The Effect of Field Training Officers on Police Use of Force

Chandon Adger, Matthew Ross, and CarlyWill Sloan *

October 9, 2023

Abstract

Over the past decade, police use of force has become an increasingly charged political issue with growing calls for reform. One of the few reforms where advocates and the policing community have reached a consensus is on the need for improved and expanded training. In this paper, we study an under-researched but nearly universal training approach whereby a recruit is paired with a senior officer during a phase referred to as "field training." In particular, we consider the link between a field training officer's prior propensity to use force 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 the universe of 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 (121 percent) leads to a 14 percent increase in their recruit's subsequent propensity to use force. The effect of having a more aggressive field training officer persists for as much as two and a half years after the recruit completes training.

*Correspondence: Adger, Texas A&M University, cadger@tamu.edu; Ross, Northeastern University, ma.ross@northeastern.edu; Sloan, United States Military at West Point, carlywill.sloan@westpoint.edu All opinions expressed in this manuscript are those of the authors and do not represent the opinions of the United States Military Academy (USMA), the United States Army, or the Department of Defense. Acknowledgments: We are grateful for useful comments from Bocar Ba, Shawn Bushway, Jesse Bruhn, Jamein Cunningham, Jennifer Doleac, Jeffery Grogger, Elizabeth Luh (twice!), John MacDonald, Steve Mello, Aurelie Ouss, Anne Piehl, and Steve Ross. We also thank participants at the 2023 AEA/ASSA Meeting, 2022 NBER Summer Institute, 2022 Al Capone Conference, 2022 Chicago/LSE Conference, 2021 Southern Economic Association Conference, and 2021 Association for Public Policy & Management Conference. Sloan and Ross are grateful to the Russell Sage Foundation (G-2011-30016) for financial support of our project "Does More Training Mitigate Disparities in Police Use of Force?". We are also especially grateful to Felipe Goncalves and Emily Weisburst for sharing DPD overtime and shift data with us. Finally, thank the the Dallas Police Department for numerous useful conversations, critical insights, and for allowing us to visit their facilities.

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). However, polls have found that nearly 90 percent of Americans support improved and expanded training in areas like de-escalation and avoiding violence (Ipsos, 2021). Policymakers and the policing community also frequently cite training as a key approach to reducing police violence. 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.

The small literature on the effectiveness of police training has historically relied primarily on surveys of officers about the impact of the training following participation in a particular course. Despite the perceived importance of training, there have been few empirical studies that consider the impact of police training on actual enforcement outcomes, such as use of force. Further, the few studies that have examined the impact of training on actual enforcement outcomes have exclusively focused on classroom-based training interventions occurring at the police Academy, or as a part of continuing education requirements (Dube et al., 2023; Johnson et al., 2021; McLean et al., 2020; Owens et al., 2018). Thus, there is little work on the effectiveness of experiential components of training, broadly referred to as "field training," on police behavior and enforcement outcomes. These 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 patrol officer 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.[1] 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 Holz et al. (2023); Rivera (2023); West (2019); Weisburst (2022) suggests that an officer's on-the-job-experiences are a significant factor in their subsequent enforcement behavior.

Despite the potential significance of supervisors, the economics literature has surprisingly overlooked their impact on workers. Additionally, much of the existing research on supervisors fails to explore the mechanisms behind their effects, largely due to challenges in obtaining data on supervisor-worker relationships and outcomes. This gap is remarkable, especially given the extensive research on peer effects, which highlights the influence of peers on decision-making in various contexts.[2] While peers certainly hold importance, most workplaces have hierarchical structures, suggesting that supervisors may have a greater impact than other colleagues. Therefore, it's crucial to thoroughly investigate how authority figures influence worker behavior. Although the existing literature on supervisor effects is limited (see Bennedsen et al., 2020; Bertrand and Schoar, 2003; Bloom and Van Reenen, 2007; Fenizia, 2022; Giorcelli, 2019; Hoffman and Tadelis, 2021; Lazear et al., 2015; Rim et al., 2022), our study is unique in examining public sector supervisors (similar to Fenizia, 2022) within a high-stakes worker decision-making environment.

In this paper, we ask whether recruits who are assigned to more aggressive field training officers (FTOs) are subsequently more likely to use force in the years following field training.

[1] Dube et al., 2023 stands out as a notable example, highlighting how a training program designed to encourage officers to thoroughly contemplate various perspectives when assessing the situations they confront results in reduced use of force and discretionary arrests.

[2] Much of this literature documents that social influence by peers shapes individual decisions in settings like schools (e.g., Anelli and Peri, 2019; Bifulco et al., 2011; Carrell et al., 2013; Sacerdote, 2001; Whitmore, 2005), work (Mas and Moretti, 2009), neighborhoods (Billings et al., 2019, Billings and Schnepel, 2020, Glaeser et al., 1996), and the military (Murphy, 2019).

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 has previously been more likely to use force, on a recruit's subsequent enforcement behavior.We characterize aggressive FTOs by constructing a measure of propensity to use force in the period prior to their being assigned a particular recruit. In constructing FTO force, we account for the fact that some FTOs may be assigned (or select into) more dangerous calls (or locations) by controlling for factors like neighborhood, characteristics of the call, and aspects related to date and time. Our primary measure captures an officer's propensity to use force above and beyond how their peers respond to a similar set of incidents. Thus, we interpret our measure of force as capturing an FTO's propensity to respond more aggressively to any given incident. However, our measure of force does not prima facie say anything about whether force was justified or necessary.[3]

In our preferred specification, we estimate that a one standard deviation increase in an FTO's prior propensity to use of force (1.485 in 1,000 or 121 percent of the sample mean) leads to a 14 percent increase in a recruit's subsequent likelihood of using force after training. Distinct from other types of training models that have been studied previously in the literature, we find that the effect of having a more forceful FTO is particularly long-lasting. Namely, we show effects persist for up to two and a half years following the culmination of training. We interpret our results as being consistent with the idea that having a more forceful FTO, results in a recruit who subsequently has a lower threshold for using force on any given call for service. We are also able to rule out several alternative mechanisms. In particular, we document that recruits assigned to a more forceful FTO are not involved

[3]Force incidents without an arrest are typically considered a proxy for potentially excessive force. Our main estimates are similar even when we restrict the dependent variable to be only incidents of force where an arrest wasn't made.

in more active forms of policing as captured through arrest rates, response times, and time spent on a given call. Further, we document that our findings are not driven by (1) recruits witnessing a force incident during field training, (2) the propensity of a recruit to subsequently answer a call for service with their FTO or another partner after training, (3) differences in reporting behavior, and (4) other FTO characteristics that are correlated with force. We also show that recruits assigned a higher force FTOs are more likely to issue less serious or unfiled arrests, but there is no change in felony or filed arrests. While Getty et al., 2014 find a descriptive correlation between the total volume of misconduct complaints filed against a recruit and their FTO, our quasi-experimental design identifies no such relationship with complaints alone or between force and complaints.[4] We argue that our results are generally consistent with the story that qualitatively less desirable use of force behavior is passed from high force FTOs to their recruits.

In settings close to our own, Holz et al. (2023) and Rivera (2023) examine peer effects related to a number of outcomes including use of force and arrest, respectively. We differ from Holz et al. (2023) by considering a fundamentally different determinant of force (i.e. force rates of supervisors). Relative to a peer getting injured on the job, we find that one's FTO has a substantially longer impact on subsequent policing behavior. Rivera (2023) documents long-term effects on arrests because of more minority peers while our focus on supervisors (FTOs) is unique and our results are particularly long-lasting. Namely, we document higher use of force for officers trained by more aggressive officers persisting for up to two and a half years following the culmination of their training. Relative to our finding of an elasticity of about 0.12, Holz et al. (2023) presents estimates that translate to an elasticity

[4]While Getty et al., 2014 also use data from the Dallas Police Department, they apply a hierarchical modeling approach that is largely descriptive in nature and does not distinguish between complaints received before, during, or after field training. Relative to Getty et al., 2014, we implement a more rigorous empirical design that allows us to obtain plausibly causal estimates to address an entirely different conceptual question, i.e. the relationship between an FTO's and a recruit's propensity to use force.

of approximately 0.9 while Rivera (2023) reports an elasticity of approximately 0.12.[5] While our findings complement the key takeaway from these papers that peers are particularly important in policing, our work emphasizes the importance of supervisors particularly during intensive periods of on-the-job training. [6]

Our main finding, that a recruit's subsequent use of force is shaped by their FTO, contributes to a broad literature that spans criminology to economics and emphasizes the characteristics of individual police officers in shaping police use of force. Criminologists were among the first to investigate heterogeneity within policing agencies (Bruinsma and Weisburd, 2014; Chan, 1997; Crank, 2004; Paoline III, 2006; Woody, 2005). Recent work in economics has focused on the role that an officer's race or their peer's race has on their propensity to use force (Ba et al., 2021, Fryer Jr, 2019, Hoekstra and Sloan, 2022, Rivera, 2023, Weisburst, 2019). Our findings also contribute to a broader literature in criminology about the impact of policing culture. Drawing on an established literature (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) state that culture is key to reducing aggressive policing. From a policy perspective, our findings represent the first rigorous empirical evidence supporting interventions aimed at reducing force by changing police culture, particularly through field training. Further, our study suggests that characteristics like aggression or prior enforcement behavior, which are typically not documented or

[5]Holz et al. (2023) reports an effect size of 9.78 percent from 95.16 percent change in the independent variable which translates to an elasticity of 0.093. Rivera (2023) reports an effect size of 6.56 percentage points or 45.87 percent relative to a dependent mean of 0.143 from a 1 standard deviation or 378.29 percent change in the independent variable which translates to an elasticity of 0.12

[6]Relative to seminal papers in the peer effects literature by Mas and Moretti (2009) and Falk and Ichino (2006), which document elasticities of 0.15 and 0.14 respectively, we document a comparable estimate of about 0.12.[7]

explicitly tracked by policing agencies, are key sources of officer heterogeneity and important determinants of aggregate rates of force.

Given the fact that classroom-based interventions (McLean et al., 2020; Owens et al., 2018; Wheller et al., 2013) have lead to meaningful but short-lived effects on enforcement behavior, our findings suggest that field training based interventions might be more effective at reducing long-term rates of force.[8] Reforms aimed at changing the composition of FTOs are also much more practical and politically palatable relative to large-scale initiatives such as defunding the police, dramatically reducing enforcement, or dramatically altering the demographic composition of the police force. Our paper suggests that a relatively low-cost policy to reduce aggregate force might be to simply select FTOs based on their prior propensity to use force. By relieving the top tercile of FTOs from their training duties, we estimate that the aggregate number of force incidents involving the 411 recruits in our sample would fall by 12.01 to 9.83 percent or from 2,731 to between 2403.20 and 2462.73 force incidents.[9] However, we note that these simple back-of-the-envelope estimates are likely too conservative because they capture only the direct effect of the reduction in FTO force on the first generation of recruits. Not captured in these estimates is the indirect effect of less forceful recruits themselves becoming FTOs and influencing future cohorts. Future work might also consider the broader impacts of changes to policing culture by leveraging

[8]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, and impartiality in decision-making (President's Task Force on 21st Century Policing, 2015). There is also a body of literature that 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.

[9]We arrive at these estimates by conducting 1,000 simulations where we replace FTOs from the highest tercile of forcefulness with those from the lowest tercile. We predict each recruit's use of force in the period after training by apportioning it using the coefficient (0.00017) from specification 3 of Table 2 multiplied by the difference between their real and simulated FTO force relative to the global mean (0.00123). In the text, we provide a range of force reduction based on a 95 percent confidence interval using the mean (2432.96) and standard error (15.19) from our policy simulation.

new approaches like marginal value of public funds (Hendren and Sprung-Keyser, 2020) to calculate aggregate welfare implications. A proper cost-benefit analysis of policies that reduce police force would account for a wide variety of impacts like officer and civilian safety, value of statistical life, deterrence, police legitimacy (i.e. community trust), and civil liability.

2 Police Officer Training and Institutional Background

According to a 2018 survey of 681 state and local law enforcement agencies, police recruits spent an average of 833 hours in basic training (at the Academy) and 508 hours in field training (Bureau of Justice Statistics, 2018). Further, the field training model used in Dallas (the San Jose Model) is the most commonly used training program in the US (President's Task Force on 21st Century Policing, 2015). Thus, we believe that the findings from our study are broadly generalizable because the training received by recruits in Dallas is representative of how the vast majority of policing agencies across the country train new officers. However, we note that recruits in Dallas tend to receive more training (both at the Academy and in the field) than the average agency and that the requirements exceed those mandated by the State of Texas. Like most agencies, however, training in Dallas is divided into two distinct phases before recruits become full-fledged police officers, i.e., Academy training (phase 1) and field training (phase 2).[10] 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.

In the first phase of training, recruits must graduate from the Dallas Police Academy. Training at the Academy lasts at least 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 them for the demands of a career in law enforcement. Next, the recruits complete legislatively mandated classroom and scenario-based training as well as a number of

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

additional courses required by Dallas PD. The legislatively mandated courses are developed by the Texas Commission on Law Enforcement (TCOLE), which is the regulatory agency governing the licensure of all peace officers in Texas. TCOLE 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) that sets both Academy and in-service training requirements. Although there is some 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 that have been broadly implemented across the country and are consistent with how Dallas PD 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 a more experienced officer (i.e., their FTO) in an apprenticeship-style model where they are gradually afforded more autonomy. FTOs have a dual responsibility of providing service in their sector while simultaneously providing on-the-job training for their assigned recruits. At the end of field training, recruits are evaluated by their FTO and, based on a successful evaluation, graduate to becoming full-fledged police officers. In our setting, nearly all recruits who successfully graduate from the Academy also successfully 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 for how law enforcement agencies in the United States approach the training of new recruits (President's Task Force on 21st Century Policing, 2015).

Key to the empirical design of our study, recruits in Dallas have no discretion in choosing their FTO. 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 at the Academy. Within a given division, recruits are then assigned to FTOs and their associated patrol sectors/beats. Command staff at Dallas PD indicated that these decisions are entirely unrelated to recruit characteristics or their performance at the Academy. In a subsequent section, we provide empirical evidence supporting the claim that, conditional on division, 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 phases, the FTO is different. The first three training phases of field training are each 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 active participants 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. Figure A.1 shows an example timeline for a recruit who began field training on December 7th, 2015.

This paper focuses on the impact that the first FTO has on the recruit's subsequent enforcement behavior. We made the decision to focus on the first FTO 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 careers. Second, police officers and command staff in Dallas noted that the first phase of field training is the most significant because it is a recruit's first exposure to providing service. Further, recruits often return to their initial FTO for their final training and evaluation phase.[11] We also focus on FTOs rather than the officer a recruit chooses to work with during their probation period 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, with each assigned a value from 1 to 4, with 1 being the highest priority. The information recorded in the CAD system is then provided to police dispatchers whose job it is to assign calls to police officers. Dispatchers assign calls to officers based on priority level (relative to other calls in the queue), proximity, and availability. Unlike in other jurisdictions, officers in Dallas do not have assigned beats. It is very common to observe an officer in many different beats during their shift. 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 Dallas PD's standards about how many officers are required for different types of calls. For example, more serious incidents

[11] If we estimate the effect of each FTO in the same regression, only the force rate of a recruit's first FTO has a large and statistically significant effect (see Table A.1). We interpret this as evidence consistent with our conversations with Dallas officers.

(such as shootings and mental health calls) may involve the dispatch of multiple officers. It is also possible for officers to observe a call and "dispatch" their car to the call, i.e., self select into a particular call. 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.[12] In general, the way that Dallas PD tracks force incidents is consistent with best practices established by criminologists and embraced by many law enforcement agencies across the county. In particular, Dallas PD officers are required to make a "Response to Resistance" entry in a proprietary database called BlueTeams. following a force incident[13] All force incidents are reviewed by a supervisor (Dallas Police Department, 2021). 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, 2021). According to police officers and command staff at Dallas PD, the penalty for not correctly reporting a force incident is extremely severe and compliance is virtually universal. Stephen Bishop, Major of Police DPD Research Division, states "Compliance with use-of-force polices is strictly enforced and widely monitored, thus under-reporting of force events is rare and not been

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

[13]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.

historically problematic."[14] 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 arrested (race, gender, age). We also categorize felony and misdemeanor arrests as filed or unfiled.[15] If an arrest is unfiled then the district attorney decided to not move forward with the case and the defendant will not be charged with a crime. We interpret unfiled arrests as resulting from lower quality police work.[16] Last, we have data on police complaints. These include both internal and external complaints against Dallas police officers. While we can link these complaints to officers, we do not have enough information to link them to specific calls for service.[17]

3 Data and Summary Statistics

3.1 Analytical Sample

Our analytical sample is derived from the universe of 3.9 million calls for service (i.e. 911 calls) received by Dallas PD from Jan 2013 to July 2019. We link these data to force re-

[14] Personal correspondence between authors and Stephen in 2023."The Dallas Police Department mandates use-of-force reporting. DPD has policies and procedures in place that are designed to guide the timely reporting of force incidents and ensure compliance. Additionally, nearly all current Dallas Police officers serving in an operational position (e.g., patrol) have been issued BWCs. Recruits have been issued BWCs by the time they graduate from academy training. All marked patrol vehicles are equipped with in-car cameras. Use-of-force incidents, BWCs, and in-car cameras receive separate month audits by officers' chain-of-command. Compliance with departmental policies is also monitored by DPD's Planning & Accreditation Unit. Compliance with use-of-force polices is strictly enforced and widely monitored, thus under-reporting of force events is rare and has not been historically problematic."

[15] We only observe whether a case is filed or not for misdemeanors and felonies. N-class arrests are typically for cases with a warrant already.

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

[17] To our knowledge, there is no "stops" dataset available in Dallas largely because Dallas does very little traffic enforcement.

ports, arrest records, Dallas County District Attorney records, and officer characteristics.[18][19] According to the Dallas Police Department, they do not keep an official historical 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.[20] 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 characterize this officer as the recruit's 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. The institution of this weighting scheme is that the calls where a recruit arrives with only one other officer (weight = 1) provide more information about the identity of their FTO relative to calls where there are many senior officers (weight = $\frac{1}{n}$) 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

[18] In linking the force and arrest records with calls for service, we do so based on the incident identifier but not the officer 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 timestamp) occurred at the scene of an incident as opposed to those occurring after a suspect is in custody and being transported to jail. The likelihood of a call for service to result in force in our sample is comparable to that reported in Weisburst, 2022, Hoekstra and Sloan, 2022, and calculated by the authors in other publicly available data. For use of force, if we also match on officer name we lose 9% of our force observations. However, our results are qualitatively similar using this alternative data restriction (coefficients of 0.000222***, 0.000229,*** 0.000170** for the specification shown in Table 2).

[19] We link arrests to Dallas District Attorney Data on filed cases using defendant name and date of offense. For each match, we block on date of offense, then measure name similarity by Jaro-Winkler distance. If there is a perfect match on name, we keep only that match. Failing that, we keep matches with a Jaro-Winkler score higher than 0.9 for both first and last name. This is a high threshold but allows some room for spelling and transcription mistakes.

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

Senior Corporal. We are reasonably confident that we have correctly identified the recruit-FTO pairings in the vast majority of our sample. However, we verify our FTO identification using another dataset documenting overtime pay.[21] In Dallas, each FTO is eligible for overtime pay to compensate for the time spent completing the necessary paperwork to evaluate a recruit after each shift. To check whether our procedure for identifying FTOs is reasonable, we verify that each officer we have flagged as an FTO is observed as receiving overtime during the training period. We find that 390 of the officer-recruit pairs that we have identified as FTO-recruit pairings also appear in the overtime pay dataset while 21 (approximately 5%) are not. Our results are robust to dropping these pairings where we fail to find the FTO in the overtime data.[22] 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 initial phase of field training. Thus, these are the senior officers who were most likely to have an impact of a recruit's subsequent policing behavior regardless of whether they were 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.[23] Although command staff emphasized that there is still some selection in terms of who was allowed to become an FTO, it was 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 a patrol officer. FTOs were also generally representative of the whole police force in terms of demographics, but perhaps a bit less diverse. Specifically, 19 percent of FTOs were Hispanic,

[21] These data also include comp. time taken instead of overtime.

[22] We estimate coefficients of 0.000228***, 0.000226***, 0.000168** for our Table 2 specifications.

[23] 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.

16 percent were Black, and 63 percent were White, compared to 20 percent, 23 percent, and 53 percent in the entire force, respectively.

3.2 Force Rate Calculation

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 an FTO's time-invariant propensity to use force on a call for service for a specific recruit. 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 FTO in the period prior to being assigned a given recruit.[24] 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, and day of the week-by night fixed effects.[25][26] The intuition behind this exercise is to create a measure that captures how likely an FTO is to use force in the period prior to being assigned a given recruit and after accounting for the fact that some officers may respond to different types of calls than others. Formally we estimate

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

[24]In practice, this means some FTOs will have more than one fixed effect. Thus, the fixed effect will be unique and estimated separately for each recruit-FTO pairing as opposed to each FTO. These fixed effects will be estimated using the pre-period data relevant to each specific pairing. For example, the fixed effect for a given FTO with a recruit assigned in a later period will leverage more data than the fixed effect for the same FTO assigned to a different recruit in an earlier period. Across different recruits, a given FTO has a remarkable similar propensity to use force. In particular, the correlation coefficient from a comparison of the overall force rate with the pair-specific rate is over 0.94.

[25]Our Table 2 are similar, but a bit smaller, if we calculate our force rate without controls or fixed effects.

[26]There are 47 priority-by-type fixed effects. The type of call is categorized by the call taker. An example type of call is burglary.

where $force_{o(r),c}$ is a binary variable equal to one if call c answered by FTO o ends in force and zero otherwise.[27] The vector X_c includes controls that characterize a call for service including indicators for the number of officers on the scene, beat, type of call (priority-by-type), calendar month, and day of the week-by-night. The coefficients of interest $\lambda_{o(r)}$ is a measure of the historic force propensity of FTO o, conditional on call characteristics, prior to being assigned a given recruit r. Since we are stacking the calls for service data for each FTO prior to being paired with each recruit and treating each pairing distinctly, the estimated fixed effects can be interpreted as an FTO's average propensity for using force that is exogenous to the particular recruit. Higher values of $\lambda_o(r)$ indicate that a FTO has historically been more aggressive (i.e., uses force more frequently) while lower values of indicate a FTO has been less aggressive (i.e., 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.

Since our analysis focuses on cohorts of new recruits that joined Dallas PD between July 2014 to December 2018 and our data spans the period from January 2013 to July 2019, the force measure assigned to each FTO-recruit pairing will rely on varying amounts of pre-period calls for service data. In particular, a pairing made in July 2014 will rely on (at most) 1.5 years of pre-period data to calculate the FTO's prior force propensity while a pairing made in December 2018 will rely on (at most) 6 years of pre-period data. In addition to these possible issues related to left truncation, force is also a relatively rare outcome of a call for service with a substantial amount of variation both across and within FTOs. We address both of these concerns by adjusting our estimates of FTO force $\lambda_{o(r)}$ using Empirical Bayes following Weisburst (2022). In particular, we construct a shrinkage factor that attenuates the estimates towards the mean for officers where we observe less pre-period data (due to truncation), observe answering fewer calls for service, or who just have a larger within-officer

[27] We denote FTO officer o as a function of recruit r since a given FTO can appear in the sample training multiple recruits. Thus, each force measure is computed using only the pre-period data for a specific recruit.

variance in their propensity to use force. The intuition of applying Empirical Bayes is that the resulting measure will vary principally on FTO force estimates which we have estimated with the highest degree of confidence.[28]

Formally, we estimate the across officer variance in FTO force, σ_A^2, and a within-officer variance, σ_W^2 [29] 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, we construct our final shrunken force rates as

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

where we multiply our shrinkage factor by our original fixed effects. We plot the distribution of police officer force rates for all 411 FTO-recruit pairs in Figure 1 for the raw and shrunken measure. As expected, the distribution of the shrunken measure is narrower (has a smaller standard deviation) relative to the unshrunken measure. Values above zero indicate that the field training police officer is more likely to use force relative to the average FTO. A number less than zero indicates that the FTO is less likely to use force relative to the average FTO. For the remainder of our analysis, we will focus on a standardized version of the shrunken FTO force measure (i.e., a z-score) for ease of interpretation.[30]

The distribution of standardized effects is shown in Figure 1b. One standard deviation increase in FTO effects is a 0.149 percentage point (121 percent compared to the sample mean of 0.123 in Table A.4). Moving from the FTO that used the least amount of force to the most is an approximate 6.5 standard deviation increase. Replacing an FTO at the

[28] We can also force FTOs to have the same number of calls before being assigned a recruit (2000 calls). Our main results (Table 2) are similar in magnitude if we make this restriction and re-estimate force rates. The limitation of this method is that we must restrict our sample of FTOs. We also acknowledge that an earlier draft of this paper contained a minor mistake that we made when calculating the degrees of freedom for our shrinkage factor. The result was a slightly more conservative measure of shrinkage. We have corrected the error and the main results from the paper remain largely the same.

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

[30] We note that our main results are robust to using the unshrunken estimates as well as a number of alternative specifications (see Table A.2).

10th percentile for one at the 90th percentile corresponds to nearly a 2.5 standard deviation change.

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 standardize the force rates as described above. Our results are shown in Figure A.2.[31] On average, FTOs use force more, 0.08 standard deviations on average, than the typical non-FTO officer, and more than the average senior Corporal or Sergeant (the ranks most likely to be FTOs). Despite these differences, our main takeaway is that there significant overlap between the distributions. 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.

3.3 Summary Statistics

We present summary statistics at the recruit level in Table A.3. 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 percent are Black, and 30 percent 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.[32] A t-test of the difference in recruit characteristics across high and low force FTOs is not statistically significant.[33]

[31] There are a few (5.5 percent) very extreme force users in our sample that we drop to create a figure that is easier to "see."

[32] High force FTOs use more force than the average FTO, and low force FTOs use force more than the average FTO.

[33] We estimate p-values of 0.22, 0.21, 0.95, 0.16 ,0.63 for test of difference across means in race, gender and age.

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 A.4. In our sample, roughly 3.7/100 calls end in any arrest, 2.2/100 end in a misdemeanor arrest and only 1.2/1000 calls end in a use of force. We characterize a call as having involved force or arrest regardless of the specific 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.[34] 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. To preview our main results, we also split the sample into calls where the recruit has a high force FTO (force measure greater than zero) or low force FTO (force measure less than zero) in Table A.4.

4 Empirical Methods

4.1 Estimation Model

The conditional random assignment of recruits to FTOs provides an ideal setting for investigating how field training shapes a recruit's subsequent policing behavior. We formally explore this question by estimating a model of the form

$$force_{r,c} = \theta_r + \beta_1 \Lambda_{o(r)} + \beta_2 X_c + \epsilon_{r,c} \tag{3}$$

where $force_{r,c}$ is a binary variable equal to one if call c ends in the use of 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 their pairing. As discussed, we shrink this measure using Empirical Bayes

[34] For force incidents, we also require the time on the 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 arrested and in police custody.

and standardize it for ease of interpretation. Thus, our coefficient of interest β_1 can be interpreted as the difference in a recruit's likelihood of using force on a given call caused by a one standard deviation increase in their FTO's prior propensity to use force. We control for possible variation across 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 recruit attributes. In our fully saturated model, this vector includes recruit 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-by night fixed effects. Standard errors are clustered at the recruit level. [35] We also note that by applying Bayesian shrinkage to our primary explanatory variable, we are accounting for estimation error in this derived variable (FTO force) in a similar way as inverse variance weighting.[36]

The model's identifying assumption is that FTO characteristics, primarily prior propensity to use force, are not correlated with recruit characteristics after conditioning on division by cohort year. Therefore, identification relies on the conditional random assignment of recruits to FTOs within a given division by cohort year. If random assignment of recruits to FTOs did not exist, we would potentially over/under state the impact of an FTO's prior force propensity since this measure might be correlated with other characteristics of a recruit that might also impact the outcome of a call. In other policing agencies where there is not random assignment of recruits to FTOs, it is reasonable to imagine selection across a number of dimensions that could potentially confound the estimates. In the next section, we will empirically demonstrate that FTO characteristics including propensity to use force are uncorrelated with recruit characteristics.

[35] 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 one-way clustering by recruit or division by cohort year.

[36] We are also robust to inverse variance weighting as well as bootstrapping our standard errors.

4.2 Research Design

We begin this section by showing that FTO characteristics are not correlated with observable characteristics of their assigned recruit which would potentially confound our main estimates. While we expect this to be true based on discussions with Dallas command staff about FTO assignments in Dallas, we also provide empirical evidence below. To begin, we regress FTO characteristics on recruit characteristics where the unit of observation is a recruit-FTO pair.[37] Each specification includes division 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 1. Of the 24 coefficients reported, only one is statistically significant at conventional levels.[38] Further, none of the p-values from a joint F-tests are statistically significant at conventional levels.

We also plot the distribution of FTO force rates by recruit characteristics in Figure A.3. Given the random assignment of recruits, we would expect these distributions to be very similar. Indeed, the distributions appear to be 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 we are justified in interpretting our results as plausibly causal, i.e. that the coefficient β_1 on the variable $\Lambda_{o(r)}$ from the prior section is the effect of an FTO's propensity to use force on a recruit's subsequent policing behavior.

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

[38] Here, we present robust standard errors. However we are robust to clustering at the division by cohort year level or FTO level, i.e. in each one coefficient and none of the F-test p-values in the tables are significant at conventional levels. See Table A.6 and Table A.5.

5 Empirical Analysis

5.1 Evidence from the Raw Data

We begin by presenting some motivating figures for our analysis. While our formal results always condition on cohort year by division (i.e. the institutional unit where random assignment occurs), we believe it is useful to consider the relationship between a FTO's propensity to use force and their recruit's subsequent use of force in the raw data. In the top panel of Figure 2 (a), we plot local average recruit use of force across different FTO force rates. Observations are grouped such that each point includes an equal number of calls. A linear fitted curve is plotted across all force rates. There are two takeaways from this figure. First, we observe both higher and lower force FTO being dispatched to calls. Second, and perhaps most important, there is a clear positive relationship between recruit use of force and their FTO's propensity to use force. The slope of the linear fit suggests that a one standard deviation increase in the FTO's propensity to use force will lead to an increase in recruit force by 0.03 percentage points.

In the bottom panel of Figure 2 (b), we again use the raw data to ask whether recruits who are assigned to a high-force FTO are more likely to select into more dangerous calls (ex-ante) that have a high probability of ending in force. If this were true, the data would suggest that the mechanism is driven by selection into more dangerous calls rather than about an officer's willingness to apply force, ceteris paribus. Another interpretation of the predicted results is that if we document a large positive relationship between FTO force and predicted recruit use of force, this could indicate that higher force FTOs lead to more predictable, or reasonable use of force, as opposed to less predictable and more undesirable use of force.

To assess this alternative explanation, we first regress a recruit's use of force on cohort year and initial assignment fixed effects. 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, and day of the week-by night. We then use that model to predict the likelihood that force would be used for a given call and add the mean use of force rate to pin down our measure.[39] [40] Although there does appear to be a slight positive correlation between predicted force and FTO force rates, the relationship is much weaker than that shown in the top panel. Given these results and out of an abundance of caution, our preferred specifications include call controls such that we can attempt to rule out the selection channel. Specifically, our fully specified model will include fixed effects for number of officers on the scene, beat, type of call (priority-by-type) year-by-month, and day of the week-by night. We will also expect that the passage of less predicable and potentially less desirable use of force from FTOs to recruits will drive our results.

Taken together, these figures provide strong supporting evidence that FTOs are an important determinant of recruit use of force. One downside of this graphical analysis is that there is potential for recruit sorting across cohort years and divisions. However, in discussion with the Dallas Police Department, we believe this type of sorting is limited and not a substantial source of bias. We now turn to our main analysis where we control for cohort year by division fixed effects as well as additional call and recruit characteristics.

5.2 Main Results

Next, we present results for the effect of FTOs in Table 2. 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.[41] The outcome variable for each column is the proportion of 911 calls that end in force. Column 1 presents our baseline specification where the coefficient on $force_{r,c}$ captures the difference between recruit use of force for recruits

[39]Intuitively, this produces a linear combination of call characteristics, where the weights are chosen based on the prediction of the likelihood of force being used.

[40]Coefficient on linear fit is 0.005 percentage points.

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

assigned to an FTO with one standard deviation higher force propensity. Our results show that recruits with FTOs that use force one standard deviation more are 0.0222 percentage points or 18 percent more likely to use force.[42]

In column 2, we add controls for recruit characteristics (age, gender, race). Given our conditional random assignment and the results in Table A.6 and Figure A.3, we would not expect recruit characteristics to alter our estimate meaningfully. Column 2 shows that recruits with FTOs that use force one standard deviation more are 0.0229 percentage points or 19 percent more likely to use force.

In column 3, we add controls for 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 generally tend to be more severe. Another way to interpret column 3, is to think of the inclusion of call controls as controlling for predictable or more desirable use of force. Given the small positive correlation in Figure 2, we should expect the magnitude of our estimates to attenuate slightly. Indeed, our estimate in column 3 is about 23 percent smaller but still economically meaningful and statistically significant at the 5 percent level. Even after controlling for call characteristics and holding constant this alternative channel, we find a one standard deviation increase (121 percent) in FTO force increases recruit force by 0.017 percentage points or 14 percent. We interpret these and the prior results as providing additional evidence in support of the preferred mechanism, i.e. that FTOs transfer information to recruits about the appropriate use of force and that our results are not only driven by more predictable use of force.

[42] In Table A.2 we show results for alternative measure of force rates. Namely, our results are robust to using the unstandardized measure, inverse hyperbolic transformation of our shrunken force measure, and the unshrunk force measure. Results are similar and statistically significant at conventional levels. While not shown directly in the table, we note that our results are similar when we restrict the outcome variable to force incidents where an arrest was and was not subsequently made. If our results were being driven purely by unjustified excessive force, one might expect the results to be larger when we construct the dependent variable from only force incidents that do not result in an arrest.

Next, we explore whether certain recruits are particularly susceptible to adopting the force behavior of their FTO. In particular, we examine heterogeneity in our main estimates across recruit characteristics like race, gender, and age. Our results are shown in Figure 3a where we report the coefficient on FTO force rate. All coefficient estimates, except for female recruits, are greater than zero. Although there is some variation across these subgroups, the main takeaway of the figure is that all of these subgroups appear to be impacted in the same manner by their FTO's prior propensity to use force.

Finally, we consider the possibility that FTOs with certain characteristics may be better able to transfer their force behavior to recruits. In particular, we test for heterogeneity by FTO characteristics in Figure 3b where each coefficient is from a separate regression. Similar to the recruit characteristics plot, every coefficient is greater than zero except for female FTOs. It is also true that younger officers seem to have larger effects in both figures. Our main takeaway from these figures is that, while there may be some variation across subgroups, the effect of FTO force is prevalent and consistent across nearly all recruits and FTOs. From a policy perspective, this is important because it shows that many different types of recruits could be influenced by reforms to field training or stricter screening of FTOs. Further, back-of-the-envelope calculations show replacing top decile FTOs with bottom decile FTOs would reduce overall force by 5 percent.

5.3 Persistent Effects and Potential Attrition

Understanding the importance of FTOs in terms of force behavior requires considering how long our treatment effect persists. In our setting, this is particularly important given Holz et al. (2023) documents only short-term effects for the same outcome but a different treatment, i.e. the effect of peer injury on use of force.

To consider how our main effects evolve over time, we estimate a model of the form

$$force_{r,c} = \theta_r + \Sigma_{t=0}^{7} \beta_t \Lambda_{o(r)} biannual_t + \beta_2 X_c + \epsilon_{r,c} \tag{4}$$

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. call and recruit characteristics). Results are shown in Figure 4. For the first two and a half years, every coefficient is greater than zero. After two and a half years the effects appear to attenuate significantly. However, we note that our sample becomes very thin as we examine further time horizons and we are likely relying on less variation in FTO force propensity. That said, it is worth nothing that most officers are promoted to senior corporal after three years of service and themselves either become a detective or FTO.

It is also reasonable to be concerned that our results potentially suffer from selective attrition bias. For example, recruits paired with lower force FTOs might be more likely to be terminated or take assignments where they no longer respond to calls for service. With respect to termination, it is highly unlikely that our results are driven by attrition along this dimension as only three recruits leave the Dallas Police Department in the first three years after training. With respect to recruits accepting alternate assignments where they no longer respond to calls for service, we find only 1% of sample exit these data in months 1-23 but an additional 9% exit in month 24 and 13% exit in months 25-30. We address this potential concern by first asking whether FTO force rates are correlated with the last date we observe a recruit in the calls for service data.[43] Regressing each recruit's exit date on FTO force, we do not find evidence that FTO force predicts when recruits stop responding to calls for service in our data. Next, we alternatively address this potential concern by limiting our estimation sample to the calls for service data occurring in months 1-23 when there is virtually no attrition in our data. Using this alternate sample in Table A.7, we find very similar results to Table 2. Given both of these results, we believe that we can confidently put aside potential concerns of attrition bias driving our results.

[43]Specifically, we regress the last date we observe a recruit in the calls data on FTO force rate using the specifications shown in columns 1, 2, and 3 of Table 2. None of the coefficients on FTO force were statistically significant at conventional levels (p-values = 0.318, 0.123, 0.104).

5.4 Randomization Inference

In this section, we provide a robustness test focusing on the calculation of standard errors in our main results, i.e. columns 1-3 of Table 2. 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 rare event, occurring in only 1 percent of calls from 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 tests 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. [44] As with our main estimates, we next construct an estimate of FTO force propensity using calls answered by the randomized FTO in the period prior to being assigned the particular recruit. We then shrink that estimate using the Empirical Bayes procedure described in the methods section and standardize the shrunken measure by subtracting the mean and dividing by the standard deviation. Using the randomized FTO's

[44] We consider an officer as eligible for being a given recruit's FTO if they are observed answering at least one call for service in the same division within 30 days of the recruit's first day assigned to patrol. We also only consider officers as eligible to be 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.

propensity to use force in the pre-period as the primary independent variable, we estimate the model from columns 1-3 of Table 2. In order to obtain p-values for a two-sided hypothesis test, we replicated this procedure 1,000 times and calculate the share of the simulations when the t-statistic exceeds the absolute value of the t-statistics from Table 2.

For the models corresponding to columns 1 and 2 of table 2, we obtain p-values for a two-sided hypothesis test using randomization inference of 0.002 and 0.001 respectively. For illustrative purposes, we also plot the distribution of t-statistics obtained from our randomization procedure in figure 5 corresponding to the fully specified model in column 3 of table 2. In our randomization procedure, we find that thirty-seven of the simulations result in a t statistic more extreme than 2.48. Thus, the associated p-value obtained through randomization inference is 0.036. We interpret these results as providing additional evidence indicating that the inference in our main results is not entirely driven by potential inference issues associated with incorrect asymptotic assumptions or rare-event bias.

6 Mechanism and Recruit Arrests

6.1 Alternative Mechanisms

Up to this point, we have interpreted our primary findings as evidence of an FTO transferring information about the appropriate threshold for applying force to their recruit. This entails an FTO teaching a recruit to behave more or less aggressively. In this section, we aim to provide additional evidence that supports this particular explanation of the underlying mechanism. To accomplish this, we have convincingly ruled out six alternative explanations. These alternative explanations include the following: (1) the possibility that omitted FTO characteristics are correlated with FTO force and a recruit's subsequent force; (2) whether our main findings are driven by more active forms of policing; (3) whether recruits are more likely to be dispatched to calls for service with their FTO even after completing training; (4) whether recruits paired with a high-force FTO are indeed witnessing force during the training period; (5) differential behavior in terms of reporting force; and (6) whether our results are driven by under-reporting of less severe force incidents. Across all of these additional estimates, we have found very little evidence of any mechanism other than knowledge transfer from the FTO to the recruit about the appropriate use of force.

First, we explore whether FTO demographics are predictive of force rates. In Figure A.4, we plot the distribution of force rates by FTO characteristics. On average White FTOs have higher force rates than Hispanic and Black FTOs but none of the distributions are statistically different from each other.[45] Figure A.4b shows that female FTOs are about one-third of a standard deviation more likely to use force than males and these two distributions are statistically different from each other (K-Smirnov p-value $= 0.002$).[46] Figure A.4a also shows that younger FTOs (i.e., less than the mean age of 50) use force 0.13 standard deviations more frequently than older FTOs and the two distributions are not statistically different from each other). Finally, we regress our force measure on FTO race, age, and gender. This regression yields an R-squared of 0.03 and an F-test p-value of 0.1360. Since we observe that force rates are correlated with other FTO characteristics, we now formally control for these measures and re-estimate the main results from column 3 of Table 2 (shown again in column 1) in columns 2, 3, and 4 of Table 4. Overall, these estimates are at least as large as our main results and statistically significant at conventional levels.

Second, we explore whether FTO force is capturing proactive policing, as opposed to a lower threshold for applying force. We construct a set of additional measures that capture other aspects of FTO behavior that we believe are associated with proactive policing. In particular, we construct measures of average FTO arrest (misdemeanor and overall) as well as response times and time spent on a call.[47] We begin by presenting the correlation between

[45] The average Black FTO has a force rate that is about 0.07 of a standard deviation lower than the average White FTO. We also note that there are few Black and Hispanic FTOs constituting only 70 and 68 recruit-FTO pairs respectively.

[46] In the full sample this difference is much less pronounced (on average female officer force rates are less than 1/10th of a standard deviation greater).

[47] Specifically, we estimate Equation (1) using arrest, misdemeanor arrest, response time, and time on calls as the outcome and shrink our estimates according to Equation (2).

FTO force, and other FTO rates in Table 3. Column 1 in Panel A reports that a one standard deviation increase in FTO force propensity leads to a 0.489 standard deviation increase in arrest propensity. We also find similar effects for other types of arrests (columns 2-5). Column 1 in Panel B considers the correlation between FTO force rates, response time, and time spent on a call. Although we find no statistically significant correlation with response time, we find that FTO's with higher force rates tend to spend less time on a call. This result is in line with our conversations with Dallas FTOs, who claim more engaged, or less "lazy", officers do not loiter at the end of calls but instead quickly respond to other calls. We also consider FTO complaints. FTO force rate is not correlated with overall levels of complaints or internal complaints. However, there is a correlation between use of force complaints and FTO use of force (see Table 3). While there is likely to be a mechanical relationship between the volume of force incidents and complaints associated with force, we cannot rule out other potential stories and so include this measure as a control in some of our models. Since we observe that FTO force is correlated with other measures that capture a more active form of policing, we now formally control for arrest rates, misdemeanor arrest rates, response time, time on call, and complaints in columns 5-9 of 4. Overall, these estimates are similar in size and significance to our main results shown again in column 1.

We include all FTO characteristics (columns 2-4) as well as other measures of proactive policing (columns 5-9 in column 10 of Table 4. Our estimate is larger in magnitude than column 1 and statistically significant at the one percent level. Together, these results illustrate that FTOs transferring information about the appropriate threshold for applying force, as opposed to other FTO characteristics or proactive policing, is the most likely explanation for our results.

Third, we ask whether our results are driven by who a recruit is dispatched with after training. First, we consider times when a recruit is dispatched with their FTO after training is complete. There are few calls where we observe FTO-recruit pairs dispatched together (16,562). If we drop those calls from our sample, our Table 2 column 3 (full controls) estimate is a similar magnitude and is statistically significant at the 5% level. We do not believe our results are driven by the continued pairing of recruits and FTOs. Next, we look at other partners (the officer a recruit is observed the most with after training) in Table 5. If we control for partner force rate, measured across the entire sample, and other partner characteristics (age, gender, race), our results shrink some, but remain similar to the estimates in Table 2 and significant at conventional levels. We note that this approach is not our preferred specification as who a recruit chooses to work with could be affected by treatment. Said another way, although including partner characteristics as controls is technically an over-control, we believe Table 5 illustrates that partners do not fully explain our results.

Fourth, we ask whether recruits paired with more aggressive FTOs might be more likely to experience force during their training periods. Said another way, do recruits and FTOs experience a correlated shock (a force incident) that might explain our results. This early exposure to a force event could drive our results. Ninety-eight recruits (24%) experienced a force incident during training. To investigate this explanation, we allow our effect to vary by whether a recruit experienced force during their training. Figure A.5 shows that results are similar no matter if a recruit experienced force during their training period. Within each specification (following specifications in Table 2), coefficients are not statistically different from each other. If anything, results are larger for recruits who did not experience force during their training. The results of this section show that results are driven by FTO force, not other characteristics of FTO policing style, other attributes of a recruit's training period, or peers after training.

Fifth, we explore whether our results are potentially driven by differential reporting patterns amongst high force FTOs that are transmitted to recruits. In conversations with both police officers and command staff at Dallas PD, we asked Dallas Police Department officers about force reporting norms within the department. Every member of the Dallas PD we spoke to reiterated that all force incidents are recorded in BlueTeams and that under reporting was unlikely because the department conducts frequent audits of reports and bodyworn/dashboard camera footage. If an officer were to be caught engaging in unreported force, they would face serious repercussions.[48]

To explore differences in report-writing behavior, we develop three measures based on the section of incident reports completed by the responding officer. [49] Following the procedure described in 3.3, we calculate measures the capture the total number of characters written in incident reports by the FTO as well as number of distinct words and a variable capturing not having entered anything. In panel C of Table 3, we show that our proxies for reporting writing are not correlated with FTO force. In Table 6, we repeat the three columns presented in Table 2 but with the addition of our reporting controls. Across each column each of the estimates are similar in magnitude and significance to Table 2.[50]

The Dallas Police department also introduced bodyworn cameras during our study period and implemented a policy of random audits. If Dallas police were systematically under-reporting force, we should expect to observe that the total volume of force would increase following the introduction of bodyworn cameras and the random auditing procedure. Figure A.6, plots the average daily number of force incidents over time.[51] We note three important dates which include the purchase of police cameras and additional training for police cameras that could affect reporting. The figure demonstrates that there is no clear change in the volume of force incidents reports across these dates. This finding is consistent with Dallas PD's statements that under-reporting is not an issue within their department.

[48]The police department's General Orders reiterate the auditing of reports and body-worn/dashboard camera footage stating "Supervisors will conduct random BWC reviews/audits of personnel assigned to them" (Dallas Police Department (2021)). It is also worth noting that, even if an officer inappropriately uses force on a citizen, the incentives are such that they are actually better off documenting the incident as opposed to potentially receiving a complaint about an unreported incident.

[49]In particular, we rely on the field "Modus Operandi (MO)" which, to our knowledge, is an open-ended text field that is completed by the officer taking the incident report.

[50]There are 10 FTOs that we cannot link to incident reports. Therefore we have approximately 10,000 fewer observations in Table 6 relative to our primary estimates.

[51]We first residualize the data (removing year-by-month fixed effects and adding back in the sample mean).

6.2 Arrest Results

So far, our paper has shown that recruits assigned to high-force field training officers (FTOs) are more likely to use force later in their careers. However, one limitation of existing research on police use of force is the difficulty in determining which interactions or calls for service warrant the use of force (a "quality" application of force) and which do not (a "not-quality" application of force). In other words, there are situations where police need to use force to ensure public safety, but it's hard to differentiate these from situations where force is not warranted.

To investigate whether high-force FTOs engage in other behaviors that could be considered quality or not from a social or policy perspective, we looked at four additional measures in Table 7. Panel A shows results for arrests that are less desirable from a social/policy perspective, while Panel B shows results for more serious and higher quality arrests. We find a one standard deviation increase in FTO force was linked to a 2-6% rise in the probability of a misdemeanor arrest and a 3-9% increase in the likelihood of an unfiled arrest. Unfiled arrests are not formally filed at the District Attorney's office, usually because of a lack of proper evidence or supporting documentation.

In Panel B, we show that recruits with high-force FTOs do not make more felony arrests or filed arrests. Panels A and B taken together show that high-force FTOs not only result in high-force recruits but also produce recruits who make less serious arrests and unfiled "lower quality" arrests but not more serious "high quality" felony arrests. In brief, we found little evidence to support the notion that high force is linked to behaviors that we might describe as "good" policing.

7 Conclusion

This paper estimates the effects of high force FTOs on recruit use of force. We compare recruits quasi-randomly assigned FTOs with higher historical force rates to those with lower force FTOs. Our results show FTOs are an important determinant of subsequent recruit force; a one standard deviation increase in FTO force increases recruit force by 14 percent. FTO effects also persist two and a half years after completing training. We also show that our results are consistent with the story of less desirable policing qualities alone being transferred from FTO to recruit. Given the broad support for reforms to police officer training, the wide availability of alternative FTOs, and the relative ease of switching FTOs, we believe our findings are consistent with FTOs 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.

Bennedsen, M., F. Pérez-González, and D. Wolfenzon (2020). Do ceos matter? evidence from hospitalization events. *The Journal of Finance 75*(4), 1877–1911.

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.

Bureau of Justice Statistics (2018). State and Local Law Enforcement Training Academies, 2018 – Statistical Tables. https://bjs.ojp.gov/sites/g/files/xyckuh236/files/media/document/slleta18st.pdf.

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.

Dallas Police Department (2021). Dallas Police Department General Orders. https://dallaspolice.net/resources/Shared

Dube, O., S. J. MacArthur, and A. Shah (2023). A cognitive view of policing. *Working Paper*.

Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *Journal of the American Statistical Association 99*(465), 96–104.

Falk, A. and A. Ichino (2006). Clean evidence on peer effects. *Journal of labor economics 24*(1), 39–57.

Fenizia, A. (2022). Managers and productivity in the public sector. *Econometrica 90*(3), 1063–1084.

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.

Hendren, N. and B. Sprung-Keyser (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics 135*(3), 1209–1318.

Hoekstra, M. and C. Sloan (2022). Does race matter for police use of force? evidence from 911 calls. *American economic review 112*(3), 827–860.

Hoffman, M. and S. Tadelis (2021). People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of Political Economy 129*(1), 243–285.

Holz, J. E., R. G. Rivera, and B. A. Ba (2023). Peer effects in police use of force. *American Economic Journal: Economic Policy 15*(2), 256–291.

Ipsos (2021). USA Today/Ipsos Crime and Safety Poll. https://www.ipsos.com/sites/default/files/ct/news/documents/2021-07/Topline-USAT-Crime-and-Safety-070821.pdf.

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-misconduct/.

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/.

Rim, N., R. Rivera, A. Kiss, and B. Ba (2022). The black-white recognition gap in award nominations. *Journal of Labor Economics*.

Rivera, R. (2023). The Effect of Minority Peers on Future Arrest Quantity and Quality. *Available at SSRN 4067011*.

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.

Weisburst, E. K. (2022). "Whose help is on the way?" The importance of individual police officers in law enforcement outcomes. *Journal of Human Resources*, 0720–11019R2.

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.

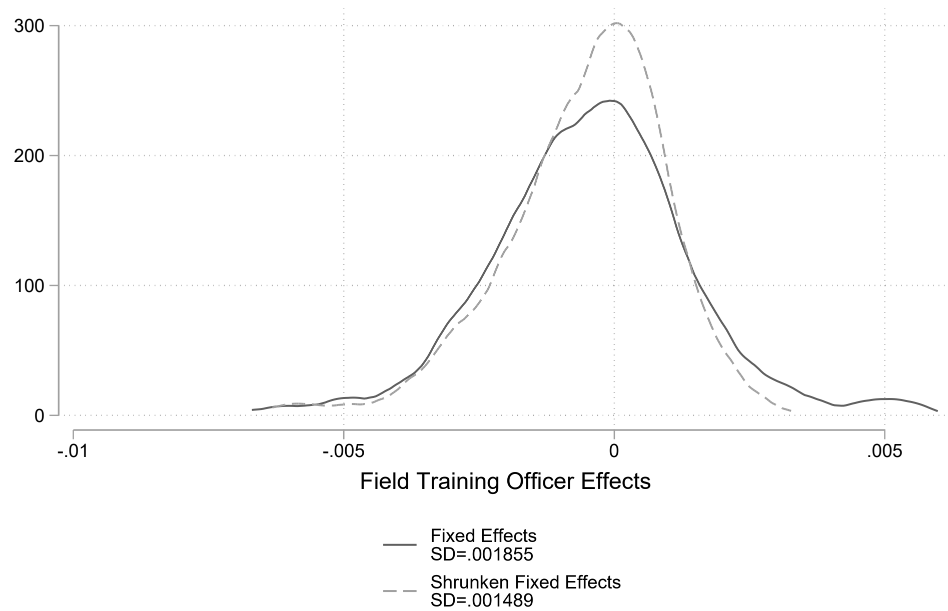
Wheller, L., P. Quinton, A. Fildes, and A. Mills (2013). The greater manchester police procedural justice training experiment. *Coventry, UK: College of Policing*.

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

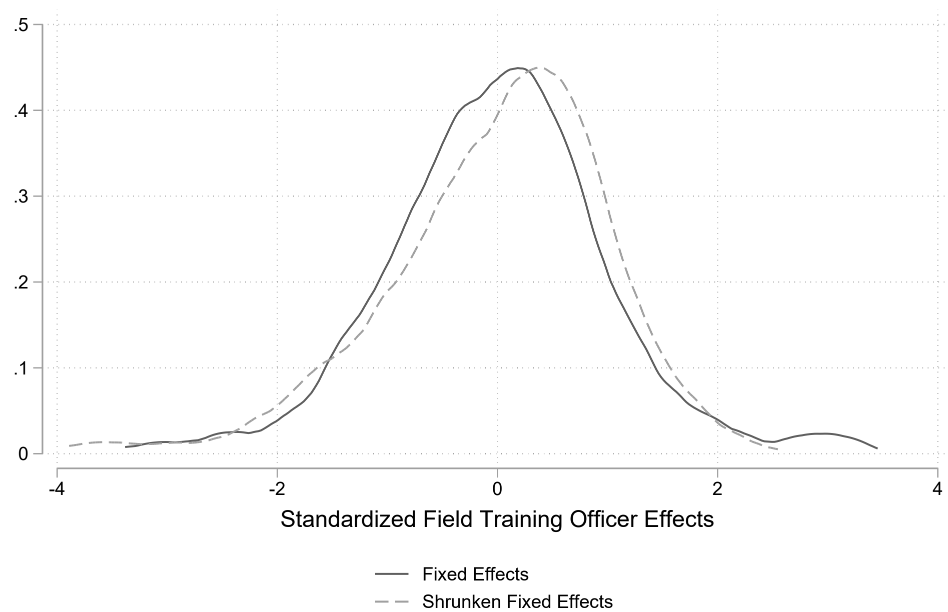
Woody, R. H. (2005). The police culture: Research implications for psychological services. *Professional Psychology: Research and Practice 36*(5), 525.

Figures and Tables

Figure 1: Density of Field Training Officer Propensity to Use Force
(a) Field Training Officer Effects



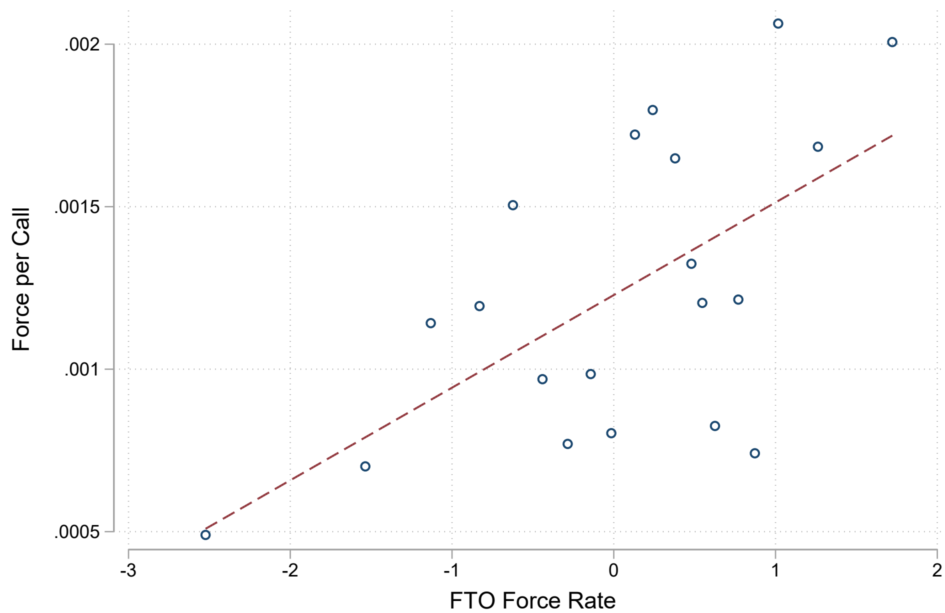
(b) Standardized Field Training Officer Effects



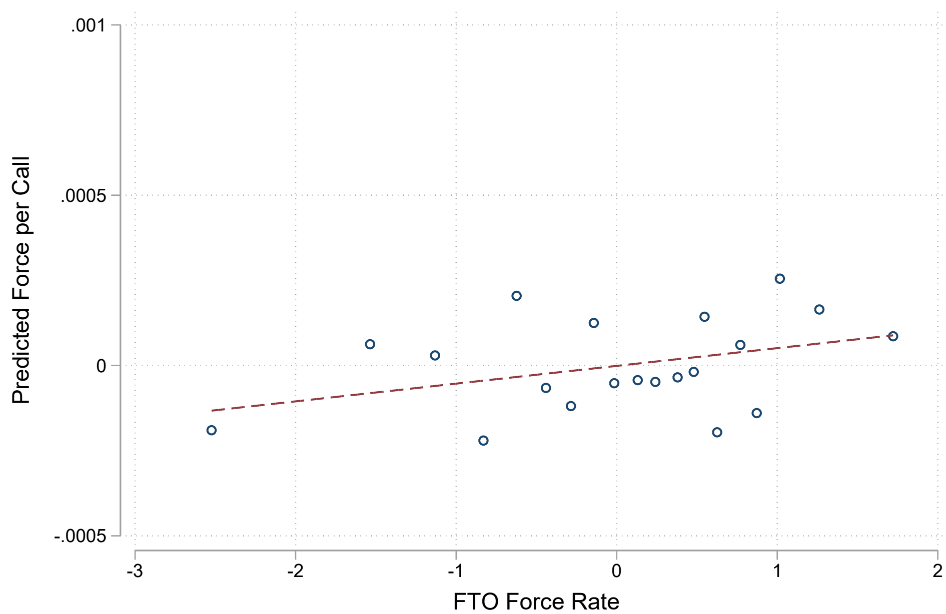
Notes: Fixed effects are calculated after accounting for the number of officers on the scene, beat, type of call (priority-by-type) year-by-month, and day of the week-by night fixed effects.

Figure 2: Recruit Actual Force and Predicted Force by Field Training Officer Effects

(a) Use of Force

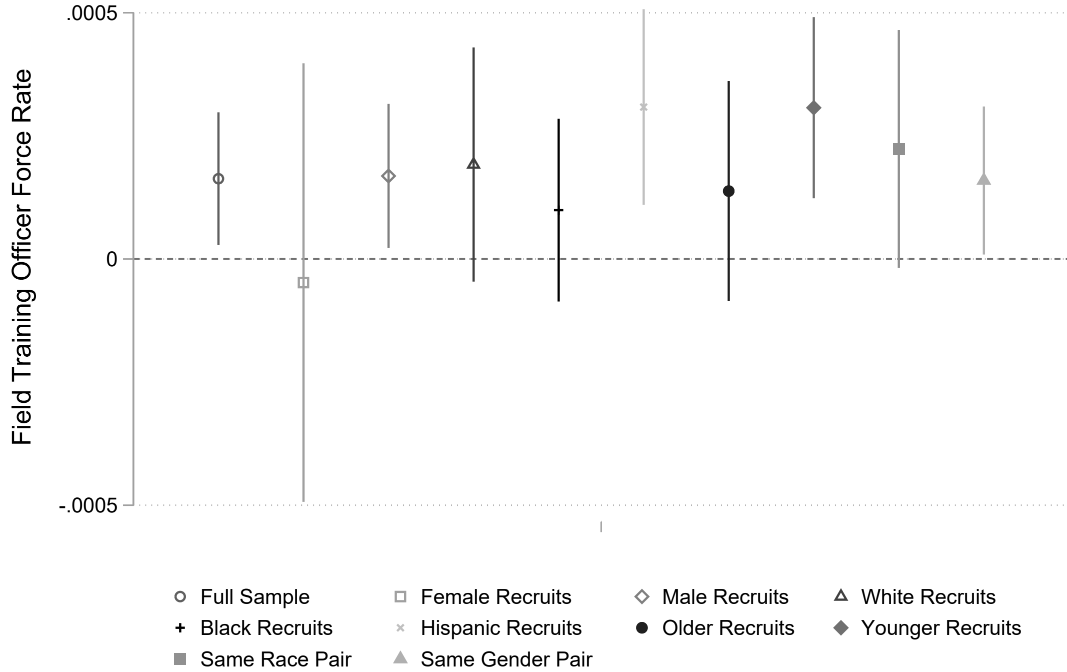


(b) Predicted Use of Force

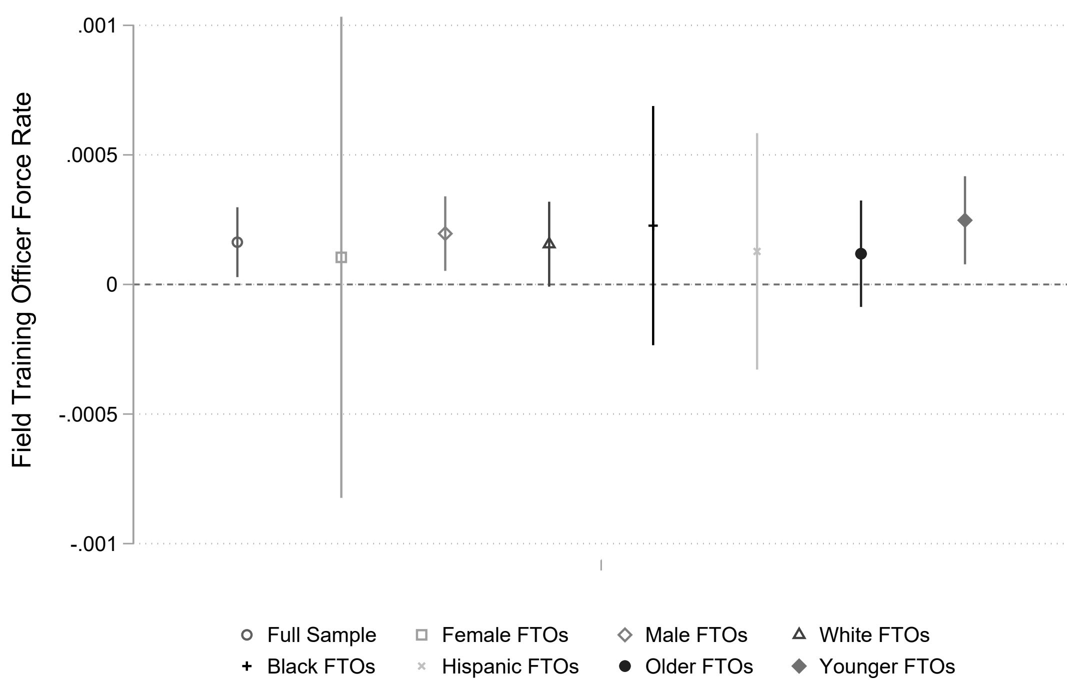


Notes: In Panel (a) we plot use of force. In Panel (b) we plot predicted use of force. The fitted line is a linear fit across all use of force rates. Observations are grouped so that each point includes an equal number of observations. Use of force is predicted using the number of officers on the scene,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

Figure 3: The Effect of Field Training Officers on Force by Recruit and FTO Subgroups



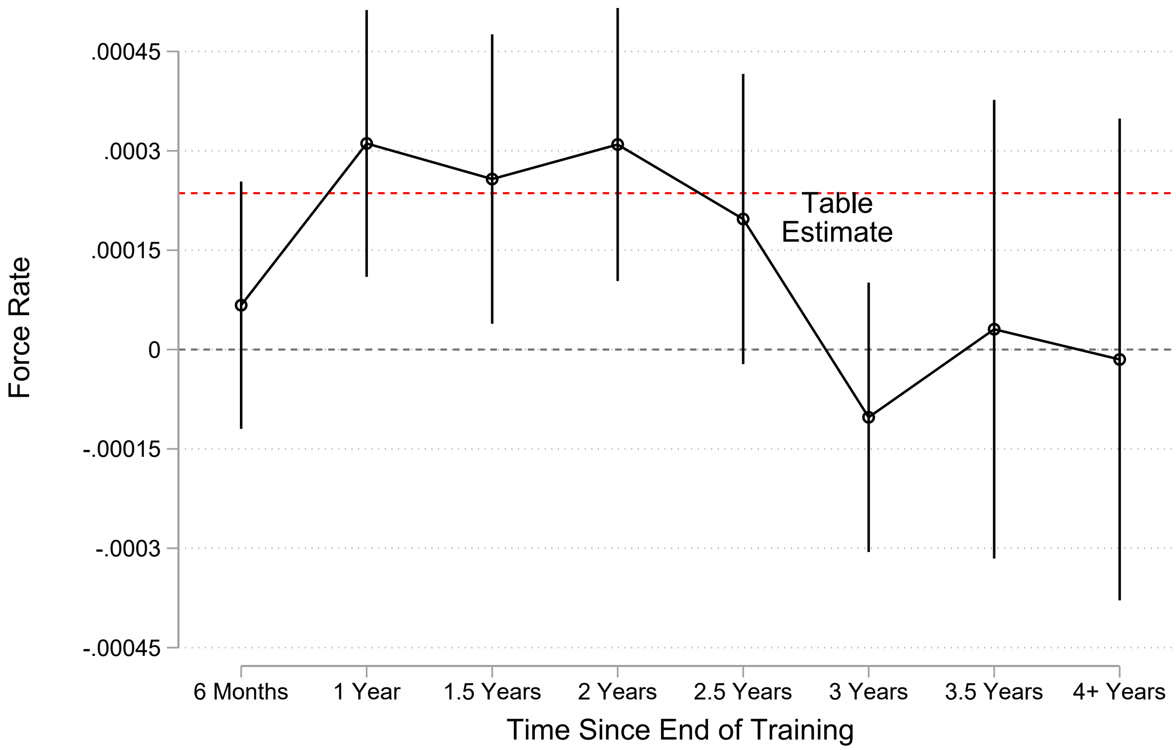
(a) Recruit Subgroups



(b) FTO Subgroups

Notes: Each coefficient is from a separate regression. Standard errors are clustered at the recruit level.

Figure 4: The Effect of Field Training Officers on Force Over Time



Notes: Sample includes all officers observed with three full years of data. The red line marks the estimate from column 3 in Table 2. Standard errors are clustered at the recruit level.

Figure 5: 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 2 by using a 1,000 randomized simulations of the data generating process. Sixteen of the simulations resulted in a t-statistic greater than our our baseline estimate, i.e. one-sided test p-value of 0.016. Thirty-six of the simulations resulted in a t-statistic with an absolute value greater than our our baseline estimate, i.e. one-sided test p-value of 0.036.

Table 1: Balance Test: Correlation between Recruit and Field Training Officer Characteristics in Calls Sample

	(1)	(2)	(3)	(4)	(5)	(6)
FTO Chars	Age	Female	Black	Hispanic	Hire Date	Force Rate
Recruit Chars						
Black	-0.9993	0.0508	0.0531	-0.0805*	0.1303	-0.2139
	(1.1958)	(0.0494)	(0.0556)	(0.0452)	(1.0730)	(0.1369)
Hispanic	-0.2953	0.0531	0.0216	-0.0176	0.3342	-0.1157
	(0.9622)	(0.0409)	(0.0450)	(0.0434)	(0.9086)	(0.1171)
Female	-0.4818	0.0423	0.0702	0.0722	0.7019	-0.1098
	(1.1198)	(0.0474)	(0.0553)	(0.0532)	(1.0309)	(0.1157)
Age	-0.0003	0.0008	0.0007	0.0006	0.0640	0.0112
	(0.0920)	(0.0039)	(0.0042)	(0.0040)	(0.0838)	(0.0094)
Observations	411	411	411	411	411	411
Div-x-Cohort FE	X	X	X	X	X	X
Outcome Mean	49.32	0.131	0.187	0.153	41.98	7.85e-10
F-Test P-Value	0.918	0.547	0.629	0.342	0.863	0.254

Notes: This table presents the results of regressing FTO characteristics on recruit characteristics. Each column is a separate regression. Robust standard errors are presented.

Table 2: The Effect of High Force Field Training Officers on Recruit Use of Force

	(1)	(2)	(3)
	Force	Force	Force
FTO Force Rate	0.000222***	0.000229***	0.000170**
	(0.0000633)	(0.0000622)	(0.0000684)
Observations	1085020	1085020	1085020
Outcome Mean	0.00123	0.00123	0.00123
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: This table presents the effect of FTO force rate on recruit use of force. Standard errors are clustered at the recruit officer level. Column 2 adds controls for recruit characteristics (age, gender, race). We add call characteristics fixed effects (number of officers on the scene, beat, type of call—priority-by-type,year-by-month, and day of the week-by night) in column 3.

Table 3: Correlation between FTO Force Rate and Other FTO Behavior

	Overall Arrest Rate	Filed Arrest Rate	Unfiled Arrest Rate	Misd. Arrest Rate	Felony Arrest Rate
Panel A: FTO Arrest					
Force Rate	0.489***	0.491***	0.386***	0.398***	0.540***
	(0.0636)	(0.0599)	(0.0611)	(0.0591)	(0.0682)
Observations	411	411	411	411	411
	Response Rate	Time on Call			
Panel B: FTO Timing					
Force Rate	-0.00132	-0.161***			
	(0.0719)	(0.0548)			
Observations	411	411			
	Complaints	UOF Complaints	Internal Complaints		
Panel C: FTO Complaints					
Force Rate	0.0319	0.138**	0.0303		
	(0.0221)	(0.0550)	(0.0184)		
Observations	411	411	411		
	Num. Characters Rate	Num. Words Rate	Write Nothing Rate		
Panel D: FTO Reporting					
Force Rate	-0.0881	-0.0857	0.0414		
	(0.0776)	(0.0757)	(0.0806)		
Observations	401	401	401		
Div-X-Cohort FE	Y	Y	Y		

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

Notes: This table presents the correlation between FTO force rate and other characteristics by regressing FTO characteristics on FTO force race. All variables are standardized. Standard errors are clustered at the field training officer level. There are 10 FTO-recruit pairs that we cannot match to incident reports (in Panel D we have 401 FTO-recruit pairs instead of 411). Response time is the number of hours between arrival time and assigned time. Time on call is the number of hours between the time an officer was enroute and when the call was cleared. We cannot link complaints to calls for service, so calculate complaints per call for service instead.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Force	Force	Force	Force	Force	Force	Force	Force	Force	Force
FTO Force Rate	0.000170**	0.000182***	0.000192***	0.000260***	0.000241***	0.000211***	0.000171**	0.000166**	0.000170**	0.000349***
	(0.0000684)	(0.0000701)	(0.0000694)	(0.0000717)	(0.0000761)	(0.0000739)	(0.0000689)	(0.0000687)	(0.0000689)	(0.0000767)
Observations	1085020	1085020	1085020	1085020	1085020	1085020	1085020	1085020	1085020	1085020
Outcome Mean	0.00123	0.00123	0.00123	0.00123	0.00123	0.00123	0.00123	0.00123	0.00123	0.00123
Assigned Div by Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Recruit Characteristics	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Call Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
FTO Gender	-	Y	-	-	-	-	-	-	-	Y
FTO Race	-		Y	-	-	-	-	-	-	Y
FTO Age	-	-	-	Y	-	-	-	-	-	Y
FTO Arrest Rate	-	-	-	-	Y	-	-	-	-	Y
FTO Misd Arrest Rate	-	-	-	-	-	Y	-	-	-	Y
FTO Response Time Rate	-	-	-	-	-	-	Y	-	-	Y
FTO Time on Call Rate	-	-	-	-	-	-	-	Y	-	Y
FTO Complaint Rate	-	-	-	-	-	-	-	-	Y	Y

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

Notes: This table presents the effect of FTO force rate on recruit use of force. Standard errors are clustered at the recruit officer level. Column 1 reports column 3 of Table 2 for reference. Columns 2-4 add FTO characteristics. Columns 5-9 add controls for other FTO rates (see Appendix B for more details). Each specification includes recruit characteristic controls, and call characteristics fixed effects (number of officers on the scene, beat, type of call—priority-by-type,year-by-month, and day of the week-by night).

Table 5: **Partner Controls**: The Effect of High Force Field Training Officers on Recruit Use of Force

	(1) Force	(2) Force	(3) Force
FTO Force Rate	0.000205***	0.000200***	0.000148**
	(0.0000519)	(0.0000461)	(0.0000610)
Observations	1084501	1084501	1084501
Outcome Mean	0.00123	0.00123	0.00123
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: This table presents the effect of FTO force rate on recruit use of force. Standard errors are clustered at the recruit officer level. Column 2 adds controls for recruit characteristics (age, gender, race). We add call characteristics fixed effects (number of officers on the scene, beat, type of call—priority-by-type,year-by-month, and day of the week-by night) in column 3. Each column includes controls for partner gender, age, race and force rate. We define a recruit's partner the officer a recruit is observed the most with after training.

Table 6: **Reporting Concerns:** The Effect of High Force Field Training Officers on Recruit Use of Force

	(1) Force	(2) Force	(3) Force	(4) Force	(5) Force	(6) Force	(7) Force	(8) Force	(9) Force
FTO Force Rate	0.000258***	0.000260***	0.000200***	0.000258***	0.000261***	0.000201***	0.000251***	0.000253***	0.000193***
	(0.0000648)	(0.0000617)	(0.0000691)	(0.0000646)	(0.0000616)	(0.0000690)	(0.0000643)	(0.0000613)	(0.0000686)
Observations	1057790	1057790	1057790	1057790	1057790	1057790	1057790	1057790	1057790
Outcome Mean	0.00124	0.00124	0.00124	0.00124	0.00124	0.00124	0.00124	0.00124	0.00124
Div-Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Recruit Chars	-	Y	Y	-	Y	Y	-	Y	Y
Call Controls	-	-	Y	-	-	Y	-	-	Y
FTO # Characters	Y	-	-	Y	-	-	Y	-	-
FTO # Words	-	Y	-	-	Y	-	-	Y	-
FTO Write Nothing	-	-	Y	-	-	Y	-	-	Y

Standard errors in parentheses

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

Notes: This table presents the effect of FTO force rate on recruit use of force. Standard errors are clustered at the recruit officer level. There are 11 FTOs who are not matched to the incident data, so the number of observations in this Table is slightly different. We also include controls for an FTO's report writing behavior (how many characters/words an officer uses in a report or how often they leave this section blank). To do so, we rely on the field "Modus Operandi (MO)" in the incident report data. This field, to our knowledge, is an open-ended text field that is completed by the officer taking the incident report. These rates are calculated similarly to our force rates. We describe these rates in greater depth in Appendix B.

Table 7: **Other Outcomes**: The Effect of High Force Field Training Officers on Recruit Arrests

	(1) Misd. Arrest	(2) Misd. Arrest	(3) Misd. Arrest	(4) Unfiled Arrest	(5) Unfiled Arrest	(6) Unfiled Arrest
Panel A: Lower Quality Arrests						
FTO Force Rate	0.00135***	0.00117***	0.000505*	0.00126***	0.00108***	0.000438**
	(0.000508)	(0.000420)	(0.000285)	(0.000428)	(0.000346)	(0.000204)
Observations	1085020	1085020	1085020	1085020	1085020	1085020
Outcome Mean	0.0215	0.0215	0.0215	0.0143	0.0143	0.0143
	(1) Felony Arrest	(2) Felony Arrest	(3) Felony Arrest	(4) Filed Arrest	(5) Filed Arrest	(6) Filed Arrest
Panel B: Higher Quality Arrests						
FTO Force Rate	0.000241	0.000265	0.0000932	0.000335	0.000367	0.000173
	(0.000144)	(0.000139)	(0.000119)	(0.000287)	(0.000255)	(0.000230)
Observations	1085020	1085020	1085020	1085020	1085020	1085020
Outcome Mean	0.00755	0.00755	0.00755	0.0148	0.0148	0.0148
Observations	Y	Y	Y	Y	Y	Y
Assigned Div by Cohort FE	-	Y	Y	-	Y	Y
Recruit Characteristics	-	-	Y	-	-	Y

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table presents the effect of FTO force rate on recruit arrests. Standard errors are clustered at the recruit officer level. Columns 2 and 5 add controls for recruit characteristics (age, gender, race). We add call characteristics fixed effects (number of officers on the scene, beat, type of call—priority-by-type,year-by-month, and day of the week-by night) in columns 3 and 6. Unfiled arrests are arrests that are not filed with the Dallas District Attorney's Office.

Appendix Tables and Figures

Figure A.1: Recruit and FTO Training Timeline

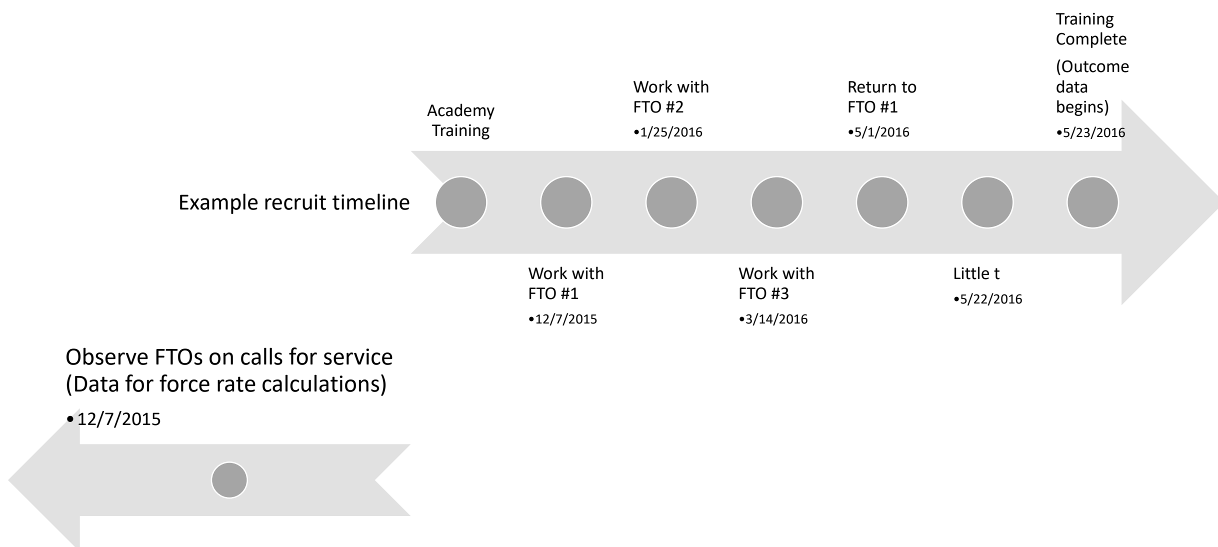
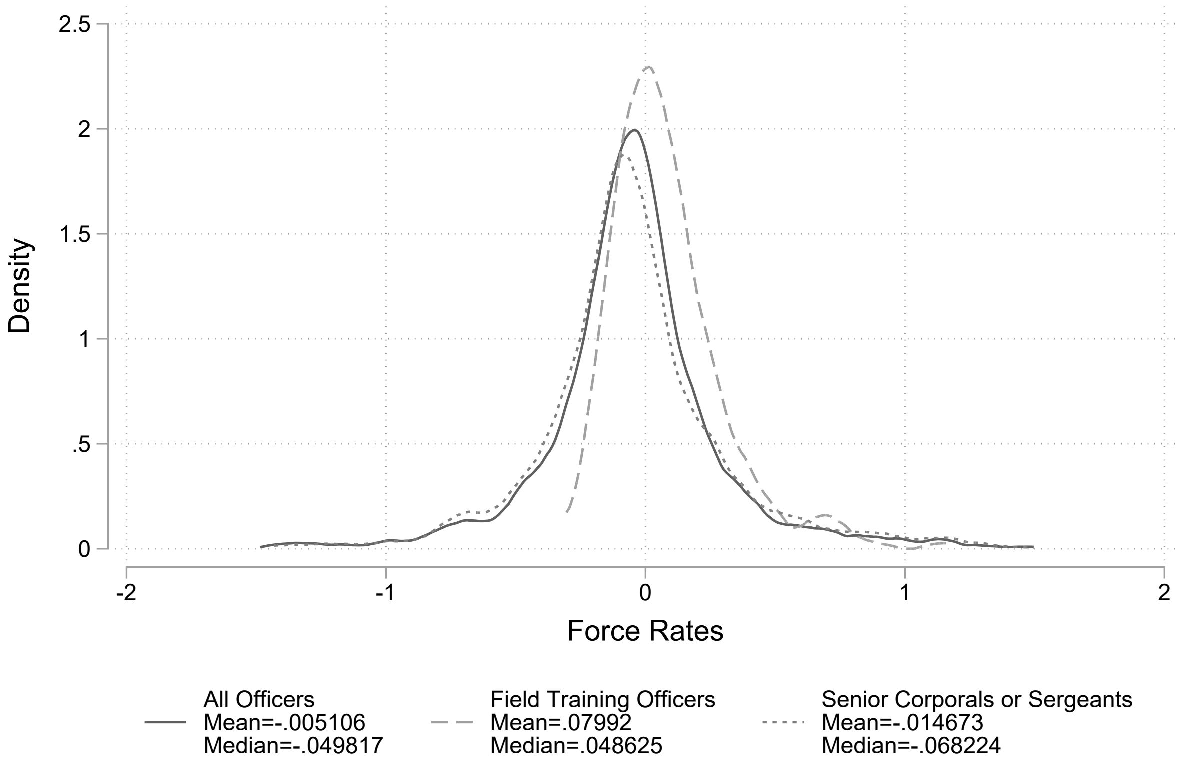
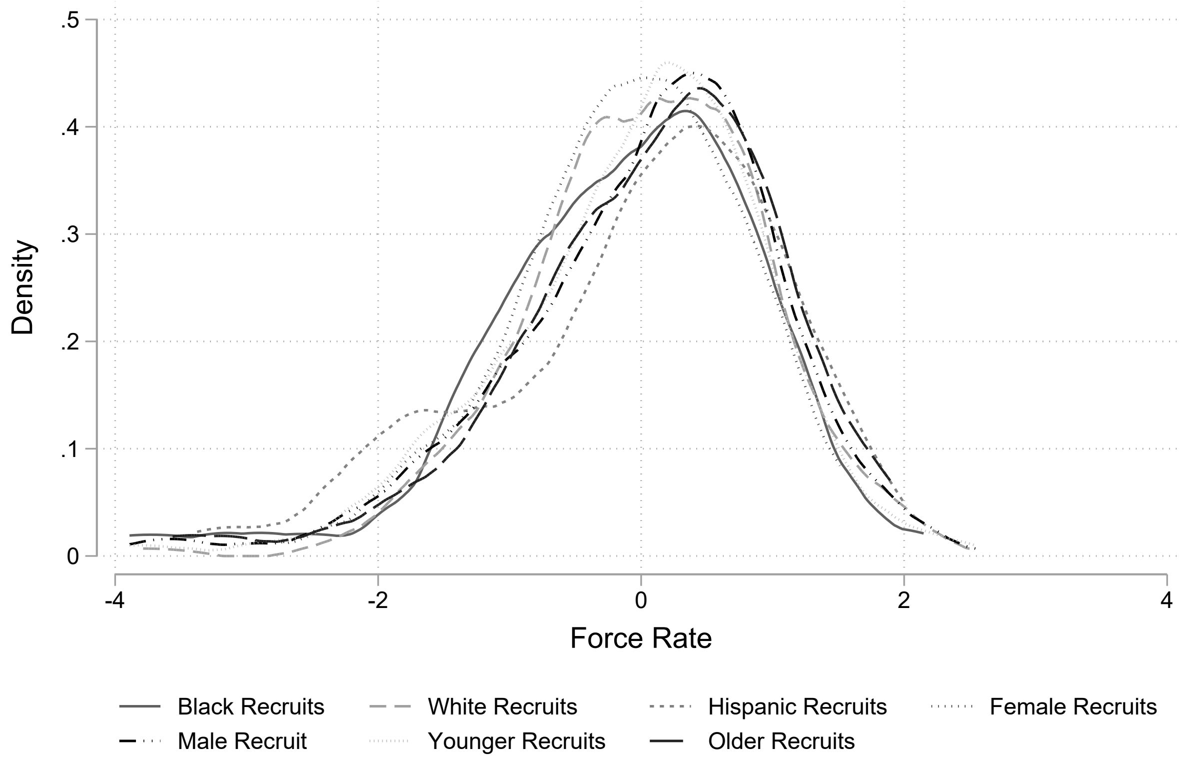


Figure A.2: Density of Officer Propensity to Use Force for All Officers



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.

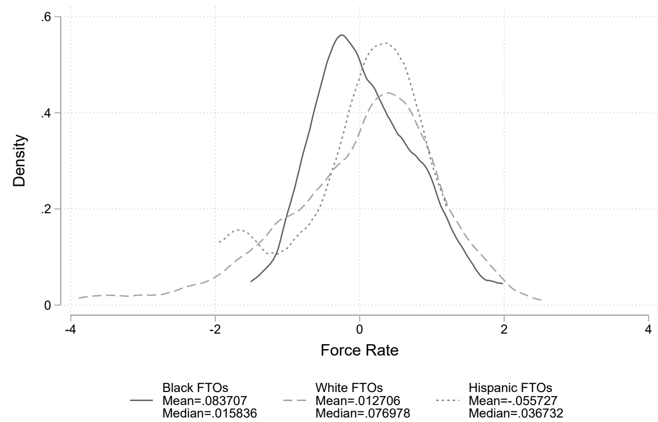
Figure A.3: Density of Field Training Officer Propensity to Use Force by Recruit Characteristics



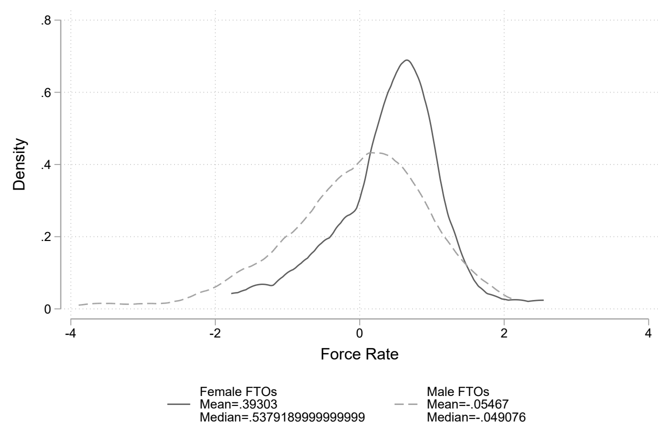
Notes: Older recruits are recruits older than the average age (36 years old).

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

(a) Field Training Officer Race



(b) Field Training Officer Gender



(c) Field Training Officer Age

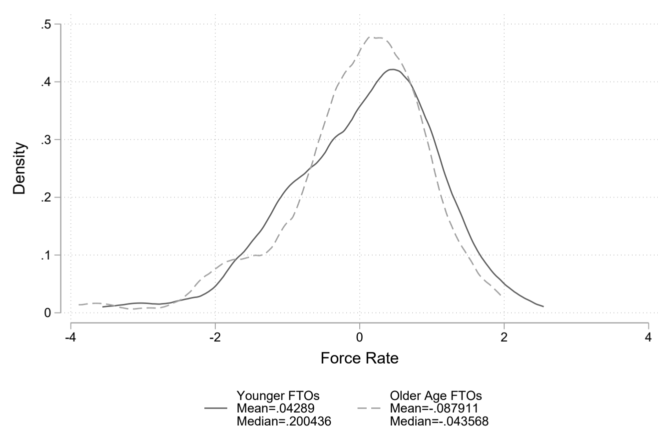
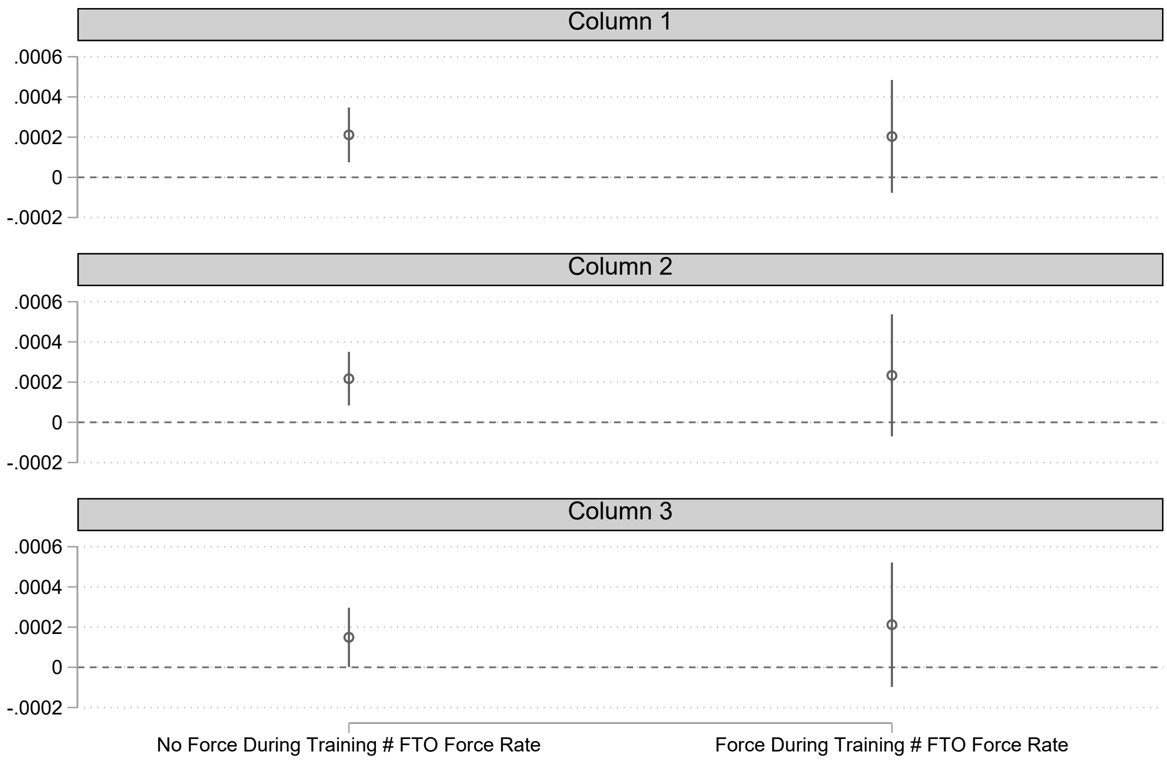
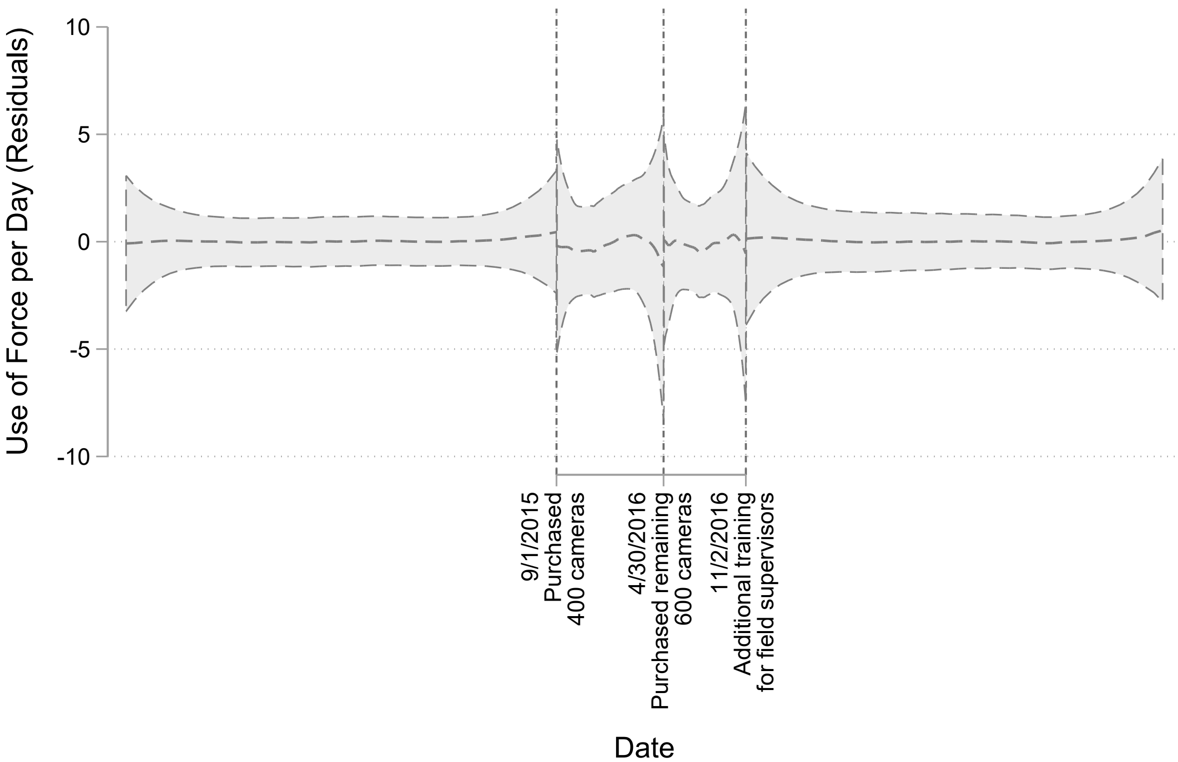


Figure A.5: The Effect of Field Training Officers on Force by Recruit Force Experience



Notes: This figure reports the effect of field training officer force rates for recruits that did and did not experience a force incident during their training period. Columns refer to the specifications used in Table 2.

Figure A.6: Police Camera Adoption



Notes: This figure plots (local linear polynomial fit with 95% confidence intervals) residualized (year-by-month fixed effects) average use of force overtime.

Table A.1: **All Three FTOs:** The Effect of High Force Field Training Officers on Recruit Use of Force

	(1) Force	(2) Force	(3) Force
FTO 1 Force Race	0.000222***	0.000235***	0.000179**
	(0.0000643)	(0.0000635)	(0.0000693)
FTO 2 Force Race	0.0000323	0.0000873	0.0000784
	(0.0000655)	(0.0000657)	(0.0000671)
FTO 3 Force Race	-0.00000219	0.0000438	0.0000594
	(0.0000659)	(0.0000673)	(0.0000666)
Observations	1085020	1085020	1085020
Outcome Mean	0.00123	0.00123	0.00123
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: This figure reports the effect of each field training officer force rate on recruit use of force. Every recruit is assigned three field training officers. The first FTO rides along with the recruit for the first seven weeks of training, the second for the second seven weeks, and the third for the third seven weeks. The recruit returns to the first FTO for final evaluation. Standard errors are clustered at the recruit level.

Table A.2: **Robustness to Different Force Measures**: The Effect of Field Training Officer Force Rate on Recruit Use of Force

	(1) Force	(2) Force	(3) Force	(4) Force	(5) Force	(6) Force	(7) Force	(8) Force	(9) Force
Theta Shrunk	0.144***	0.148***	0.110**						
	(0.0411)	(0.0404)	(0.0444)						
IHS Theta Shrunk				0.144***	0.148***	0.110**			
				(0.0411)	(0.0404)	(0.0444)			
Unshrunken Force Rate							0.122***	0.123***	0.0915**
							(0.0337)	(0.0333)	(0.0357)
Observations	1085020	1085020	1085020	1085020	1085020	1085020	1085020	1085020	1085020
Outcome Mean	0.00123	0.00123	0.00123	0.00123	0.00123	0.00123	0.00123	0.00123	0.00123
Assigned Div by Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Recruit Characteristics	-	Y	Y	-	Y	Y	-	Y	Y
Call Controls	-	-	Y	-	-	Y	-	-	Y

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

Notes: This table presents the effect of field training officers and recruit use of force. Theta shrunk is our shrunken, but unstandardized measure. IHS Theta Shrunk is the inverse hyperbolic sine of our force measure. Unshrunken Force rate is our unshrunken and unstandardized force rate measure. Standard errors are clustered at the recruit level.

Table A.3: Officer Level Summary Statistics

	(1) Entire Sample	(2) High Force Trainer	(3) Low Force Trainer
White	0.438	0.447	0.432
	(0.497)	(0.499)	(0.496)
Black	0.207	0.171	0.232
	(0.406)	(0.377)	(0.423)
Hispanic	0.302	0.312	0.295
	(0.460)	(0.465)	(0.457)
Female	0.178	0.147	0.199
	(0.383)	(0.355)	(0.400)
Age	35.84	36.23	35.57
	(5.466)	(5.662)	(5.318)
Observations	411	170	241

Notes: This table reports officer-level (i.e., one observation per FTO-recruit pair) summary statistics. High Force FTOs have higher force rates than the average FTO, and Low Force FTOs have lower force rates than the average FTO. Standard deviations are shown in parentheses.

Table A.4: Call Level Summary Statistics (Outcomes)

	(1) Entire Sample	(2) High Force Trainer	(3) Low Force Trainer
Force	0.00123	0.00140	0.00110
	(0.0351)	(0.0374)	(0.0332)
All Arrests	0.0370	0.0372	0.0368
	(0.189)	(0.189)	(0.188)
Misd. Arrest	0.0215	0.0221	0.0211
	(0.145)	(0.147)	(0.144)
Felony Arrest	0.00755	0.00757	0.00753
	(0.0865)	(0.0867)	(0.0865)
Filed Arrest	0.0148	0.0149	0.0148
	(0.121)	(0.121)	(0.121)
Unfiled Arrest	0.0143	0.0149	0.0138
	(0.119)	(0.121)	(0.117)
Observations	1085020	456943	628077

Notes: This table reports call-level summary statistics (means). High Force FTOs have higher force rates than the average FTO, and Low Force FTOs have lower force rates than the average FTO. There are three types of arrests in our dataset: felony, misdemeanor, or n-class. Most n-class arrests are for outstanding warrants. We also categorize felony and misdemeanor arrests as filed or unfiled. If an arrest is unfiled then the district attorney decided to not move forward with the case and the defendant will not be charged with a crime.

Table A.5: **Robust to Alternative Standard Errors** Balance Test: Correlation between Recruit and Field Training Officer Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
FTO Chars	Age	Female	Black	Hispanic	Hire Date	Force Rate
Recruit Chars						
Black	-0.9993	0.0508	0.0531	-0.0805*	0.1303	-0.2139
	(1.5174)	(0.0519)	(0.0585)	(0.0466)	(1.2405)	(0.1409)
Hispanic	-0.2953	0.0531	0.0216	-0.0176	0.3342	-0.1157
	(1.0342)	(0.0390)	(0.0439)	(0.0446)	(1.1043)	(0.1185)
Female	-0.4818	0.0423	0.0702	0.0722	0.7019	-0.1098
	(1.1615)	(0.0481)	(0.0519)	(0.0467)	(1.1401)	(0.1630)
Age	-0.0003	0.0008	0.0007	0.0006	0.0640	0.0112
	(0.0951)	(0.0034)	(0.0041)	(0.0032)	(0.0814)	(0.0113)
Observations	411	411	411	411	411	411
Div-x-Cohort FE	X	X	X	X	X	X
Outcome Mean	49.32	0.131	0.187	0.153	41.98	7.85e-10
F-Test P-Value	0.908	0.309	0.523	0.248	0.879	0.314

Notes: This table presents the results of regressing FTO characteristics on recruit characteristics (i.e, replicating Table 1). Each column is a separate regression. Standard errors are clustered at the initial assignment-by-cohort level.

Table A.6: **Robust to Alternative Standard Errors** Balance Test: Correlation between Recruit and Field Training Officer Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
FTO Chars	Age	Female	Black	Hispanic	Hire Date	Force Rate
Recruit Chars						
Black	-0.9993	0.0508	0.0531	-0.0805*	0.1303	-0.2139
	(1.1435)	(0.0491)	(0.0597)	(0.0471)	(1.1402)	(0.1408)
Hispanic	-0.2953	0.0531	0.0216	-0.0176	0.3342	-0.1157
	(1.0012)	(0.0442)	(0.0472)	(0.0453)	(0.9105)	(0.1165)
Female	-0.4818	0.0423	0.0702	0.0722	0.7019	-0.1098
	(1.0613)	(0.0438)	(0.0591)	(0.0535)	(0.9032)	(0.1112)
Age	-0.0003	0.0008	0.0007	0.0006	0.0640	0.0112
	(0.0826)	(0.0036)	(0.0040)	(0.0040)	(0.0796)	(0.0092)
Observations	411	411	411	411	411	411
Div-x-Cohort FE	X	X	X	X	X	X
Outcome Mean	49.32	0.131	0.187	0.153	41.98	7.85e-10
F-Test P-Value	0.909	0.561	0.710	0.356	0.740	0.169

Notes: This table presents the results of regressing FTO characteristics on recruit characteristics (i.e, replicating Table 1). Each column is a separate regression. Standard errors are clustered at the field training officer level.

Table A.7: **Robustness Attrition**: The Effect of High Force Field Training Officers on Recruit Use of Force

	(1)	(2)	(3)
	Force	Force	Force
FTO Force Rate	0.000248***	0.000256***	0.000208***
	(0.0000674)	(0.0000665)	(0.0000717)
Observations	772444	772444	772444
Outcome Mean	0.00128	0.00128	0.00128
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: This table presents the effect of FTO force rate on recruit use of force. The estimation sample is calls for service within two years of training when there is virtually no attrition. Standard errors are clustered at the recruit officer level. Column 2 adds controls for recruit characteristics (age, gender, race). We add call characteristics fixed effects (number of officers on the scene, beat, type of call—priority-by-type,year-by-month, and day of the week-by night) in column 3.

Appendix B: Other Field Training Officer Rates

To better understand the mechanism behind our results, and to rule out explanations such as reporting, we calculate field training officer propensity to make arrests, respond to calls in a timely manner, and write up informative reports in a manner similar to our force rate calculations. Namely, we estimate Equation 1 using arrest, misdemeanor arrests, felony arrests, filed arrests, unfiled arrests, response time, and time spent on a call as our outcome. We then shrink our FTO-recruit pair estimates of $\lambda_{o(r)}$ according to Equation 2.

To address whether our results are drive by officer reporting, we also estimate field training officer propensity to write wordy reports. For example, it is reasonable to believe that officers that are more likely to write informative and lengthy reports are also the most likely to report force. Unfortunately we do not have incident reports written by officers for each 911 call. To measure officer wordiness we rely on a separate data set of incident reports. In this data set, we observe 401 of our 411 field training officer-recruit pairs. We attempt to estimate our $\lambda_{o(r)}$'s in a very similar manner, although we do not have exactly the same controls as in our 911 data set. In Equation 1, we control for watch instead of night, and we do not control for the number of officer responding to a call. Despite these differences, we believe our analysis is very similar to the what we perform in the 911 sample.

Finally, we also link field training officers to their complaints. Unfortunately, this data set is the most incomplete of the outcome measures. We only observe type of complaint and the date it was filed (not the date the incident occurred). Therefore, we simply take the number of complaints per FTO before they are assigned to a recruit and divide this by the number of calls their respond to during this time period to calculate our rates.

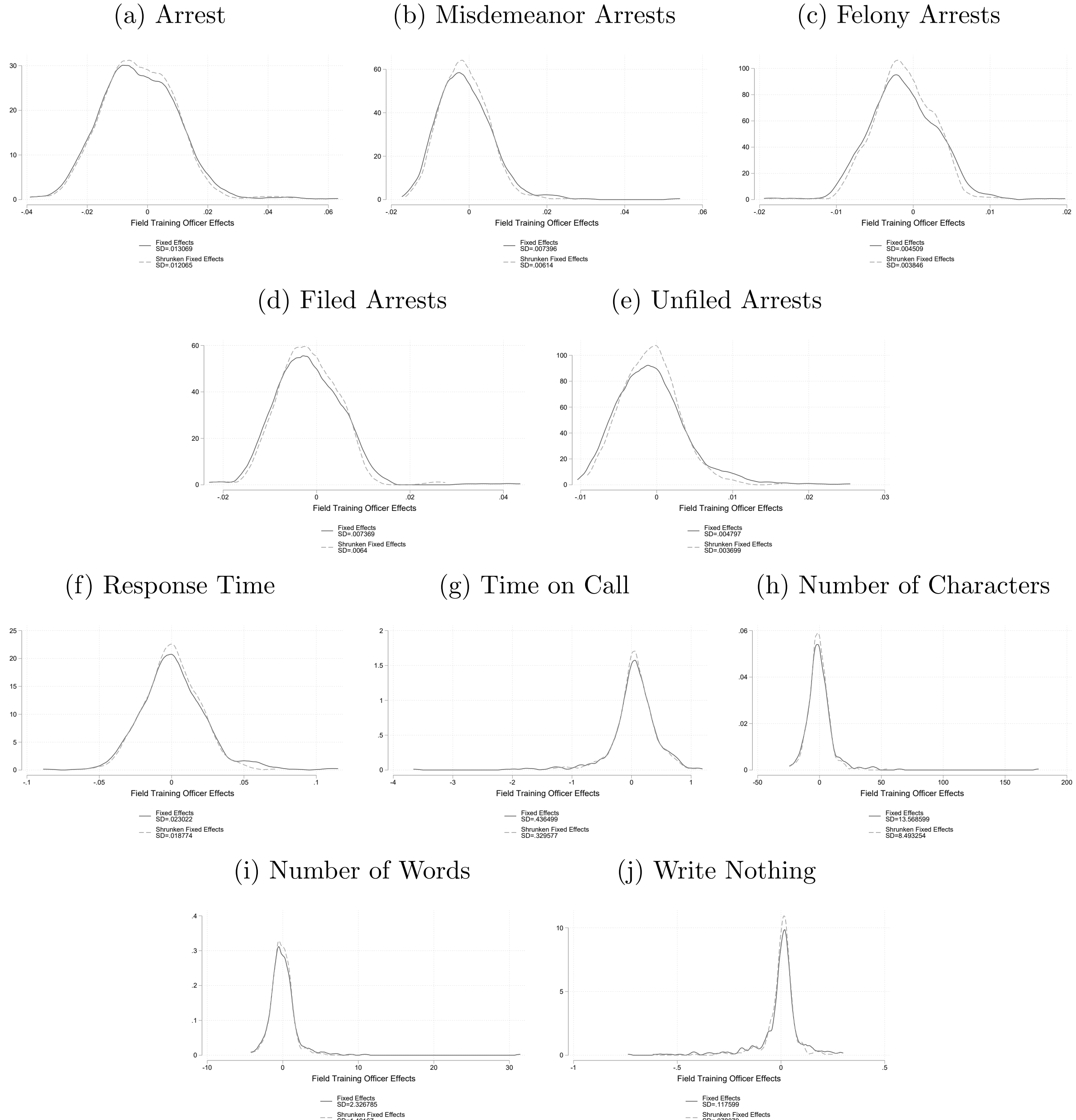
The results of these calculations are show in B.1. Figures B.1a , B.1b, B.1c, B.1d, B.1e show the distribution for our unshrunken and shrunken measures for a field training officer's propensity to make different types of arrests. Both distributions have a longer right tail, indicating that there are some officers with much higher arrest rates than the average field training officer. Further, there is substantial variation in our arrests rates. A one standard deviation increase in officer effects corresponds to a 32% (0.012/.037)and 27% (0.006/ 0.0215) increase in arrest or misdemeanor arrest rates.

Figures B.1c and B.1d show results for our measures of time use (measured in hours). A one standard deviation increase in response time is 1 minute (0.02 hours) or 17% (.02/.12 hours) increase. A one standard deviation increase in time on a call is 0.359 hours or a 150% (.359/.24) increase.

Next, we consider how many words an officer uses when writing up an incident in Figures B.1e, B.1f, and B.1j . Unsurprisingly, the two distributions for number of words and characters look similar. The average number of characters used in an incident report is 43.19976 and there are a few officers that are very wordy. A one standard deviation increase in number of characters used is an increase of 8.5 characters or 19%. The average number of words in a report is 7.31. A one standard deviation increase in wordiness is an increase of 1.39 words or a 20% (1.4/7.31) increase. Finally, we consider an officers propensity to write nothing. On average 10 % of incidents don't have a description. A one standard deviation increase in writing nothing corresponds to a 81% (0.08/0.099) increase in writing nothing. Together these figures show substantial variation in other officer behaviors.

Finally, we consider officer complaints. Complaints are fairly common for FTOs in our sample; 84% of FTOs have at least one complaint. We also consider internal complaints (complaints filed by a DPD employee) and use of force complaints separately. FTOs receive a complaint about once for every 500 calls.

Figure B.1: Other Field Training Officer Rates



Notes: This figure plots the distribution of field training officer effects for arrests, response time, time on a call and measures of wordiness. Response time is the number of hours between arrival time and assigned time. Time on call is the number of hours between the time an officer was enroute and when the call was cleared.