Senior Corporal however, to our knowledge, this is not strictly enforced. This is particularly true during out sample period when there were a large number of requirements. As of 2019, 50 percent of officers held the rank of senior corporal or above. Given these statistics,

[7] In linking the force records with calls for service, we do so based on the incident identifier but not the officer badge number. Although our results are generally robust to linking this data on both incident and badge number, we have taken a conceptual stance that it is more correct to associate an incident resulting in force with every officer on the scene. This is because one officer may influence another officer on the scene. We also restrict force incidents to those we are confident (based on the time stamp) occurred on the scene of an incident as opposed to those occurring after a suspect is in custody. We attach arrests to calls for service in a similar fashion such that all officers on the scene are associated with a given arrest regardless of whether they are the specific officer listed in the arrest report.

[8] The first seven weeks after the Academy are phase one, the second seven weeks phase two, the third seven weeks phase three, and the last three weeks are phase four.

we have a reasonable amount of confidence to believe that we have correctly identified the recruit-FTO pairings in the vast majority of our sample. Furthermore, we do not feel that misidentification of these pairings creates any bias in our subsequent results. In particular, we are confident that these are the senior officers that recruits have actually shadowed on the largest number of calls during their phase one field training. Thus, these are the senior officers most likely to have an impact of a recruit's subsequent policing behavior regardless of whether they are the true administratively assigned FTO.

Since our analysis focuses primarily on the impact of the first FTO, we only provide summary statistics related to that pairing. Police Officers are eligible for promotion to Senior Corporal after three years of service. According to Dallas command staff, most officers who stay with the force for three years should expect a promotion.[9] In interviews with Dallas command staff, it was stated that there is barely enough FTOs to meet the volume of new recruits who replaced retiring officers in our sample period. Although command staff emphasized that there is still some selection in terms of who they allow to become an FTO, it is not necessarily a position reserved for only highly experienced or exceptionally talented officers. According to our data, the average age of a FTO is 48. This is three years younger than the average age of an average patrol officer. FTOs are also generally representative of the whole police force in terms of demographics, but perhaps a bit less diverse. Specifically, 19 percent of FTOs are Hispanic, 16 percent are Black, and 63 percent are White, compared to 20 percent, 23 percent, and 53 percent in the entire force, respectively.

Next, we assign each of the 411 FTOs-recruit pairs a force rate based on the FTO's propensity to use force in the period prior to being assigned a given recruit. To do so, we estimate pair-specific fixed effects, which represent a police officer's time-invariant propensity to use force on a call for service. Specifically, we regress an indicator for a call resulting in force on a fixed effect for each recruit-FTO pair using only calls for service answered by the

[9]There are four main positions within the Dallas Police Department. Officers begin with the rank of Police Officer and then can advance to Senior Corporal, Sergeant, and finally Lieutenant. Each promotion entails a pay raise.

FTO in the period prior to being assigned a given recruit. In estimating this fixed effect, we also control for important call characteristics such as the number of officers on the scene, beat, type of call (priority-by-type) year-by-month, hour, and day of the week fixed effects.[10] The intuition behind this exercise is to create a measure that captures how likely an FTO is to use force after accounting for the fact that some officers may respond to different types of calls than others. Formally we estimate:

$$force_c = \lambda_{o,r} + \beta_1 X_c + \epsilon_{o,r,c} \tag{1}$$

where $force_c$ is a binary variable equal to one if call c ends in force and zero for all other observations. X_c includes fixed effects for the number of officers on the scene, beat, type of call (priority-by-type) year-by-month, hour, and day of the week. The estimate of interest $\lambda_{o,r}$ captures an recruit-FTO pairing that summarizes a FTO's o average propensity for using force before being assigned a given recruit r. Higher values of λ_o, r indicate a FTO is more aggressive or uses force more frequently, and lower values of λ_o, r indicate a FTO is less aggressive and uses less force. We cluster standard errors on the FTO, rather than the recruit-FTO pair since some FTOs appear more than once with different recruits.

To estimate a more conservative measure of force usage, we adjust our estimate $\lambda_{o,r}$ using Empirical Bayes following Weisburst (2019). To do so, we construct a shrinkage factor that attenuates the estimates for officers with few calls for service or small within-officer variance towards the sample mean. Specifically, we estimate across officer variance, σ_A^2 and a within-officer variance, σ_W^2 [11] Next, we use our two variance measures and the number of observations per officer to estimate a shrinkage factor $\frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma_W^2}{N_{o,r}}}$. Finally, to estimate our final shrunken force rates by multiplying our shrinkage factor by our original fixed effects such

[10] In practice, this means some FTOs will have more than one fixed effect. However, across different recruit's FTOs have remarkable similar estimate force rates. The correlation coefficient from a comparison of the two measures is 0.77.

[11] Formally our we calculate within officer residual variance as $\sigma_W^2 = E(\epsilon_{o,r}^2)$.

that

$$\Lambda_{o,r} = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma_W^2}{N_{o,r}}} * \lambda_{o,r} \tag{2}$$

where we plot the distribution of police officer force rates for all 411 FTO-recruit pairs in Figure 1. Figure 1 (a) plots the distribution of force rates for our raw and shrunken measure. As expected, the distribution of the shrunken effects is narrower (has a smaller standard deviation). Values above zero indicate that the field training police officer is more likely to use force than the average FTO. A number less than zero indicates that the FTO is less likely to use force than the average. In our analysis, we also categorize FTOs as high force or low force. We define high force as an officer with a fixed effect greater than zero. Said another way, a FTO is called high force if they use force more than the average FTO.

For the rest of our analysis, we will focus on a standardized, shrunken measure of FTO force for ease of interpretation. However, we note that our results are generally robust to using the unshrunken estimates as well as a number of alternative specifications. The distribution of standardized effects is shown in Figure 1 (b). One standard deviation increase in FTO effects is a 0.0091 percentage point, or 9 percent increase in average use of force compared to the sample mean of 0.10576 percent. Moving from the FTO that uses the least amount of force to the most is nearly a 7 standard deviation increase, and replacing an FTO at the 10th percentile for one at the 90th percentile represents a 20 percent increase in average force.

Interestingly, when we consider if FTO demographic characteristics are predictive of force rates, we find that gender, race, and age are not meaningful predictors of an FTO's propensity to use force. In Figure 2, we plot the distribution of force rates by FTO characteristics. The main take away of this graph is that the distributions look remarkably similar to each other, indicating that, at least for FTOs in our sample, force rates cannot be easily predicted by FTO race, gender or age. We interpret this as evidence suggesting that our measure captures an otherwise unobservable aspect of an FTO's enforcement behavior which is likely related

to aggression.

Finally, we compare FTO force rates to the force rates of other patrol officers in Dallas. To do so, we first construct a force rate for each officer using our entire sample of calls for service. Next, we shrink and standardized the force rates as described above. Our results are shown in Figure 3.[12] On average, FTOs use force less, 1 standard deviation on average, than the typical non-recruit officer, and less than the average senior Corporal or Sergeant (the ranks most likely to be FTOs). Even though the distribution of force rates for FTOs is to the left of the other two samples, there is still significant overlap between the distributions. Our takeaway from these figures is that while FTOs may be selected on force usage to a certain extent, their propensity to use force does not make them outliers relative to all other patrol officers.

We present summary statistics at the recruit level in Table 1. As noted, there are 411 recruits in our sample. This translates to roughly 90 new recruits each year. The average recruit is much younger than the average FTO. Most recruits are White (44 percent), 21% are Black, and 30% are Hispanic. Given the conditional random assignment of recruits to FTOs, we would expect that recruit characteristics shouldn't differ across the type of FTO. Although these summary statistics do not reflect the exact comparison we use in our formal tests of balance where we control for cohort year by division fixed effects, it is worth noting that recruit characteristics look remarkably similar across high and low force FTOs. A t-tests of the difference in recruit characteristics across high and low force FTOs is not statistically significant.[13]

In our main analysis, we evaluate recruit behavior after field training using data on their subsequent calls for service. Summary statistics at the call level are presented in Table 2. In our sample, roughly 3.5/100 calls end in an arrest and only 1/1000 calls end in a use of force. We characterize a call as having involved force or arrest regardless of the specific

[12]There are a few (4 percent) very extreme force users in our sample that we drop to create a figure that is easier to "see".

[13]We estimate p-values of 0.2025, 0.7134 ,0.3353, 0.3032, 0.1686 for test of difference across means in race, gender and age.

officer who used force or made the arrest. This conceptual decision was motivated by possible endogenity in terms of the specific officers on the scene of an incident and who actually ends up using force. However, it is worth noting this restriction improved the precision of our estimates, but that our results are generally robust to using measures of force and arrest that are linked both to specific incidents and officers. [14] Our call data also includes other important characteristics that may impact police officer behavior on the scene. Specifically, we observe the call type, priority (a measure of urgency and severity), location, date and time.

4 Empirical Methods

4.1 Estimation Model

The conditional random assignment of FTO to recruits provides an ideal context for investigating how FTOs shape the subsequent policing behavior of their recruits. To do so, we estimate the effect of FTO force rates on a recruit's subsequent use of force and arrests. Specifically, we estimate:

$$force_c = \theta_r + \beta_1 \Lambda_{o,r} + \beta_2 X_c + \epsilon_{r,d,c} \tag{3}$$

Where $force_c$ is a binary variable equal to one if call c ends in force. Our primary variable of interest $\Lambda_{o,r}$ represents the propensity of a recruit's FTO to use force in the period prior to the pairing. As discussed, we shrink this measure using Empirical Bayes and standardize it for ease of interpretation. Thus, our coefficient of interest β_1 can be interpreted as the change in a recruit's likelihood of using force on a given call from a one standard deviation increase in their FTO's prior propensity to use force. We control for possible variation across

[14] For force incidents, we also require the time on force report to be between when the first officer arrived and the call was cleared. This sample restriction was made because we suspect some force incidents occur after a suspect is in custody. Again, this sample restriction improved the precision of our estimates but our results are robust to alternative unrestricted specifications.

recruits in their initial assignment over time by including θ_r representing a set of 38 Academy cohort year by division fixed effects. To control for variation across calls, we also include X_c representing a vector of call and officer level attributes. In our fully saturated model, this vector includes recruit characteristics (age gender, race), field trainer characteristics (age, gender, race), geographic fixed effects (beat), call characteristics (priority, call type), number of officers dispatched, as well as year-by-month, and day of the week fixed effects. Standard errors are clustered at the recruit level.[15]

The model's identifying assumption is that FTO characteristics, primarily force rates, are not correlated with recruit characteristics after controlling for division by cohort year fixed effects. Identification, therefore relies on the conditional random assignment of recruits to FTOs. The identifying assumption could fail if being assigned a high force FTO is correlated with other factors that also alter a call's outcome. For instance, high force FTOs may choose to work with recruits they believe are similar to them in other police departments. In this case, we would overstate the influence of FTOs. We avoid this problem by using an environment where FTOs are conditionally randomly assigned. In the next section, we empirically demonstrate that FTO characteristics including propensity to use force is uncorrelated with recruit characteristics.

4.2 Research Design

We begin this section by showing that FTO characteristics are not correlated with confounding factors. While we expect this to be true based on discussions with Dallas command staff about FTO assignments in Dallas, we also provide empirical evidence. To begin, we regress FTO characteristics on recruit characteristics where the unit of observation is a recruit-FTO pair.[16] Standard errors are clustered at the FTO level.[17] Each specification includes division

[15] We are also robust (i.e. statistically significant at the 5 percent level or less.) to two-way clustering by recruit and FTO as well as recruit and division by cohort year.

[16] As discussed, a recruit-FTO paid means that there is one observation per recruit but each FTO can be assigned to multiple recruits over the sample period.

[17] Robust standard error also yield very similar results.

by cohort year fixed effects because we believe that FTOs are as-good-as randomly assigned to recruits within cohorts and divisions. Specifically, we investigate whether FTO age, race, and force rate are correlated with recruit age, race, gender, and hire date (measured in years). The results of this test are shown in Table 3. Of the 35 coefficients reported, only one is statistically significant at conventional levels.[18] Further, to observe whether FTOs are assigned similar recruits, we conduct F-tests of joint significance. None of the p-values from the F-tests are statistically significant at conventional levels.

We also plot the distribution of FTO force rates by recruit characteristics in Figure 4. Given the random assignment of recruits, we would expect that these distribution would be very similar. Indeed, the distributions appear to very similar and a Kolmogorov-Smirnov test also fails to estimate statistically significant differences across the distributions. These results indicate that recruit characteristics are generally orthogonal to FTO characteristics and are consistent with the institutional background that recruits are as good as randomly assigned to FTOs. Thus, we feel that our estimates are plausibly causal in nature and that the coefficient β_1 on the variable $\Lambda_{o,r}$ from the prior section can be interpreted as the effect of an FTO's propensity to use force on a recruit's subsequent policing behavior.

5 Main Results

5.1 Field Training Officer Force and Subsequent Recruit Behavior

Next, we present results for the effect of training officers in Table 4. Each specification includes cohort year by division fixed effects, and standard errors are clustered at the recruit level. Our results are also robust to two-way clustering at the recruit badge and FTO level, as well as recruit and division-by-cohort year level.[19] The outcome variable for each column

[18] Here we cluster at the FTO level, however we are robust (i.e. only one coefficient in the Table is significant at conventional levels) to clustering at the division by cohort year level or simply estimating robust standard errors.

[19] Namely, column 1 is significant at the 5 percent level and columns 2 & 3 are significant at the 1 percent level.

is the proportion of 911 calls that end in force. Column 1 presents our baseline specification where the coefficient on $force_c$ captures the difference between recruit use of force for recruits assigned to a FTO with one standard deviation higher force propensity. Our results show that recruits with FTOs that use force one standard deviation more are 0.0131 percentage points or 12 percent more likely to use force.[20]

In column 2, we add controls for recruit characteristics (age, gender, race). Given our conditional random assignment and the results in Table 3 and Figure 4, we would not expect recruit characteristics to alter our estimate meaningfully. Column 2 shows that recruits with FTOs that use force 1 standard deviation more are 0.0140 percentage points or 13 percent more likely to use force. In Column 3, we add controls for FTO characteristics (age, gender, race) call characteristics (the same used to predict use of force). Even if recruits are indeed randomly assigned to FTOs, it is possible that the inclusion of call controls could alter our treatment effect. This is because assignment to a high force FTO could cause recruits to work in areas where calls are more severe to begin with. If this story were accurate, then adding these controls would cause our estimate to shrink. However, the estimates in Column 3 are remarkably similar to the ones presented in Column 2. Our preferred estimate shows that a one standard deviation increase in FTO use of force increases recruit force by 0.0125 percentage point or 12 percent. These findings are consistent with the story that our results are driven by force changes, not the type of calls recruits respond to.

We also assess whether our results are driven by recruits with higher force FTOs responding to calls that are ex-ante more likely to end in force by considering predicted force. To begin, we regress recruit use of force on cohort year and initial assignment fixed effect. We then regress these residuals on every covariate we observe for each call. These include the number of officers on the scene, beat, type of call (priority-by-type) year-by-month, hour, and day of the week. We use the resulting regression equation to predict the likelihood force

[20]In Table A0 we show results for alternative measure of force rates. Namely, our results are robust to using the unstandardized measure, a binary variable for above or below average force, inverse hyperbolic transformation of our shrunken force measure, and the unshrunk force measure.

would be used for each call. We then add the mean use of force rate for the full sample to each predicted value. Intuitively, this produces a linear combination of exogenous call characteristics, where the weights are chosen as to best predict the likelihood of force being used. Then, we consider whether FTO force is correlated with predicted recruit force. If recruits are responding to similar call across FTO type, predicted use of force should be similar across FTO force rates. The results of this exercise are shown in Figure 5(a). Observations are grouped so that each point includes an equal number of calls. There is no visible strong relationship between FTO effects and predicted use of force.[21] We also plot our actual use of force against our FTO effects in Figure 5(b). In contrast to our predicted results, here we see strong upward relationship, reiterating our findings in Table 4.

Next, we investigate whether being assigned a higher force FTO affects recruit arrests. These results are shown in Table 5. The layout of this table is similar to Table 4. Column 1 presents our baseline results, Column 2 adds officer characteristic controls, and Column 3 adds FTO and call controls. In all three columns our coefficients are similar in magnitude and significance. Our estimates show that one standard deviation harsher FTO does not impact recruit arrests. All coefficients are close to zero and are insignificant at conventional levels. For example, in Column 3, recruits with a one standard deviation higher force officer increase arrests by 0.0002 or about 0.6 percent. Using our column 3 estimate we can also rule out effects larger than a 3 percent increase in arrests.

It is also possible that by considering effects for any arrest, we are masking important heterogeneities. For example, perhaps recruits with higher force FTO make more serious (e.g. felony) arrests or arrest more minority civilians. To examine this, we estimate column 3 using different subgroups of arrests as an outcome.[22] Our results are shown in Figure 6. No estimate shown in Figure 6 is statistically different from zero, and none vary greatly from our original estimate. These results show that recruit with higher force FTOs do not make

[21] Coefficient on linear fit is 0.000043 with a p-value of 0.349 . The correlation coefficient for the two terms is 0.0096.

[22] For example, in the "Arrestee White" specification our outcome variable is 1 if the recruit arrests a White civilian and zero otherwise.

more arrests of any type.

Our results for recruit force and arrests indicate that high force FTOs are an essential determinant of recruit use of force up to a year after training. However, high force FTOs do not lead to increases in arrests. These results are consistent with recruits with high-force FTOs policing more aggressively, but not necessarily engaging in more active forms of policing. For example, consider a scenario where recruits are more likely to proactively engage with civilians on calls for service or more likely to be sent to dangerous calls after being trained by a high force FTO. If this story were true, then we should estimate an increase in force and arrests. However, our results are consistent with the story that high-force recruits are just as likely to engage with citizens or go to particularly dangerous calls, but are more likely to behave aggressively on those calls.

Finally, we explore whether certain recruits are particularly susceptible to adopting the force behavior of their FTO. To do so, we examine the effect of higher-force FTOs on force by recruit race, gender, age and whether fielding training officer and recruit are the same race. Our results are show in Figure 7. Here we report the coefficient on the indicator variable for high-force FTOs. All coefficients are greater than zero and although there is some variation across subgroups, the main takeaway of the graph is that all subgroups may experience an increase in force. No estimates are statistically different from our full sample effect. From a policy perspective this is important because it shows that many different types of recruits could be influenced by reform targeting FTOs.

5.2 Persistent Effects

Understanding the importance of field training officers also requires considering how long our treatment effect persists. In our setting this is particularly possible given Holz et al. (2020) documents only short-term effect for a somewhat similar treatment (peer injury). To accurately assess the effects of fielding training officers over time, we first need to restrict our sample. Namely, we restrict to the set of officers that we can observe three years after the

end of training. This is because some officers that are hired in recent cohorts are truncated, meaning that we can only observe them for less than three years because our data ends in 2019. We also drop officers that attrit, but this number (3 percent or 6 officers) of the early cohorts is very small. Our results are also robust to the inclusion of attriters.[23]

First, we present our main results from our new three-year sample in Table 6. The main takeaway from this table is that our estimates for all three columns are similar in magnitude to Table 4, but a bit larger, and are significant at conventional levels. This give us confidence that our results are not driven by compositional change and that it is sensible to consider the effect of field training officers over time in this sample.

To consider how the effects evolve over time we estimate a model of the following form

$$force_c = \theta_r + \Sigma_{t=0}^{5} \beta_t \Lambda_{o,r} biannual_t + \beta_2 X_c + \epsilon_{r,d,c} \tag{4}$$

where *biannual* is a binary variable that takes on a value of 1 t 6 month periods after the end of training. We also add separate fixed effects for *biannual*. All other terms are unchanged from Equation 1 and column 3 controls are included (i.e. calls, recruit and field training officer characteristics). Results are shown in Figure 8. Every coefficient is greater than zero, and there is no evidence that effects fade over time. Rather, increases in force appear to persist at least 3 years after the end of training.

5.3 Randomization Inference

In this section, we provide a robustness test focusing on the main results of the empirical estimates, i.e. columns 1-3 of Table 4. In our main estimates, we follow standard approaches by clustering our standard errors at the recruit level and note that we are also robust to two-way clustering at the recruit and division by year level (Bertrand et al., 2004). The concern motivating the robustness test in this section is that our outcome variable (force by call) is a

[23]Coefficients are 0.0001518, 0.0001856 , 0.0001913 for columns, 1,2 and 3 if the 6 officers are not dropped.

rare event representing only 0.106 percent of calls in our sample. In cases where an outcome variable is a rare event, standard asymptotic assumptions related to the distribution of point estimates and associated standard errors may be inappropriate. Here, we use randomization inference to construct an empirical distribution of point estimates and reassess the validity of the hypothesis test conducted for our primary set of estimates.

As discussed in Efron (2004), randomization inference is most appropriate to non-experimental settings when researchers are able to replicate the data generating process of the observed data. In our institutional setting, recruits from a given Academy cohort are randomly assigned to FTOs by command staff within their respective division. As discussed previously, our balancing tests support that this source of variation is as-good-as random. Thus, our randomization procedure attempts to replicate this variation in constructing an empirical distribution of point estimates and associated standard errors. For each recruit in our sample, we randomly draw an FTO from the set of eligible officers we observe working in the recruit's respective division. [24] Next, we construct an estimate of FTO propensity to use force using calls from the period prior to being assigned that particular recruit and shrink that estimate using an empirical Bayes procedure (see the full description in the methods section). Finally, we standardize FTO force by subtracting the mean and dividing by the standard deviation before estimating the model from columns 1-3 of Table 4. In order to obtain p-values for a two-sided hypothesis test, we replicated this procedure 1,000 times and calculate the share of these simulations where the t-statistic exceeds the absolute value of the t-statistics from Table 4.

For the models corresponding to columns 1 and 2 of table 4, we obtain p-values for a two-sided hypothesis test using randomization inference of 0.024 and 0.018 respectively. For illustrative purposes, we also plot the distribution of t-statistics obtained from our randomization procedure in figure 9 corresponding to our preferred specification from column 3 of

[24] We consider an officer as eligible for being a given recruit's FTO if they are observed working in the same division within 30 days of the recruit's first day assigned to patrol. We also only consider officers as eligible for being an FTO if they have a rank of police officer or higher though we note that we are robust to imposing a more stringent rank requirement of senior corporal or above.

table 4. In our randomization procedure, we find that none of the simulations result in a t statistic above 4.34. Thus, the associated p-value obtained through randomization inference is less than 1/1000. We interpret these results as providing additional evidence suggesting that our main results are not driven by potential inference issues associated with a rare outcome.

The results of the policy simulation are shown in figure 10 where each marker represents a 95 percent confidence interval around the estimated probability of any given recruit in our sample using force in any given month in the three years subsequent to field training. Each of the thirteen policy simulations, represented by the labels on the x-axis, were conducted by replacing FTOs from a given point in the FTO force distribution with a random draw of eligible FTOs from the lowest quartile of the distribution. As discussed, the predicted mean and standard error were computed by conducting 100 bootstrapped simulations so that we are were to obtain a confidence interval around each of the estimates. As shown below, there is a sharp decline in the probability of force when the simulation focuses on FTOs from points above the 90th percentile in aggregate force propensity. These estimates represent an approximate decline of 6 percent in recruit force (i.e. about 0.1 fewer force incidents per officer or 1.24 fewer force incidents per month). Although our simulation does not seem to indicate that there is any marginal benefit to replacing FTOs from below the 90th percentile, we note that our approach here is extremely conservative because it does not consider dynamic effects. In particular, we do not account for the fact that less forceful FTOs lead to recruits being less forceful and those recruits then become FTOs themselves. Rather than including these dynamic effects, we simply take the set of eligible FTOs and their respective force rates as fixed when we substitute high for low force FTOs so long as there is an alternative eligible FTO available for a given recruit. Thus, we interpret these estimates as an absolute lower-bound on the policy impact because these dyanmic effects could be quite large. It is worth noting that the policy proposed in this section is effectively a costless intervention because we have simply swapped high for low-force FTOs

whenever possible. Therefore, view these lower-bound estimates as indicative of an effective and extremely practical tool for policing administrators seeking to reduce aggregate rates of force.

6 Conclusion

This paper estimates the effects of high force field training officers on recruit use of force. To do so, we compared recruits quasi-randomly assigned field training officers with higher historical force rates to those quasi-randomly assigned lower force field training officers. Our results show field training officers are an important determinant of subsequent recruit force; a one standard deviation increase in field training officer force increases recruit force by 12 percent. Field training officer effects also persist three years after the completion of training. We find no evidence of increases in arrests. Given the broad support for reforms to police officer training, the wide availability of alternative field training officers, and the relative ease of switching field training officers, we believe our findings are consistent with field training officers being a viable avenue for reducing aggressive policing.

References

Anelli, M. and G. Peri (2019). The effects of high school peers' gender on college major, college performance and income. *The Economic Journal 129*(618), 553–602.

Ba, B. A., D. Knox, J. Mummolo, and R. Rivera (2021). The role of officer race and gender in police-civilian interactions in Chicago. *Science 371*(6530), 696–702.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics 119*(1), 249–275.

Bertrand, M. and A. Schoar (2003). Managing with style: The effect of managers on firm policies. *The Quarterly journal of economics 118*(4), 1169–1208.

Bifulco, R., J. M. Fletcher, and S. L. Ross (2011). The effect of classmate characteristics on post-secondary outcomes: Evidence from the add health. *American Economic Journal: Economic Policy 3*(1), 25–53.

Billings, S. B., D. J. Deming, and S. L. Ross (2019). Partners in crime. *American Economic Journal: Applied Economics 11*(1), 126–150.

Billings, S. B. and K. T. Schnepel (2020). Hanging out with the usual suspects: Neighborhood peer effects and recidivism. *Journal of Human Resources*.

Bloom, N. and J. Van Reenen (2007). Measuring and explaining management practices across firms and countries. *The quarterly journal of Economics 122*(4), 1351–1408.

Bruinsma, G. and D. Weisburd (2014). *Encyclopedia of criminology and criminal justice*. Springer Reference.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica 81*(3), 855–882.

Chan, J. B. (1997). *Changing police culture: Policing in a multicultural society*. Cambridge University Press.

Crank, J. P. (2004). Understanding police culture.

Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis.

Journal of the American Statistical Association 99(465), 96–104.

Fenizia, A. (2021). Managers and productivity in the public sector. *Working Paper*.

Fryer Jr, R. G. (2019). An empirical analysis of racial differences in police use of force. *Journal of Political Economy 127*(3), 1210–1261.

Gallup Panel (2020). Americans' Views of the Need for Changes in Policing. `https://news.gallup.com/poll/315962/americans-say-policing-needs-major-changes.aspx`.

Getty, R. M., J. L. Worrall, and R. G. Morris (2014). How far from the tree does the apple fall? Field training officers, their trainees, and allegations of misconduct. *Crime & Delinquency 62*(6), 821–839.

Giorcelli, M. (2019). The long-term effects of management and technology transfers. *American Economic Review 109*(1), 121–52.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (1996). Crime and social interactions. *Quarterly Journal of Economics 111*(2), 507–548.

Hoekstra, M. and C. Sloan (2020). Does race matter for police use of force? evidence from 911 calls. *American Economic Review forthcoming*.

Holz, J. E., R. G. Rivera, and B. A. Ba (2020). Peer effects in police use of force. *Working Paper*.

Johnson, H., S. Ross, S. Mello, and M. Ross (2021). Experience, formalized training, and police discrimination. *Working Paper*.

Lazear, E. P., K. L. Shaw, and C. T. Stanton (2015). The value of bosses. *Journal of Labor Economics 33*(4), 823–861.

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–45.

McLean, K., S. E. Wolfe, J. Rojek, G. P. Alpert, and M. R. Smith (2020). Randomized controlled trial of social interaction police training. *Criminology & Public Policy 19*(3), 805–832.

Murphy, F. X. (2019). Does increased exposure to peers with adverse characteristics reduce workplace performance? Evidence from a natural experiment in the US army. *Journal of*

Labor Economics 37(2), 435–466.

Owens, E., D. Weisburd, K. L. Amendola, and G. P. Alpert (2018). Can you build a better cop? experimental evidence on supervision, training, and policing in the community. *Criminology & Public Policy 17*(1), 41–87.

Paoline III, E. (2006). The myth of a monolithic police culture. *Demystifying crime and criminal justice*, 81–88.

Paoline III, E. A. (2003). Taking stock: Toward a richer understanding of police culture. *Journal of criminal justice 31*(3), 199–214.

Paoline III, E. A. and J. M. Gau (2018). Police occupational culture: Testing the monolithic model. *Justice quarterly 35*(4), 670–698.

Pew Research Center (2020). Majority of public favors giving civilians the power to sue police officers for misconduct. `https://www.pewresearch.org/politics/2020/07/09/majority-of-public-favors-giving-civilians-the-power-to-sue-police-officers-for-misco`

President's Task Force on 21st Century Policing (2015). Final report of the president's task force on 21st century policing.

Rakich, N. (2020). How americans feel about 'defunding the police'. `https://fivethirtyeight.com/features/americans-like-the-ideas-behind-defunding-the-police-more-than-the-slogan-itself/`.

Rosenbaum, D. P. and D. S. Lawrence (2017). Teaching procedural justice and communication skills during police–community encounters: Results of a randomized control trial with police recruits. *Journal of Experimental Criminology 13*(3), 293–319.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics 116*(2), 681–704.

Schaefer, B. and T. Hughes (2016). Honing interpersonal necessary tactics (hint): An evaluation of procedural justice training. *Louisville, KY: Southern Police Institute, University of Louisville*.

Skogan, W. G., M. Van Craen, and C. Hennessy (2015). Training police for procedural

justice. *Journal of experimental criminology 11*(3), 319–334.

Skolnick, J. H. (1966). *Justice without trial: Law enforcement in democratic society*.

Warners, Ronald (2020). The field training experience: Perspectives of field training officers and trainees. https://www.policechiefmagazine.org/the-field-training-experience-perspectives-of-field-training-officers-and-trainees/.

Weisburst, E. K. (2019). Police use of force as an extension of arrests: Examining disparities across civilian and officer race. In *AEA Papers and Proceedings*, Volume 109, pp. 152–56.

West, J. (2019). Learning by doing in law enforcement. *Working Paper*.

Westley, W. A. (1970). *Violence and the police: A sociological study of law, custom, and morality*, Volume 28. MIT press Cambridge, MA.

Wheller, L., P. Quinton, A. Fildes, and A. Mills (2013). The greater manchester police procedural justice training experiment. *Coventry, UK: College of Policing*.

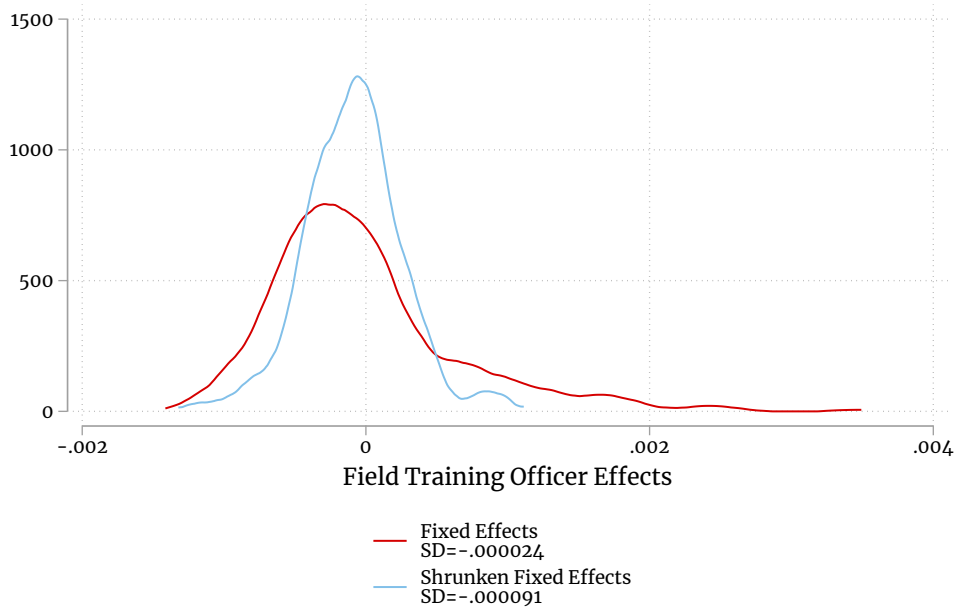
Whitmore, D. (2005). Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review 95*(2), 199–203.

Woody, R. H. (2005). The police culture: Research implications for psychological services. *Professional Psychology: Research and Practice 36*(5), 525.

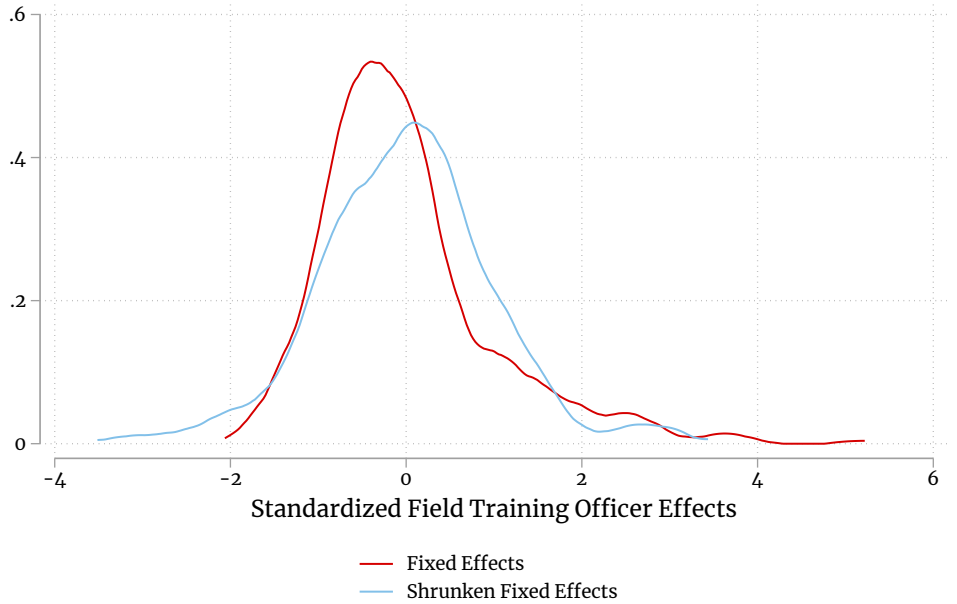
Figures

Figure 1: Density of Field Training Officer Propensity to Use Force

(a) Field Training Officer Effects

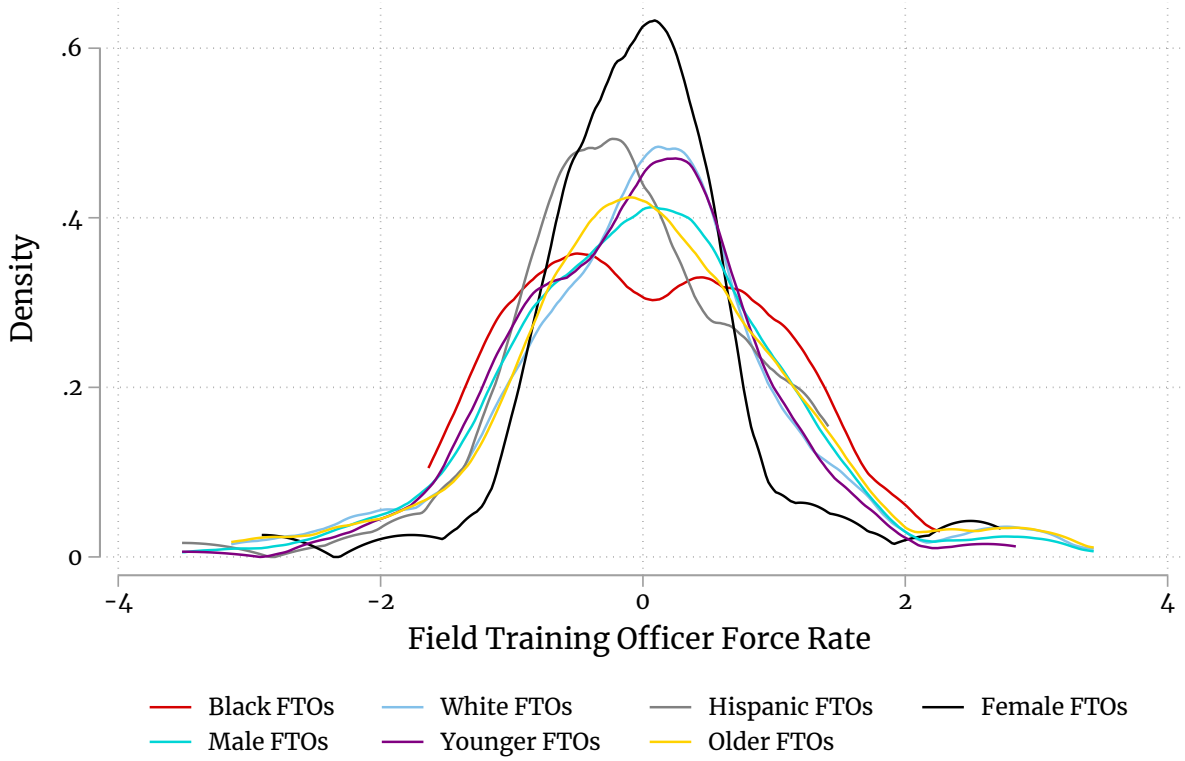


(b) Standardized Field Training Officer Effects



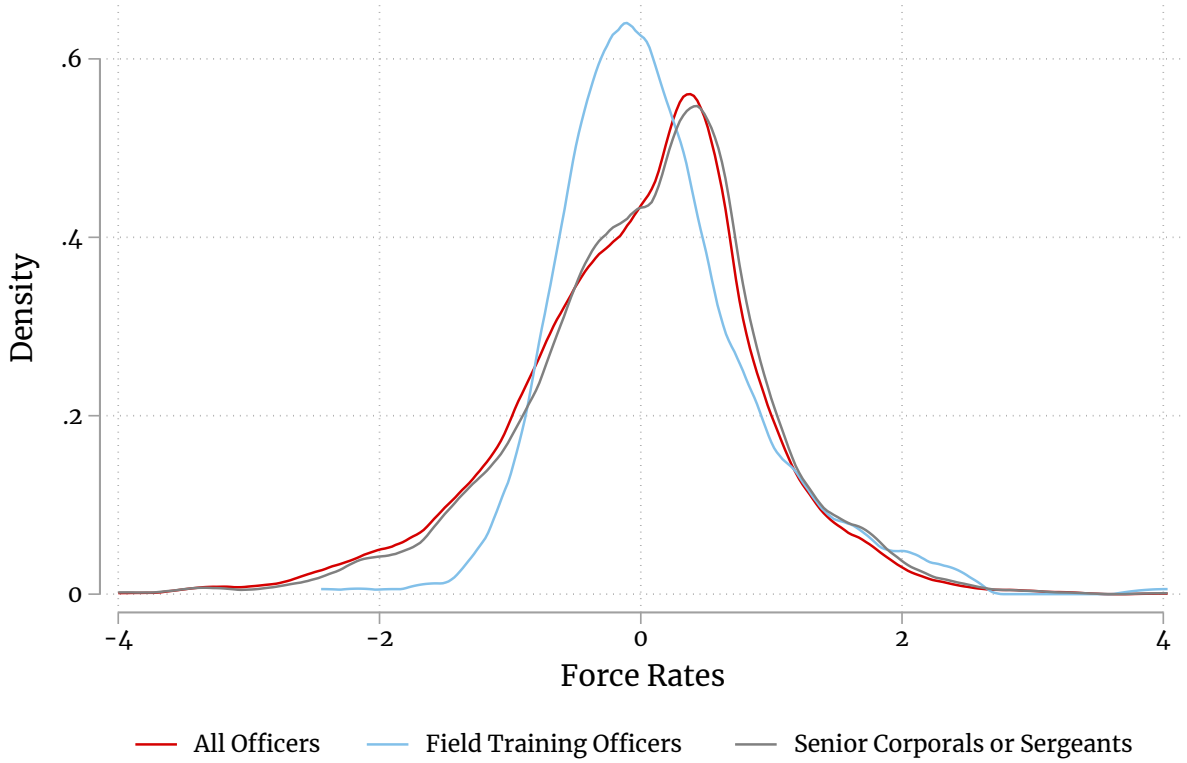
Notes: These figures plot the distribution of field training officer fixed effects. Fixed effects are calculated after accounting for the number of officers on the scene, beat, type of call (priority-by-type) year-by-month, hour, and day of the week fixed effects.

Figure 2: Density of Field Training Officer Propensity to Use Force by Field Training Officer Characteristics



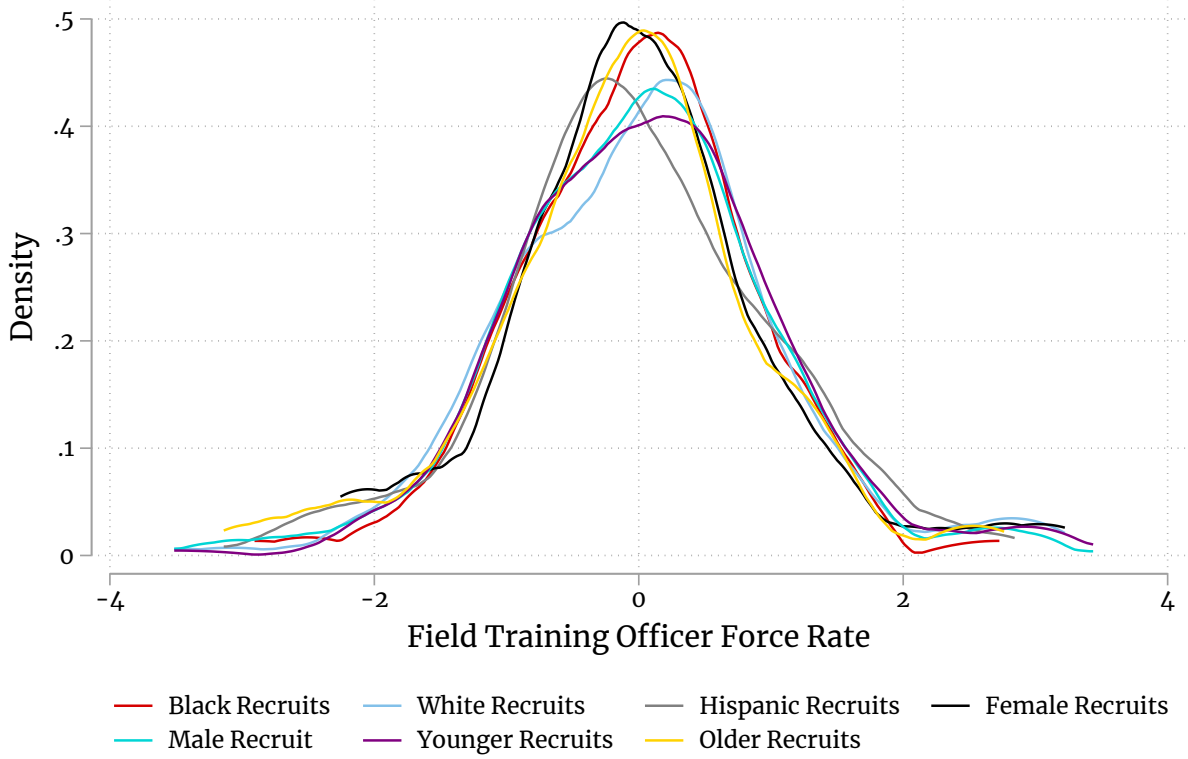
Notes: This figure plots the distribution of field training officer effects by field training officer characteristics.

Figure 3: Density of Officer Propensity to Use Force for All Officers



Notes: This figure plots the distribution of police officer effects in the full sample of calls. Senior Corporals and Sergeants are the most frequent rank of field training officers.

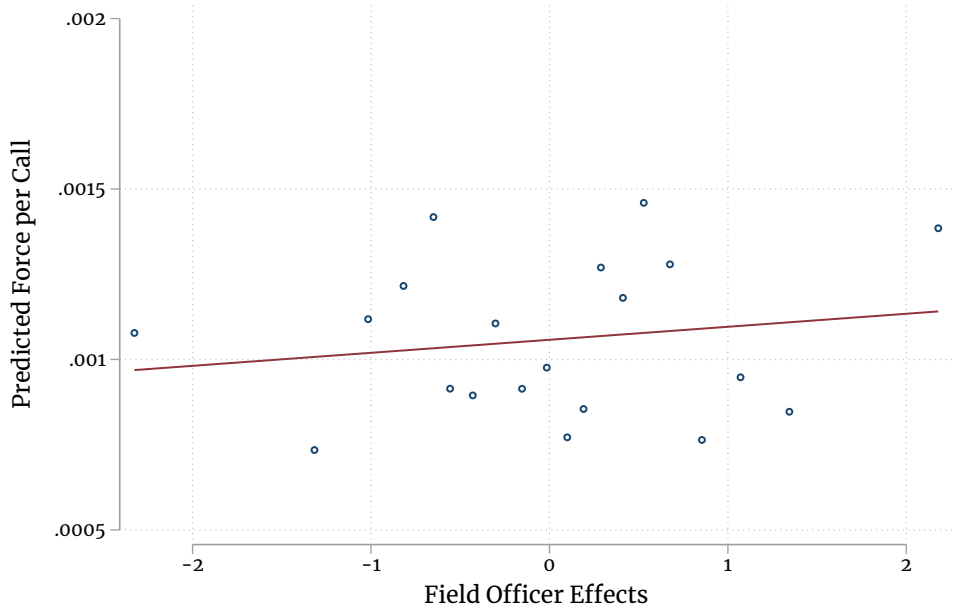
Figure 4: Density of Field Training Officer Propensity to Use Force by Recruit Characteristics



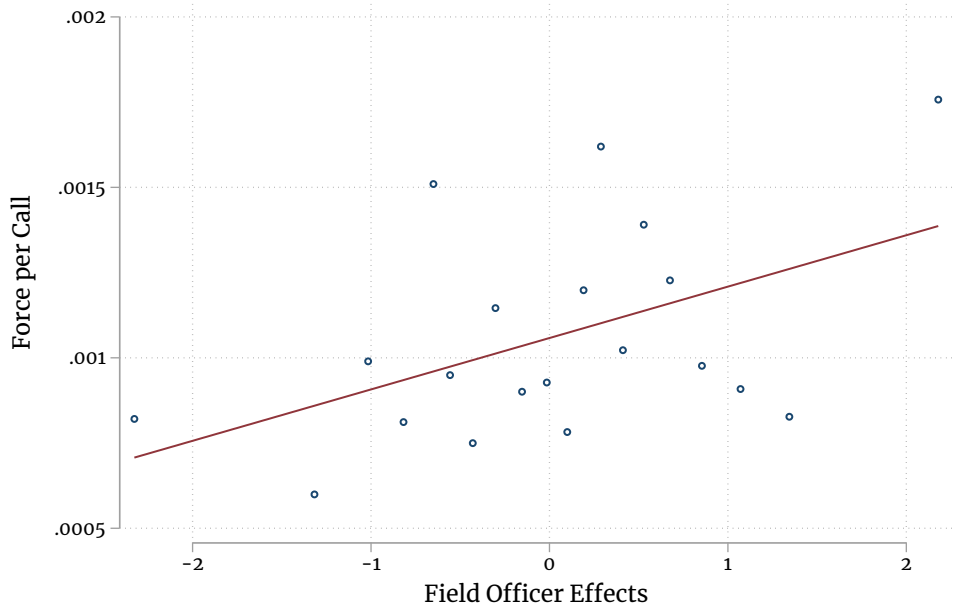
Notes: This figure plots the distribution of field training officer effects by recruit characteristics.

Figure 5: Recruit Predicted Force and Actual Force by Field Training Officer Effects

(a) Predicted Use of Force

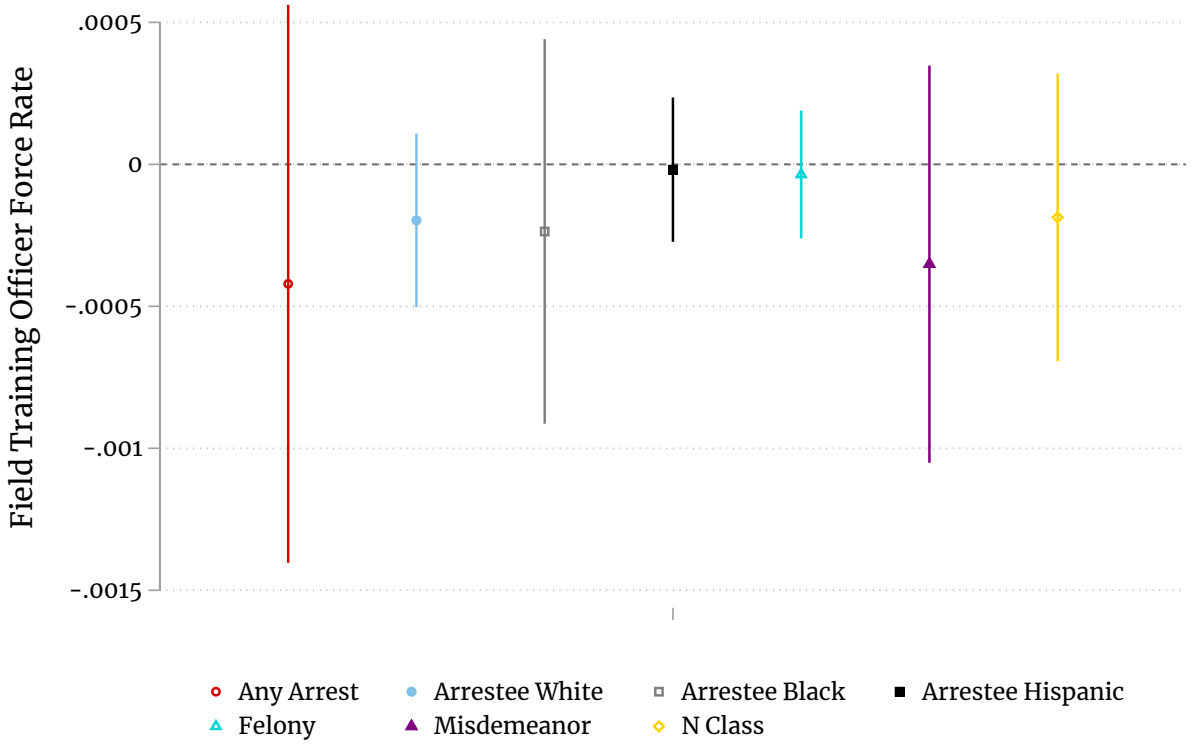


(b) Use of Force



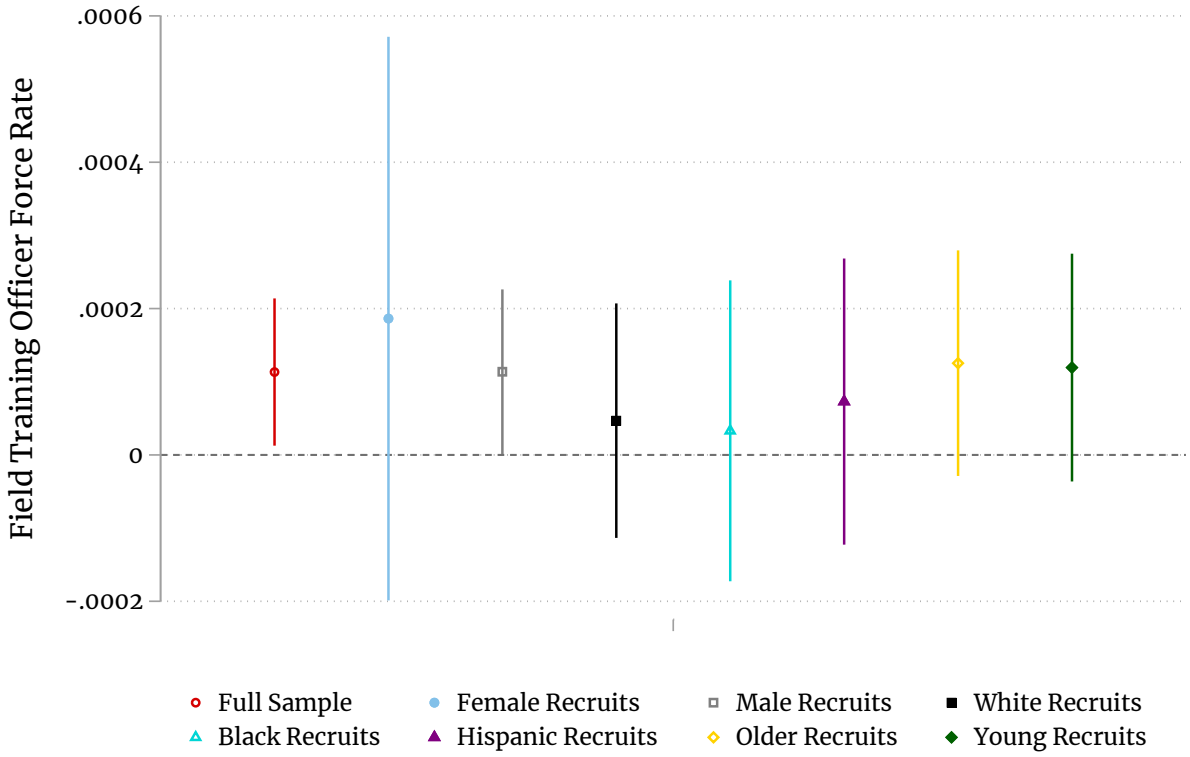
Notes: In Panel (a) we plot use of force. In Panel (b) we plot predicted use of force. The fitted line is a linear fit across all use of force rates. Observations are grouped so that each point includes an equal number of observations. Use of force is predicted using the number of officers on the scene, beat, type of call (priority-by-type) year-by-month, hour, and day of the week fixed effect.

Figure 6: The Effect of Field Training Officers on Different Types of Arrests



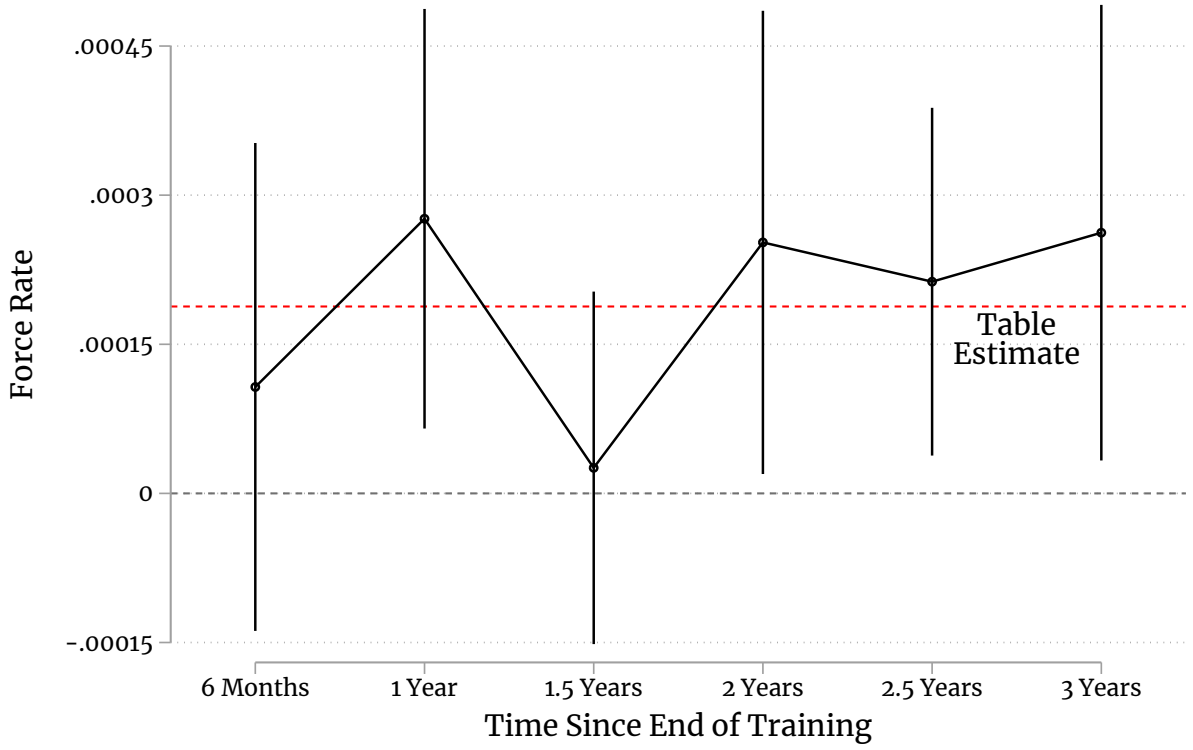
Notes: This figure reports the effect of field training officer on different types of arrests. Each coefficient is from a separate regression. Standard errors are clustered at the recruit level.

Figure 7: The Effect of Field Training Officers on Force by Recruit Subgroups



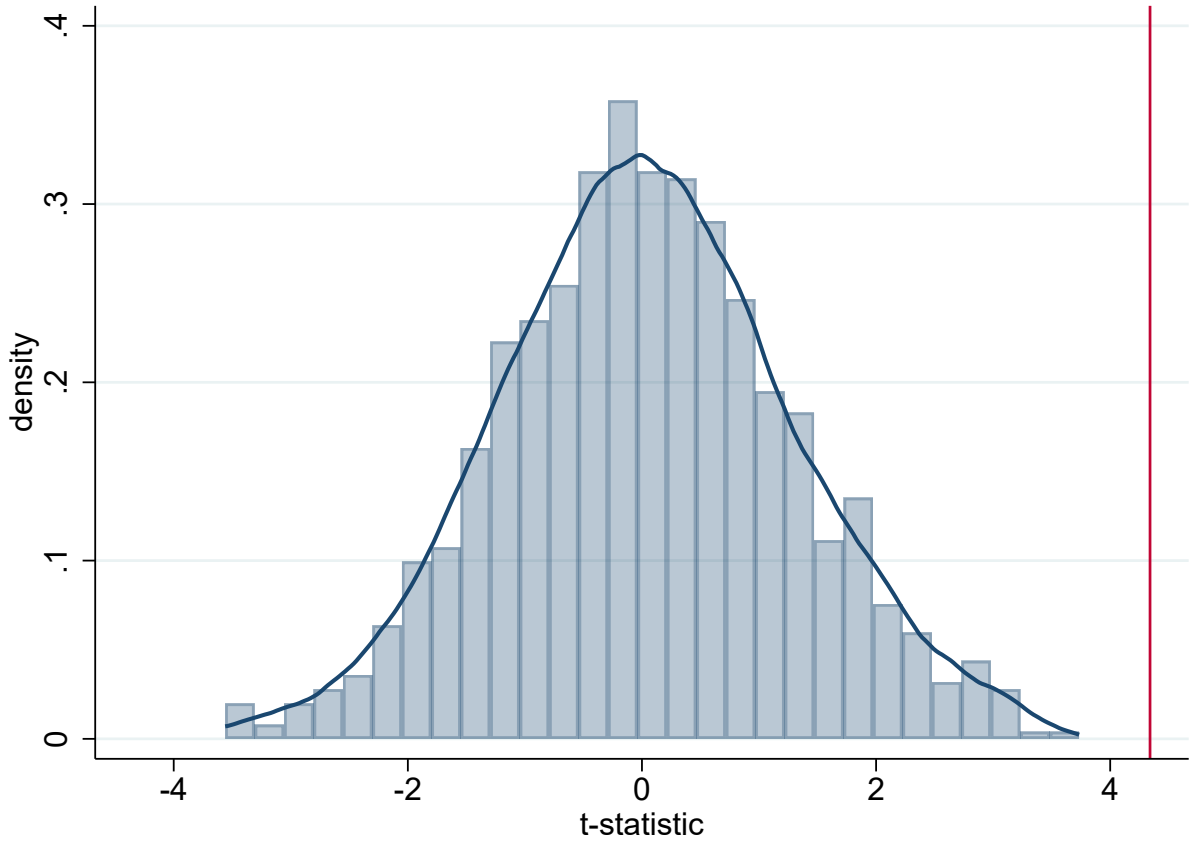
Notes: This figure reports the effect of field training officer force rates by recruit subgroup. Each coefficient is from a separate regression. Standard errors are clustered at the recruit level.

Figure 8: The Effect of Field Training Officers on Force Over Time



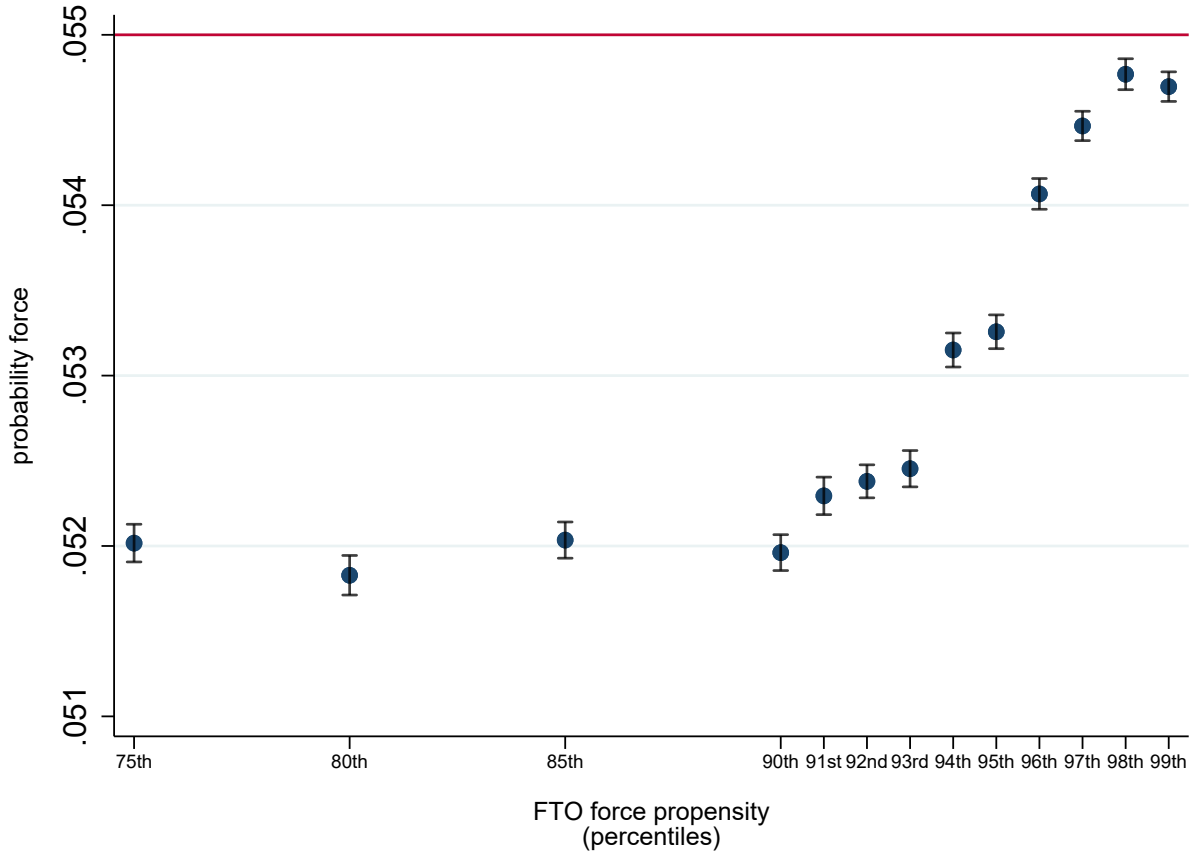
Notes: This figure reports the effect of field training officer force rates over time. Sample includes all officers observed with three full years of data. The red line marks the estimate from column 3 in Table 6. Standard errors are clustered at the recruit level.

Figure 9: Empirical Distribution of T-Statistics from Randomization Inference



Notes: This figure reports the empirical distribution of t-statistics from estimating the regression shown in column 3 of Table 4 by using a 1,000 randomized simulations of the data generating process. None of the simulations result in a t-statistic greater than our baseline estimate of 4.34 indicating a p-value of less than 1/1000.

Figure 10: Policy Simulation Replacing High with Low-Force FTOs



Notes: This figure plots the mean and 95% confidence interval for 13 policy simulations involving 100 bootstrapped draws where high-force FTOs are replaced with low-force FTOs. The probabilities on the y-axis are calculated per recruit in any given month in the period following field training.

Tables

Table 1: Officer-level Summary Statistics

	(1) Entire Sample	(2) High Force Trainer	(3) Low Force Trainer
White	0.438	0.478	0.413
	(0.497)	(0.501)	(0.493)
Black	0.207	0.197	0.213
	(0.406)	(0.399)	(0.410)
Hispanic	0.302	0.274	0.319
	(0.460)	(0.447)	(0.467)
Female	0.178	0.153	0.193
	(0.383)	(0.361)	(0.395)
Age	35.84	35.37	36.13
	(5.466)	(5.608)	(5.366)
Observations	411	157	254

Table 2: Call Level Summary Statistics

	(1) Entire Sample	(2) High Force Trainer	(3) Low Force Trainer
Force	0.00106	0.00120	0.000970
	(0.0325)	(0.0347)	(0.0311)
Arrest	0.0348	0.0345	0.0349
	(0.183)	(0.183)	(0.184)
Priority	2.520	2.513	2.524
	(0.828)	(0.831)	(0.826)
Observations	1085020	405679	679341

Table 3: Balance Test: Correlation between Recruit and Field Training Officer Characteristics

	Field Training Officer Characteristics						
	(1) Age	(2) Female	(3) White	(4) Black	(5) Hispanic	(6) Hire Date	(7) Force Rate
Recruit Characteristics							
White	-1.0720	-0.1124	-0.1555	-0.0052	0.0952	0.7714	0.1537
	(1.7427)	(0.0845)	(0.0997)	(0.0688)	(0.0705)	(1.9804)	(0.2887)
Black	-1.4251	-0.0514	-0.0800	0.0090	0.0061	0.9156	0.0504
	(1.9739)	(0.0939)	(0.1163)	(0.0951)	(0.0801)	(2.2187)	(0.2987)
Hispanic	-1.3049	-0.0561	-0.1512	0.0158	0.0730	1.1466	0.1082
	(1.8763)	(0.0874)	(0.1094)	(0.0921)	(0.0806)	(2.0828)	(0.2987)
Female	-0.1773	0.0293	-0.0923	0.0502	0.0560	0.7939	0.0084
	(0.9826)	(0.0433)	(0.0639)	(0.0524)	(0.0538)	(0.9008)	(0.1214)
Age	0.0502	-0.0000	0.0027	0.0003	0.0004	0.0293	-0.0193**
	(0.0798)	(0.0033)	(0.0049)	(0.0038)	(0.0042)	(0.0801)	(0.0093)
Observations	411	411	411	411	411	411	411
Div. by Cohort FE	Y	Y	Y	Y	Y	Y	Y
Outcome Mean	49.45	0.129	0.628	0.170	0.165	41.91	3.12e-10
F-Test P-Value	0.961	0.515	0.338	0.928	0.389	0.877	0.361

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: Standard errors are clustered at the field training officer level.

Table 4: The Effect of High Force Field Training Officers on Recruit Use of Force

	(1) Force	(2) Force	(3) Force
Field Training Officer Force Rate	0.000131**	0.000140***	0.000125***
	(0.0000533)	(0.0000529)	(0.0000288)
Observations	1085020	1085020	1085020
Outcome Mean	0.00106	0.00106	0.00106
Assigned Div by Cohort FE	Y	Y	Y
Recruit Characteristics	-	Y	Y
Call Controls	-	-	Y

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: Standard errors are clustered at the field training officer level. Column 2 adds controls for recruit characteristics (age, gender, race). We add controls for FTO characteristics (age, gender, race) and call characteristics fixed effects (number of officers on the scene, beat, type of call—priority-by-type,year-by-month, hour, and day of the week).

Table 5: The Effect of High Force Field Training Officers on Recruit Arrests

	(1) Arrest per Call	(2) Arrest per Call	(3) Arrest per Call
Field Training Officer Force Rate	0.000255	0.000403	0.000200
	(0.000678)	(0.000515)	(0.000462)
Observations	1085020	1085020	1085020
Outcome Mean	0.0348	0.0348	0.0348
Assigned Div by Cohort FE	Y	Y	Y
Recruit Characteristics	-	Y	Y
Call Controls	-	-	Y

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: Standard errors are clustered at the field training officer level. Column 2 adds controls for recruit characteristics (age, gender, race). We add controls for FTO characteristics (age, gender, race) and call characteristics fixed effects (number of officers on the scene, beat, type of call—priority-by-type,year-by-month, hour, and day of the week).

Table 6: Three Year Sample: The Effect Field Training Officers on Recruit Force

	(1) Force	(2) Force	(3) Force
Field Training Officer Force Rate	0.000146*	0.000164**	0.000188***
	(0.0000778)	(0.0000738)	(0.0000460)
Observations	474297	474297	474297
Outcome Mean	0.00102	0.00102	0.00102
Assigned Div by Cohort FE	Y	Y	Y
Recruit Characteristics	-	Y	Y
Call Controls	-	-	Y

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: Standard errors are clustered at the field training officer level. Column 2 adds controls for recruit characteristics (age, gender, race). We add controls for FTO characteristics (age, gender, race) and call characteristics fixed effects (number of officers on the scene, beat, type of call—priority-by-type,year-by-month, hour, and day of the week).

Appendix

Table A0: Robustness: The Effect of High Force Field Training Officers on Recruit Use of Force – Different Force Measures

	(1) Force	(2) Force	(3) Force	(4) Force	(5) Force	(6) Force	(7) Force	(8) Force	(9) Force	(10) Force	(11) Force	(12) Force
Theta Shrunk	0.369**	0.393***	0.353***									
	(0.150)	(0.149)	(0.0809)									
Above Avg Theta Shrunk				0.000272**	0.000251**	0.000164**						
				(0.000123)	(0.000127)	(0.0000798)						
IHS Theta Shrunk							0.369**	0.393***	0.353***			
							(0.150)	(0.149)	(0.0809)			
Unshrunk Force Rate										0.150*	0.164**	0.143***
										(0.0818)	(0.0823)	(0.0457)
Observations	1085020	1085020	1085020	1085020	1085020	1085020	1085020	1085020	1085020	1085020	1085020	1085020
Outcome Mean	0.00106	0.00106	0.00106	0.00106	0.00106	0.00106	0.00106	0.00106	0.00106	0.00106	0.00106	0.00106
Assigned Div by Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Recruit Characteristics	-	Y	Y	-	Y	Y	-	Y	Y	-	Y	Y
Call Controls	-	-	Y	-	-	Y	-	-	Y	-	-	Y

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$