The Effect of Field Training Officers on Police Use of Force

Chandon Adger, Matthew Ross, and CarlyWill Sloan *

January 23, 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 (138 percent) is associated with a 12 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 Jeffery Grogger, Elizabeth Luh (twice!), Aurelie Ouss, Steve Ross, Jamein Cunningham, the Dallas Police Department, Steve Mello, John MacDonald, Shawn Bushway, Anne Piehl, and participants at the 2022 NBER Summer Institute, 2022 Al Capone Conference, 2022 Chicago/LSE Conference, 2021 Southern Economic Association Conference as well as the Association for Public Policy & Management Conference. Sloan and Ross are also 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 grateful to Felipe Goncalves and Emily Weisburst for sharing DPD overtime and shift data with us.

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 that has sought to evaluate the effectiveness of police training has historically relied 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 (McLean et al., 2020; Johnson et al., 2021; 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. Apprenticeship-style models of field training are nearly universal among policing agencies in the United States and consist of an inexperienced recruit, who has graduated the police academy, partnering

with a more experienced officer on patrol for about six months. Although the few studies that have focused on classroom style training have found little to no long-term effects on subsequent enforcement behavior, there are two key reasons to believe that field training might have a larger and more persistent impact. First, the law enforcement community generally characterizes their field training as "the most important stage in the process of becoming an independent police officer" (Warners, Ronald, 2020). Second, recent work by Holz et al. (2020); West (2019); Weisburst (2022) suggests that an officer's on-the-job-experiences are a significant factor in their subsequent enforcement behavior.

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

In our preferred specification, we estimate that a one standard deviation increase (138 percent) in an FTO's prior propensity to use of force is associated with a 12 percent increase in a recruit's subsequent likelihood of using force after training. Distinct from other types of training model 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 that effects persist for at least two and a half years after the completion 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 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 report-writing behavior, and (4) other FTO characteristics which are correlated with force. Last we 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. We argue that these results are generally consistent with the story that less desirable use of force behavior is passed from high force FTOs to recruits.

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 (Crank, 2004; Chan, 1997; Paoline III, 2006; Bruinsma and Weisburd, 2014; 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, Weisburst, 2019, Fryer Jr, 2019, Hoekstra and Sloan, 2020, Rivera, 2022). 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." Pao-

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

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

The two prior works closest to our own are Holz et al., 2020 and Getty et al., 2014. In Holz et al. (2020), the authors examine the impact of an officer's peer being injured on the job and find that other officers are more likely to use force, receive complaints, and injure suspects in the week following their peer's injury. We differ from Holz et al., 2020 by considering a fundamentally different determinant of force (i.e. FTO force propensity). Relative to a

peer getting injured on the job, we find that one's FTO has a substantially longer impact on subsequent policing behavior. Namely, we document higher use of force for officers trained by more aggressive officers for as much as three years after the completion of their training. In Getty et al., 2014, the authors also use data from the Dallas Police Department to study the impact of FTOs on subsequent allegations of misconduct. Using hierarchical modeling, the authors find that 26.5% of the overall variation in a recruit's subsequent complaints are related to their FTO. They do not ask whether having an FTO with more complaints causes a recruit to acquire more complaints. Relative to Getty et al., 2014, we implement a more rigorous empirical design which 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.

The findings of this paper have several important policy implications. First, by demonstrating that FTOs are an important determinant to a recruit's subsequent propensity to use force, we have identified a particularly promising target for policy interventions aimed at reducing aggregate rates of police use of force. To our knowledge, there has never been any targeted interventions aimed at substituting high for low-force FTOs in an effort to reduce aggregate rates of police force. Given our findings and the fact that classroom based procedural justice interventions (McLean et al., 2020; Owens et al., 2018; Wheller et al., 2013) have lead to large but only short-lived effects on enforcement behavior, there is reason to believe that reforms to field training might have larger and longer-lasting impacts on aggregate force. [1] Second, our findings are conservative in that we do not attempt to explore the dynamic effects of changing the composition of FTOs. In particular, our findings suggest that selecting less-forceful FTOs in any given year would lead to less forceful recruits who

[1]Procedural justice refers to the idea of fairness in the processes that resolve disputes and allocate resources. It is a concept that, when embraced, promotes positive organizational change and bolsters better relationships. Procedural justice speaks to four principles, often referred to as the four pillars: fairness in the processes, transparency in action, opportunities for voice, impartiality in decision making (COPS, 2020). There is also 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.

will then subsequently become FTOs in the future. Thus, the dynamic effects of altering the composition of FTOs in any given year will potentially have large effects on aggregate force that will last very far into the future. Third, changing the composition of FTOs is a much more cost-effective (even potentially costless) intervention relative to alternatives that involve classroom or simulation based training. Finally, reforms aimed at changing the composition of FTOs are much more practical and politically palatable relative to large-scale initiatives such as defunding the police, dramatically reducing enforcement policies, or attempts to substantially alter the composition of the police force.

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). 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 a full-fledged police officer, i.e. Academy training (phase 1) and field training (phase 2).[2] 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

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

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 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) which 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 which 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 more a experienced officer (i.e. their FTO) in, what could be characterized as, 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.

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. Here, 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 phase, the FTO is different. The first three training phases of field training are seven weeks long. The final evaluation phase is conducted by the initial (i.e. phase one) FTO and lasts three weeks. When field training begins, recruits are instructed to take on a more observational role. As training progresses, they are given more autonomy and become an active participant in responding to calls for service. For example, in the early weeks of field training, a recruit may simply watch a FTO respond to a call for service. In later phases, the recruit may lead the response under the guidance and observation of their FTO. FTOs also conduct frequent, often daily, evaluations of recruits. According to command staff in Dallas, these evaluations are largely used to provide the recruit with extensive feedback on their performance.

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

This paper focuses on the impact that the first FTO has on the recruit's subsequent

enforcement behavior. We made 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 career. 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.[3] We also focus on FTOs rather than the officer a recruit chooses to work with during "little t" because FTOs are conditionally randomly assigned. In contrast, a recruit may select their partner during "little t."

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

[3]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.

the dispatch of multiple officers. Once an officer responds to a call, the officer is afforded a significant amount of discretion in how they handle an incident in terms of their decision to make an arrest or use force.

To measure officers use of force, we link 911 calls to force reports.[4] 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[5] All force incidents are reviewed by a supervisor (Dallas Police Department General Orders, 2015). According to the Dallas Police "The physical control techniques used may range from the use of handcuffs in an arrest, strikes with an impact weapon, or the use of a firearm" (Dallas Police Department, 2019). 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. 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. 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. [6]

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

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

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

3 Data and Summary Statistics

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 reports, arrest records, Dallas County District Attorney records, and officer characteristics.[7][8] 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.[9] 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

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

[8] 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.

[9] 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.

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 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.[10] 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.[11] 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.[12] 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

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

[11] We estimate coefficients of 0.0002068***, 0.0002038***, and 0.0001519** for our Table 4 specifications.

[12] 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.

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, 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.[13] 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. 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

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

[13]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 0.84.

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.[14] 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 is 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 join 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

[14] 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.[15]

Formally, we estimate the across officer variance in FTO force, σ_A^2 and a within-officer variance, σ_W^2 [16] 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.[17]

The distribution of standardized effects is shown in Figure 1b. One standard deviation increase in FTO effects is a 0.1706 percentage point which represents a 138 percent increase in average use of force compared to the sample mean of 0.1234 percent. Moving from the FTO that used the least amount of force to the most is an approximate 6 standard deviation increase. Replacing an FTO at the 10th percentile for one at the 90th percentile represents a 338 percent increase in average force.

Finally, we compare FTO force rates to the force rates of other patrol officers in Dallas.

[15] We can also force FTOs to have the same number of calls before being assigned a recruit (2000 calls). Our main results (Table 4) are similar in magnitude if we make this restriction and re-estimate force rates. The limitation of this method is that we must limit our sample of FTOs.

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

[17] We note that our main results are robust to using the unshrunken estimates as well as a number of alternative specifications including a binary indicating high/low force FTOs (see Table A.4).

To do so, we first construct a force rate for each officer using our entire sample of calls for service. Next, we shrink and standardized the force rates as described above. Our results are shown in Figure A.2.[18] On average, FTOs use force more, 0.45 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 1. As noted, there are 411 recruits in our sample. This translates to roughly 90 new recruits each year. The average recruit is much younger than the average FTO. Most recruits are White (44 percent), 21 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. A t-tests of the difference in recruit characteristics across high and low force FTOs is not statistically significant.[19]

In our main analysis, we evaluate recruit behavior after field training using data on their subsequent calls for service. Summary statistics at the call level are presented in Table 2. In our sample, roughly 3.7/100 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

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

[19]We estimate p-values of 0.5147, 0.3586,0.3399 , 0.1800, 0.1645 for test of difference across means in race, gender and age.

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.[20] Our call data also includes other important characteristics that may impact police officer behavior on the scene. Specifically, we observe the call type, priority (a measure of urgency and severity), location, date and time.

4 Empirical Methods

4.1 Estimation Model

The conditional random assignment of 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 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 associated with 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

[20] 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.

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.[21]

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.

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.[22] 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

[21] 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.

[22] 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.

investigate whether FTO age, race, and force rate are correlated with recruit age, race, gender, and hire date (measured in years). The results of this test are shown in Table 3. Of the 35 coefficients reported, only one is statistically significant at conventional levels.[23] 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.

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,

[23] 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 three coefficients and none of the F-test p-values in the tables are significant at conventional levels. See Table 3 and Table A.2.

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 (138 percent) in the FTO's propensity to use force will lead to an increase in recruit force by 0.025 percentage points or approximately 20 percent.[24]

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.[25] [26] 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

[24]The p-value from this regression is less than 0.001

[25]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.

[26]Coefficient on linear fit is 0.0046 percentage points with a p-value of 0.116 . The correlation coefficient for the two terms is 0.0127.

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 4. Each specification includes cohort year by division fixed effects, and standard errors are clustered at the recruit level. Our results are also robust to two-way clustering at the recruit badge and FTO level, as well as recruit and division-by-cohort year level.[27] 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 assigned to an FTO with one standard deviation higher force propensity (138 percent) . Our results show that recruits with FTOs that use force one standard deviation more are 0.0199 percentage points or 16 percent more likely to use force.[28]

In column 2, we add controls for recruit characteristics (age, gender, race). Given our conditional random assignment and the results in Table A.3 and Figure A.3, we would not expect recruit characteristics to alter our estimate meaningfully. Column 2 shows that

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

[28] In Table A.4 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. We also explore using a binary treatment variable, and the results are similar but not always statistically significant at conventional levels.

recruits with FTOs that use force one standard deviation more are 0.0205 percentage points or 17 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 24 percent smaller but still economically meaningful and statistically significant at the 5 percent level. Even after controlling call characteristics and holding constant this alternative channel, we find a one standard deviation increase (138 percent) in FTO force increases recruit force by 0.015 percentage points or 12 percent. We interpret these and the prior results as providing additional evidence in support of 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.

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 that associated with 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%.

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. (2020) 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.[29] 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, we find very similar results to Table A.5. Given both of these results, we believe that we can confidently put aside potential concerns of attrition bias driving our results.

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 4. In our main estimates, we follow standard approaches by clustering our standard errors at the recruit level and note that we are also robust to two-way clustering at the recruit and division by year level (Bertrand et al., 2004). The concern motivating the robustness test in this section is that our outcome variable (force by call) is a rare event, occurring in only 0.106 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,

[29] Specifically, we regress the last date we observe a recruit in the calls data on FTO force rate using the specifications shown in column 1, 2, and 3 of Table 4. None of the coefficients on FTO force were statistically significant at conventional levels (p-values = 0.364, 0.166, 0.147).

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. [30] 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 propensity to use force in the pre-period as the primary independent variable, we estimate the model from columns 1-3 of Table 4. In order to obtain p-values for a two-sided hypothesis test, we replicated this procedure 1,000 times and calculate the share of the simulations when the t-statistic exceeds the absolute value of the t-statistics from Table 4.

For the models corresponding to columns 1 and 2 of table 4, we obtain p-values for a two-sided hypothesis test using randomization inference of 0.01 and 0.013 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 4. In our randomization procedure, we find that eighty of the simulations result in a t statistic more extreme than 2.137. Thus, the associated p-value obtained through randomization

[30] 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.

inference is 0.08. 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 incorrect asymptotic assumptions or rare-event bias.

6 Mechanism and Recruit Arrests

6.1 Alternative Mechanisms

Until this point, we have interpreted our main results an FTO transferring information about the appropriate threshold for applying force to their recruit, i.e. an FTO teaching a recruit to behave more/less aggressively. In this section, we will provide additional evidence in support of that particular explanation of the underlying mechanism by convincingly ruling out six alternative explanations. These alternative explanations include: (1) omitted FTO characteristics are correlated with FTO force and a recruit's subsequent force; (2) whether our main results 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 they complete training; (4) whether recruits paired with a high-force FTO are actually 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 find 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.1, 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.[31] Figure A.1b shows that female FTOs are nearly half a standard deviation more likely to use force than males and these two distributions are

[31] The average Black FTO has a force rate that is about 1/10 of a standard deviation higher 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

statistically different from each other (K-Smirnov p-value = 0.001). Figure A.1a also shows that younger FTOs (i.e. less than the mean age of 50) use force 0.15 standard deviations more frequently than older FTOs and the two distributions are statistically different from each other (K-Smirnov p-value = 0.09).[32] Finally, we regress our force measure on FTO race, age, and gender. This regression yields an R-squared of 0.05 and an F-test p-value of 0.02. 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 4 (shown again in column 1) in columns 2, 3, and 4 of Table 6. 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.[33] We begin by presenting the correlation between FTO force, and other FTO rates in Table 5. Column 1 in Panel A reports that a one standard deviation increase in FTO force propensity is associated with a 0.58 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. 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 and time on call in columns 5-8 of 6. Overall, these estimates are similar in

[32]We also regress our force rate on a FTO age and find that 10 years of age translate to a 0.11 standard deviation decrease in an officer's force rate, however this estimate is not significant at conventional levels.

[33]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).

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 5-8) in column 9 of 6. 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 4 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 7. 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 4 and significant at conventional levels. We note that this approach is not our preferred specification as who a recruit chooses to work could be affected by treatment. Said another way, although including partner characteristics as controls is technically an over-control, we believe it 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. 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 6 shows that results are similar no matter if a recruit experienced force during their training period. Within each specification (following specifications in Table 4), 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.[34] To explore differences in report-writing behavior, we develop three measures based on the section of incident reports completed by the responding officer. [35] 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 5, we show that our proxies for reporting writing are not correlated with FTO force. In Table 9, we repeat the three columns presented in Table 4 but with the addition of our reporting controls. Across each column each of the estimates are similar in magnitude and significance to Table 4.[36]

Finally and relevant to the prior discussion, we again explore under-reporting by focusing on a subset of the most "reportable" force incidents, i.e. those where a weapon was discharged.[37] A limitation of our data set is that the information on the type of force used

[34]The police department's General Orders reiterate the auditing of reports and bodyworn/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.

[35]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.

[36]There are 10 FTOs that we cannot link to incident reports. Therefore we have approximately 10 thousand fewer observations in Table 9 relative to our primary estimates.

[37]Specifically we flag force incidents where a firearm, taser, spray, or pepperball was used.

is missing for 70% of our linked force incidents. For the following analysis we drop those force incidents with missing information. When replace our outcome variable with one for *only* the most severe force incidents, our results are remarkably similar (see Table A.6).[38] Further our results are not driven by events with only one officers, where there is arguably less incentive to report.[39]

6.2 Arrest Results

Thurs far, our paper has demonstrated that recruits assigned to a high-force FTO are more likely to use force later in their careers. One limitation of most existing work on police use of force, is that it is arguably impossible for the researcher to determine which interactions or calls for service warrant use of force (a"good" application of force) and which do not (a "bad" application of force). Put differently, there are clearly some dangerous incidents when we want to empower police to use force in the name of preserving public safety. In an effort to explore whether high-force FTOs are engaged in other behaviors that might be considered "good" vs. "bad" from a social or policy perspect, we consider four additional measures in Table 8. Panel A presents results for arrests that are arguably less desirable from a social/policy perspective, and Panel B presents results for more serious and higher quality arrests. Our results follow the columns presented in Table 4 where we demonstrate that an increase of one standard deviation in FTO force was associated with a 3-5% increase in the likelihood of a misdemeanor arrest and a 3-7% increase in the likelihood of an unfiled arrests. Unfiled arrest are arrests that are made, but not formally filed at the District Attorney's office. This typically happens when an arrest is made for frivolous reasons or without proper evidence or supporting documentation. In Panel B, we find that recruits with higher force FTOs do not make more felony arrests or filed arrests. Together, Panels A and B provide

[38] Given how rare our outcome measure is, we shift from beat to sector fixed effects, and estimate fixed effects for whether a recruit is above average age (instead of a fixed effect for each year).

[39] Our results are still statistically significant at conventional levels and larger when we drop calls with only one officer.

evidence that high-force FTOs not only lead to high-force recruits but that they also produce recruits that make more petty arrests and unfiled "low quality" arrests but not more serious "high quality" felony arrests. In short, we find very little evidence in support of the idea that high force is associated with the types of behaviors we might characterize 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 12 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.

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%20Documents/General-Orders.pdf.

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

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

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

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

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

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

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

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

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

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-misco`

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

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

Rivera, R. (2022). 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.

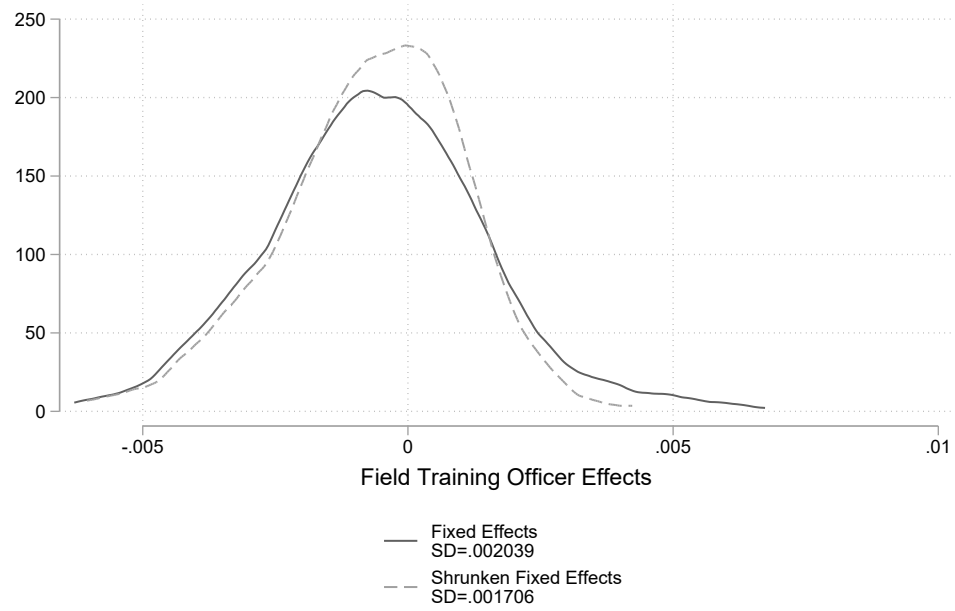
Woody, R. H. (2005). The police culture: Research implications for psychological services.

Professional Psychology: Research and Practice 36(5), 525.

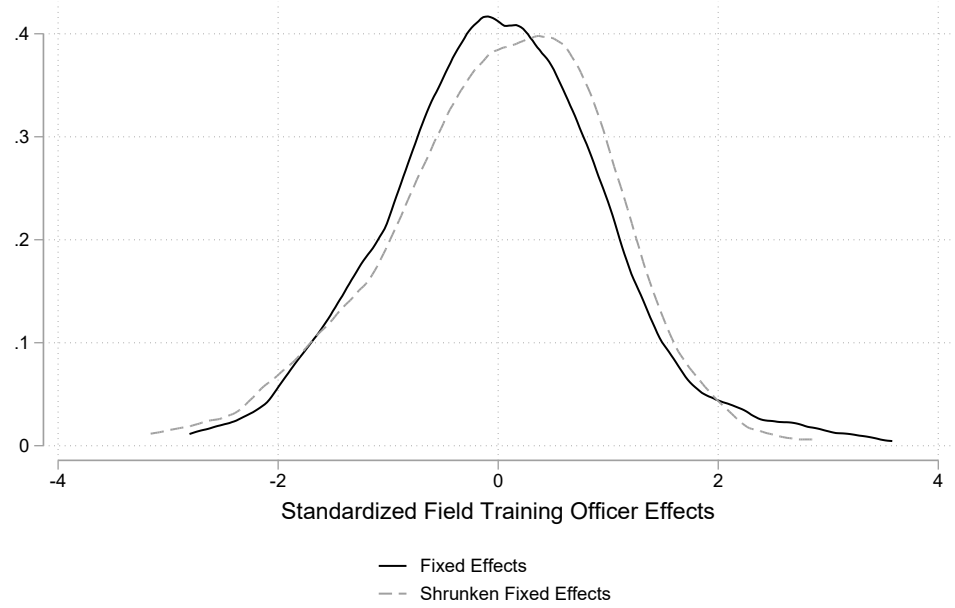
Figures

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

(a) Field Training Officer Effects



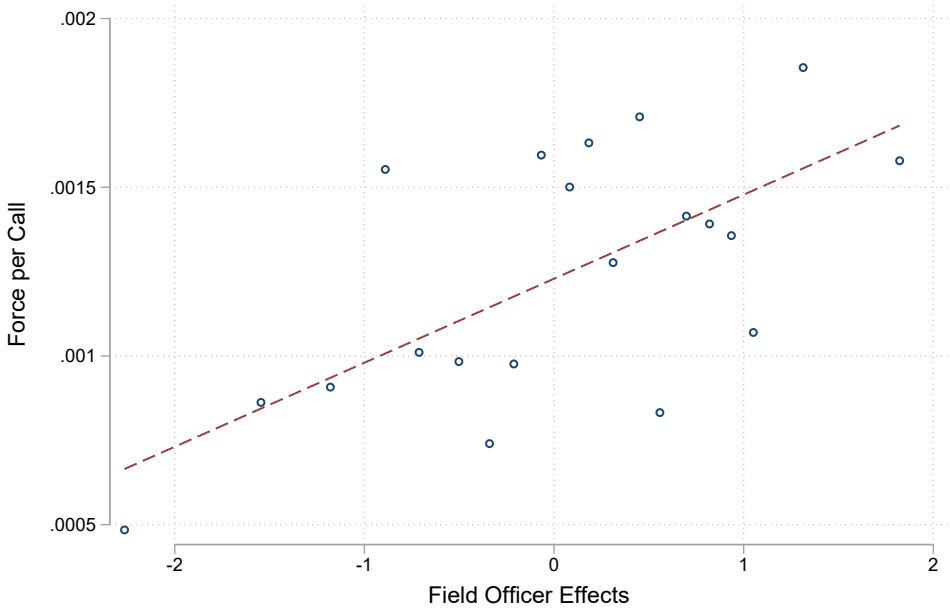
(b) Standardized Field Training Officer Effects



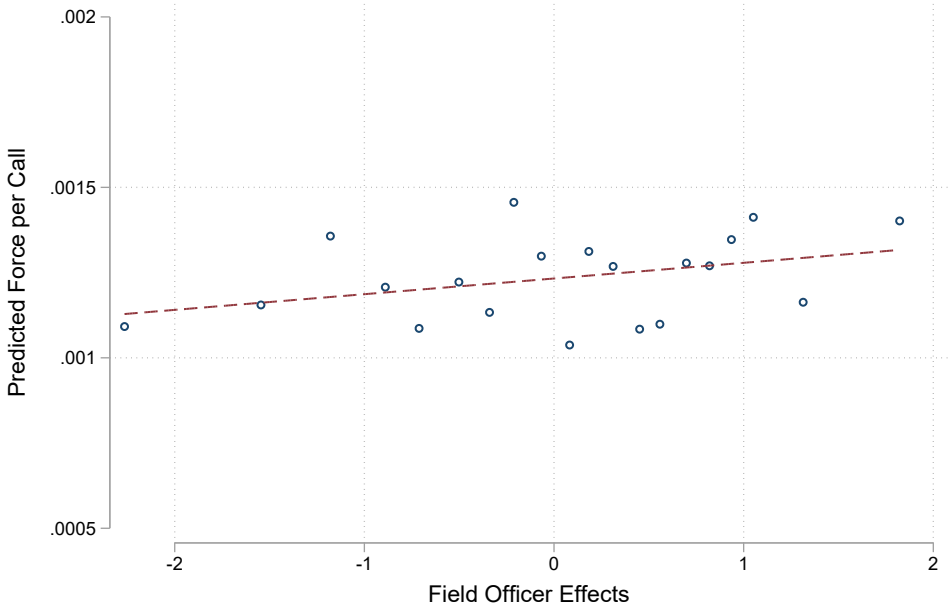
Notes: These figures plot the distribution of field training officer fixed effects. Fixed effects are calculated after accounting for the number of officers on the scene, beat, type of call (priority-by-type) year-by-month, and day of the week-by night fixed effects.

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

(a) Use of Force

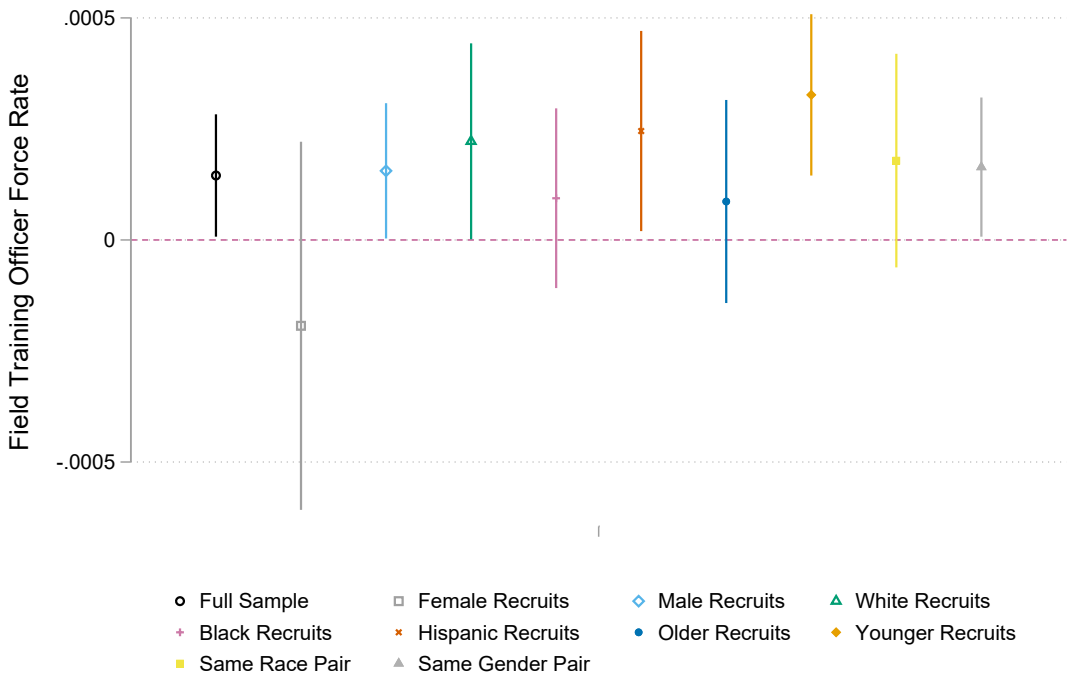


(b) Predicted Use of Force

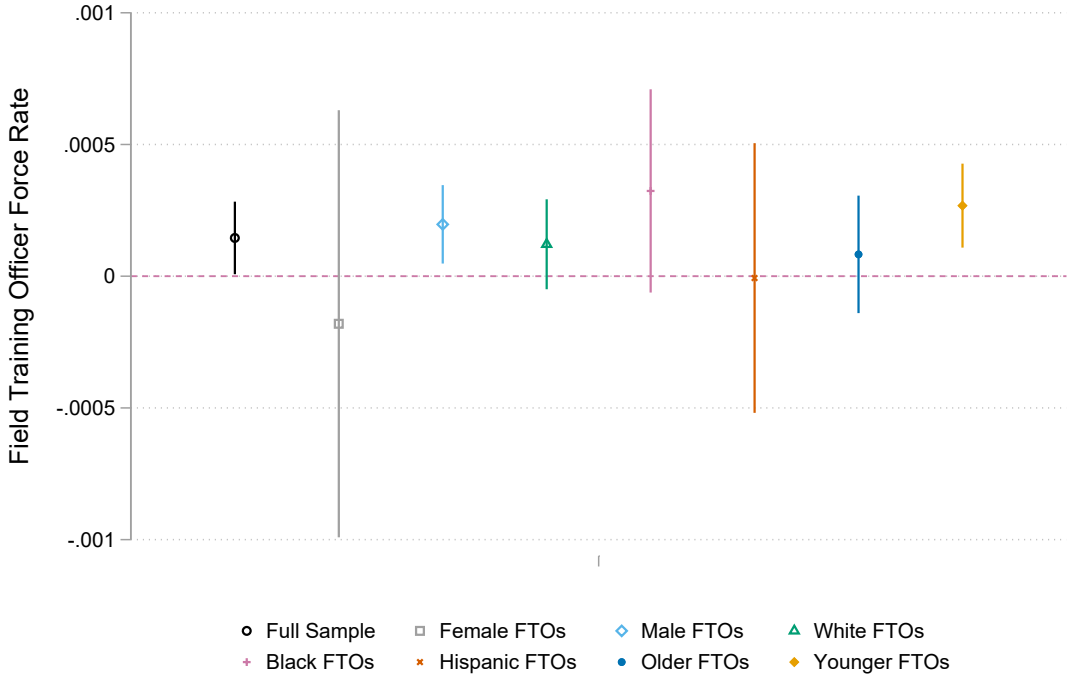


Notes: In Panel (a) we plot use of force. In Panel (b) we plot predicted use of force. The fitted line is a linear fit across all use of force rates. Observations are grouped so that each point includes an equal number of observations. Use of force is predicted using the number of officers on the scene,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

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



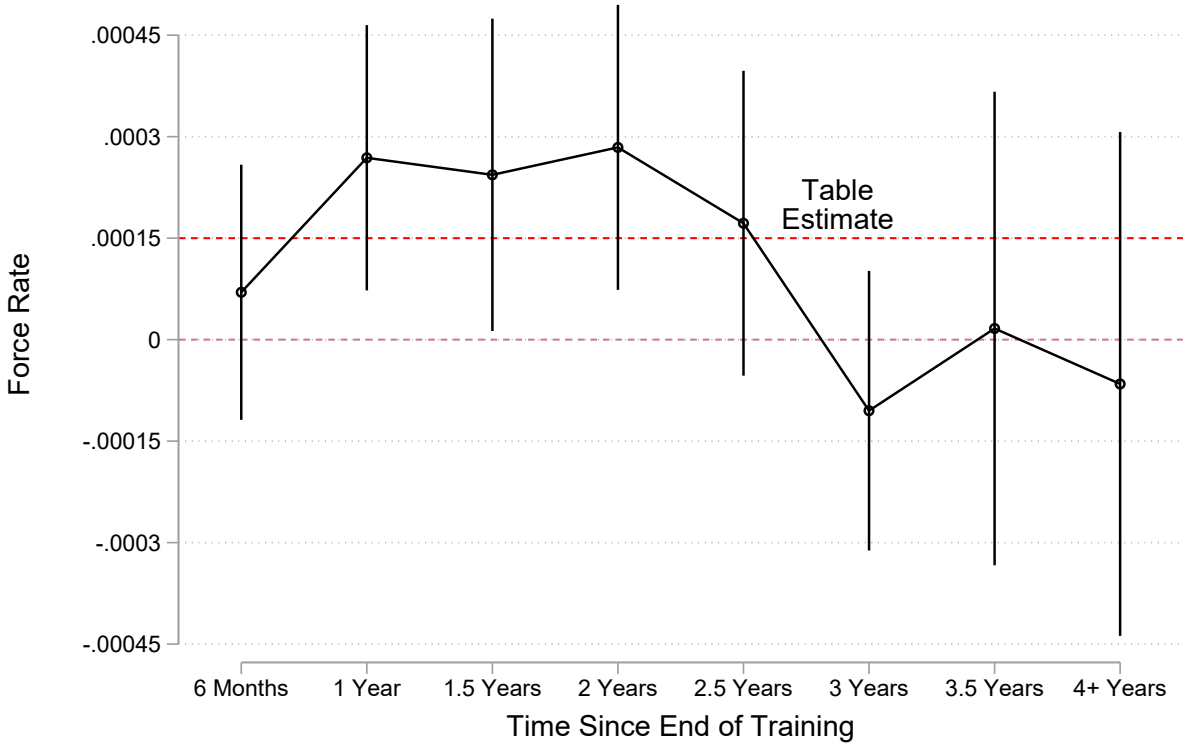
(a) Recruit Subgroups



(b) FTO Subgroups

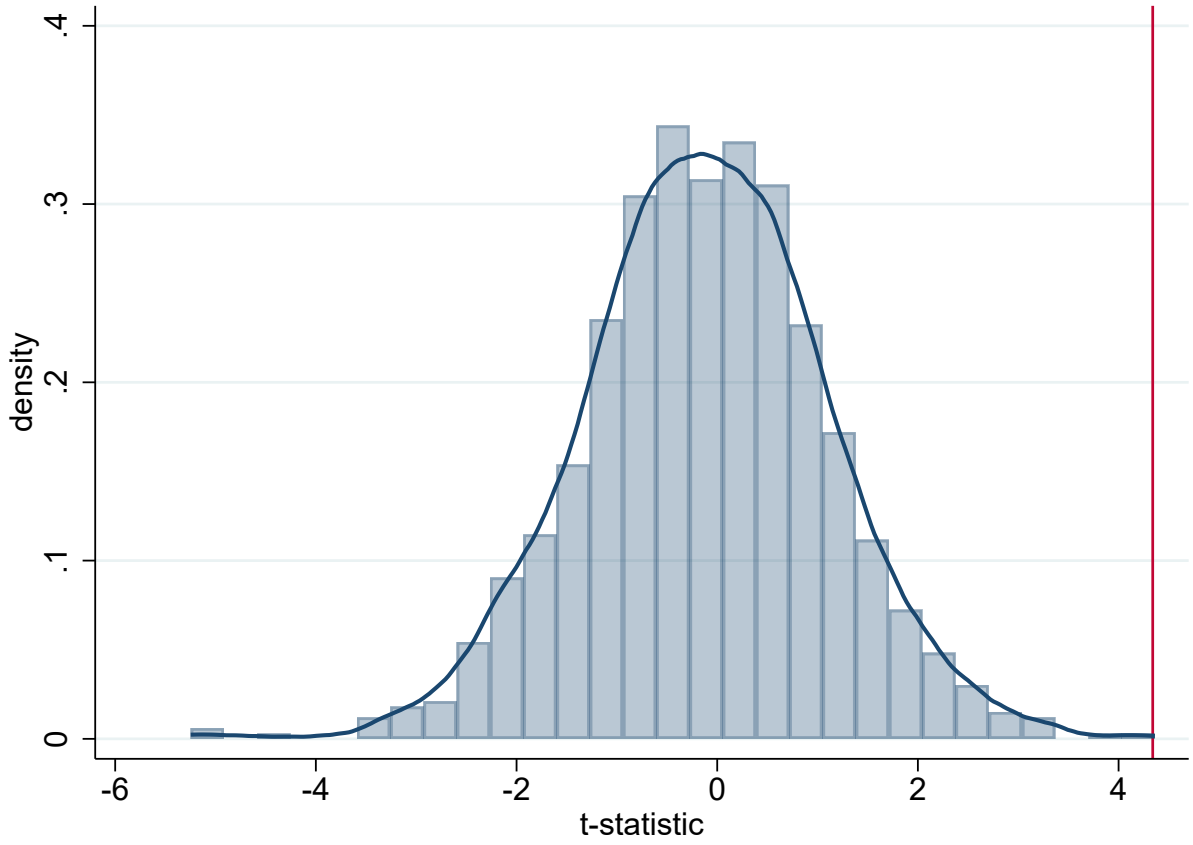
Notes: This figure reports the effect of field training officer force rates by recruit and FTO subgroup. Each coefficient is from a separate regression. Standard errors are clustered at the recruit level.

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



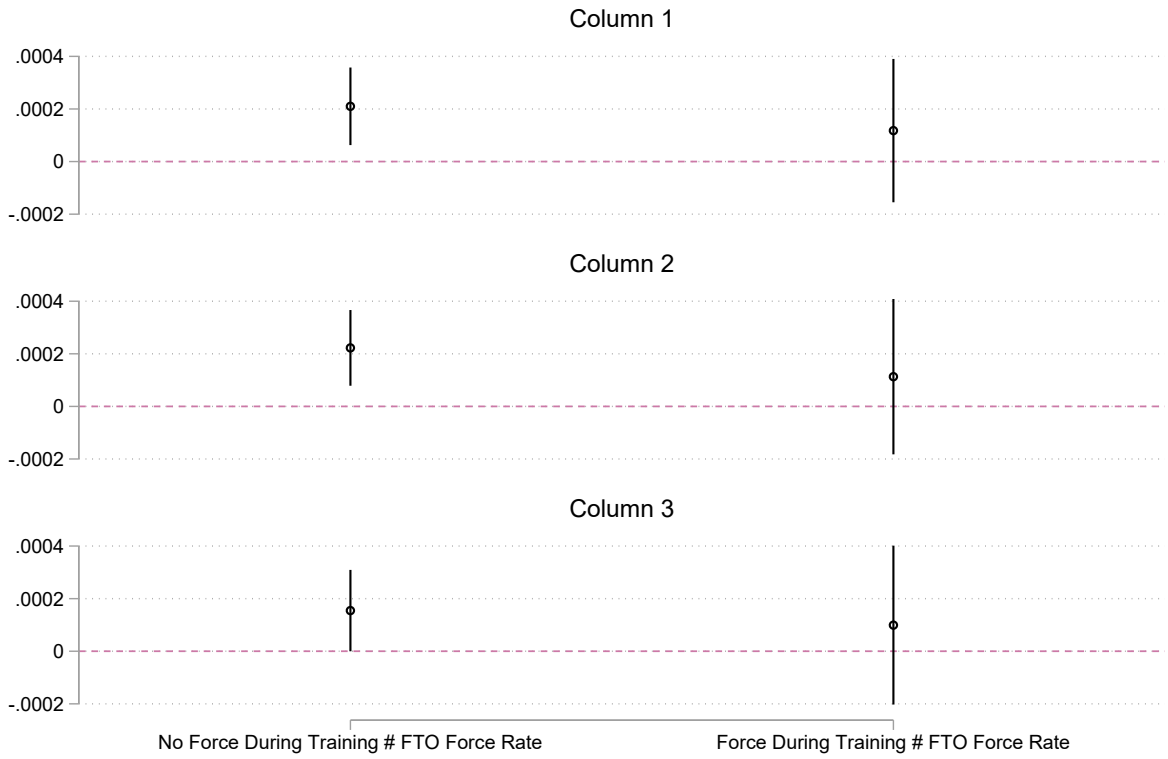
Notes: This figure reports the effect of field training officer force rates over time. Sample includes all officers observed with three full years of data. The red line marks the estimate from column 3 in Table 4. Standard errors are clustered at the recruit level.

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 4 by using a 1,000 randomized simulations of the data generating process. Eighty of the simulations resulted in a t-statistic more extreme than our baseline estimate of 2.137 indicating a p-value of less than 0.08.

Figure 6: 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 4.

Tables

Table 1: Officer-level Summary Statistics

	(1) Entire Sample	(2) High Force Trainer	(3) Low Force Trainer
White	0.438	0.418	0.451
	(0.497)	(0.495)	(0.499)
Black	0.207	0.184	0.221
	(0.406)	(0.388)	(0.416)
Hispanic	0.302	0.329	0.285
	(0.460)	(0.471)	(0.452)
Female	0.178	0.146	0.198
	(0.383)	(0.354)	(0.399)
Age	35.84	36.32	35.55
	(5.466)	(5.948)	(5.132)
Observations	411	158	253

Table 2: Call Level Summary Statistics

	(1) Entire Sample	(2) High Force Trainer	(3) Low Force Trainer
Force	0.00123	0.00141	0.00112
	(0.0351)	(0.0375)	(0.0334)
All Arrests	0.0370	0.0374	0.0367
	(0.189)	(0.190)	(0.188)
Misd. Arrest	0.0215	0.0221	0.0211
	(0.145)	(0.147)	(0.144)
Felony Arrest	0.00755	0.00755	0.00754
	(0.0865)	(0.0866)	(0.0865)
Filed Arrest	0.0148	0.0149	0.0148
	(0.121)	(0.121)	(0.121)
Unfiled Arrest	0.0143	0.0149	0.0139
	(0.119)	(0.121)	(0.117)
Observations	1085020	421901	663119

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				FTO Characteristics			
	Age	Female	White	Black	Hispanic	Hire Date	Force Rate
Recruit Chars							
White	-1.0720	-0.1124	-0.1555	-0.0052	0.0952	281.5651	-0.3147
	(1.8062)	(0.0879)	(0.1044)	(0.0831)	(0.0728)	(758.3525)	(0.2124)
Black	-1.4251	-0.0514	-0.0800	0.0090	0.0061	334.1880	-0.4693**
	(1.9980)	(0.0971)	(0.1141)	(0.0921)	(0.0793)	(797.1520)	(0.2370)
Hispanic	-1.3049	-0.0561	-0.1512	0.0158	0.0730	418.5266	-0.3657
	(1.8585)	(0.0916)	(0.1088)	(0.0888)	(0.0754)	(765.9610)	(0.2304)
Female	-0.1773	0.0293	-0.0923	0.0502	0.0560	289.7770	-0.1121
	(1.0851)	(0.0472)	(0.0648)	(0.0523)	(0.0537)	(377.1587)	(0.1179)
Age	0.0502	-0.0000	0.0027	0.0003	0.0004	10.6934	0.0051
	(0.0894)	(0.0035)	(0.0051)	(0.0041)	(0.0040)	(30.9955)	(0.0098)
Observations	411	411	411	411	411	411	411
Div-x-Cohort FE	X	X	X	X	X	X	X
Outcome Mean	49.45	0.129	0.628	0.170	0.165	15298	-1.51e-09
F-Test P-Value	0.972	0.482	0.357	0.942	0.359	0.941	0.391

Notes: Robust standard errors are presented.

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

	(1) Force	(2) Force	(3) Force
FTO Force Rate	0.000199***	0.000205***	0.000150**
	(0.0000663)	(0.0000651)	(0.0000702)
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: 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. A one standard deviation increase in FTO use of force is a 138 percent increase.

Table 5: 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	0.584***	0.596***	0.453***	0.498***	0.625***
	(0.0551)	(0.0516)	(0.0557)	(0.0541)	(0.0533)
Observations	411	411	411	411	411

	Response Rate	Time on Call
Panel B:		
FTO Timing		
Force Rate	-0.0299	-0.168***
	(0.0743)	(0.0473)
Observations	411	411

	Num. Characters Rate	Num. Words Rate	Write Nothing Rate
Panel C:			
FTO Reporting			
Force Rate	-0.119*	-0.106	0.0317
	(0.0698)	(0.0670)	(0.0701)
Observations	401	401	401
Div-X-Cohort FE	Y	Y	Y

Standard errors in parentheses

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

Notes: 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 C 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. A one standard deviation increase in FTO use of force is a 138 percent increase.

Table 6: **Mechanisms:** 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	0.000150**	0.000158**	0.000159**	0.000237***	0.000255***	0.000217***	0.000151**	0.000136*	0.000337***
	(0.0000702)	(0.0000715)	(0.0000695)	(0.0000743)	(0.0000838)	(0.0000793)	(0.0000707)	(0.0000709)	(0.0000790)
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	Y	Y	Y
Call Controls	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

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

Notes: Standard errors are clustered at the recruit officer level. A one standard deviation increase in FTO use of force is a 138 percent increase.

Table 7: **Partner Controls**: The Effect of High Force Field Training Officers on Recruit Use of Force

	(1) Force	(2) Force	(3) Force
FTO Force Rate	0.000170***	0.000165***	0.000110*
	(0.0000532)	(0.0000450)	(0.0000571)
Observations	1083209	1083209	1083209
Outcome Mean	0.00123	0.00123	0.00123
Assigned Div by Cohort FE	Y	Y	Y
Recruit Characteristics	-	Y	Y
Call Controls	-	-	Y
Partner Controls	Y	Y	Y

Standard errors in parentheses

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

Notes: 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.

Table 8: 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						
Force Rate	0.00118**	0.00103**	0.000568*	0.00105**	0.000913***	0.000441**
	(0.000484)	(0.000412)	(0.000290)	(0.000411)	(0.000339)	(0.000204)
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						
Force Rate	0.000210	0.000224	0.0000871	0.000360	0.000365	0.000235
	(0.000136)	(0.000130)	(0.000113)	(0.000267)	(0.000241)	(0.000227)
Outcome Mean	0.00755	0.00755	0.00755	0.0148	0.0148	0.0148
Observations	1085020	1085020	1085020	1085020	1085020	1085020
Assigned Div by Cohort FE	Y	Y	Y	Y	Y	Y
Recruit Characteristics	-	Y	Y	-	Y	Y
Call Controls	-	-	Y	-	-	Y

Standard errors in parentheses

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

Notes: Standard errors are clustered at the recruit officer level. Columns 2 and 4 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 columns 3 and 6. Unfiled arrests are arrests that are not filed with the Dallas District Attorney's Office.

Table 9: **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	0.000232***	0.000237***	0.000186***	0.000231***	0.000237***	0.000185***	0.000220***	0.000225***	0.000174**
	(0.0000680)	(0.0000648)	(0.0000711)	(0.0000677)	(0.0000646)	(0.0000708)	(0.0000670)	(0.0000636)	(0.0000697)
Observations	1074474	1074474	1074474	1074474	1074474	1074474	1074474	1074474	1074474
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

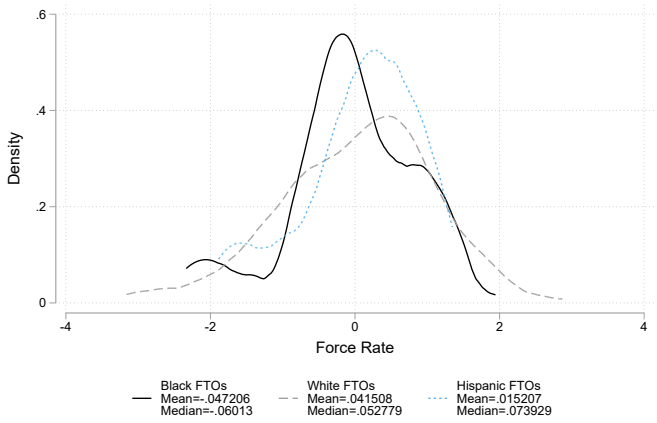
* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: 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 1074474 instead of 1085020. A one standard deviation increase in FTO use of force is a 138 percent increase.

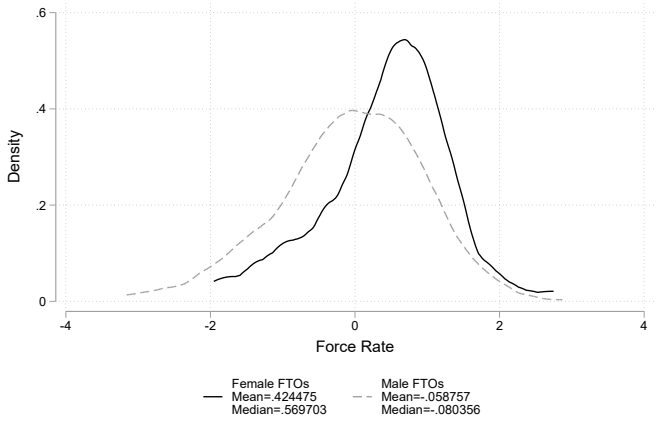
A.2 Appendix

Figure A.1: Density of Field Training Officer Propensity to Use Force by Field Training Officer Characteristics

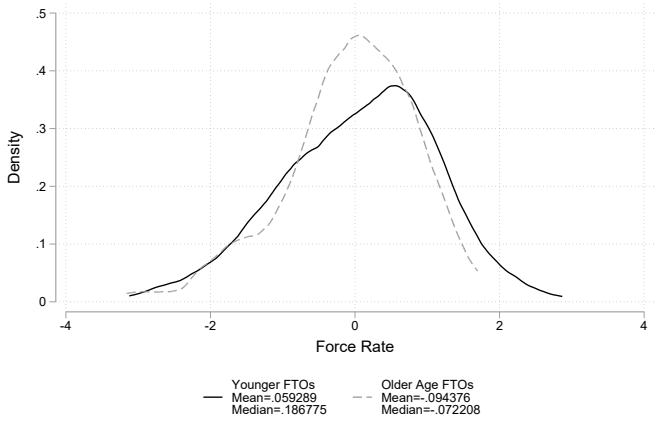
(a) Field Training Officer Race



(b) Field Training Officer Gender

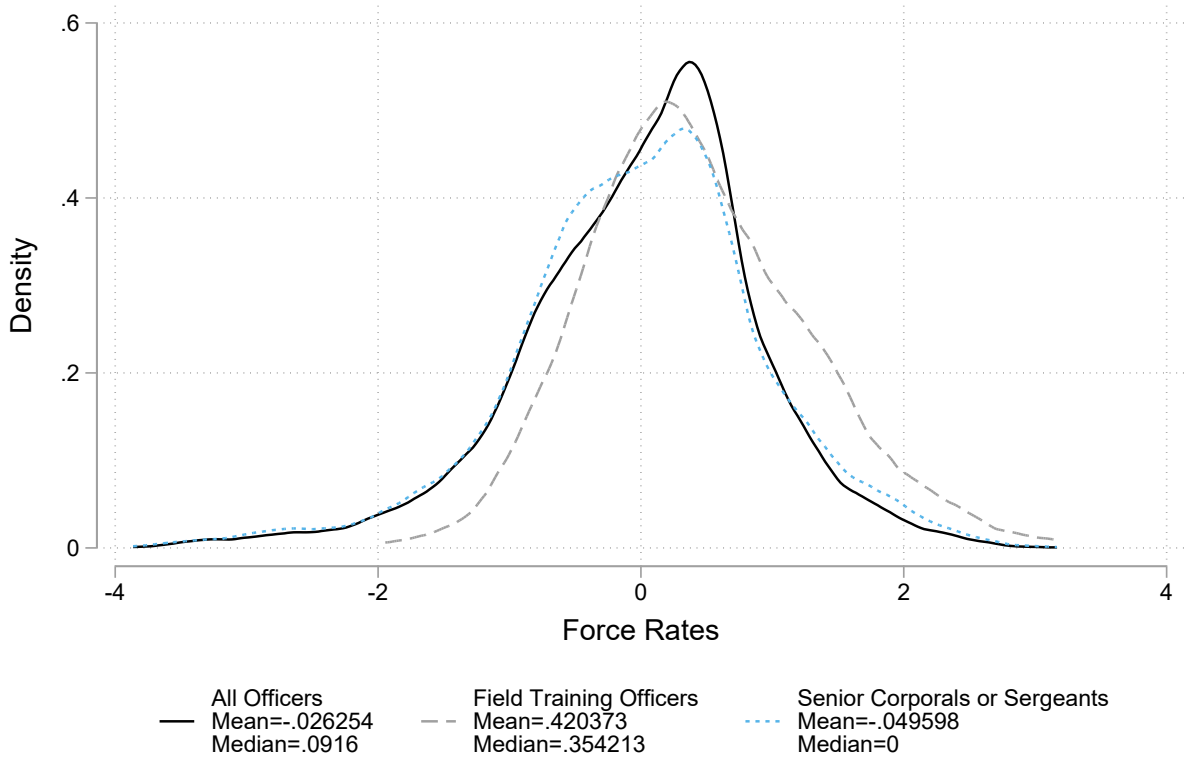


(c) Field Training Officer Age



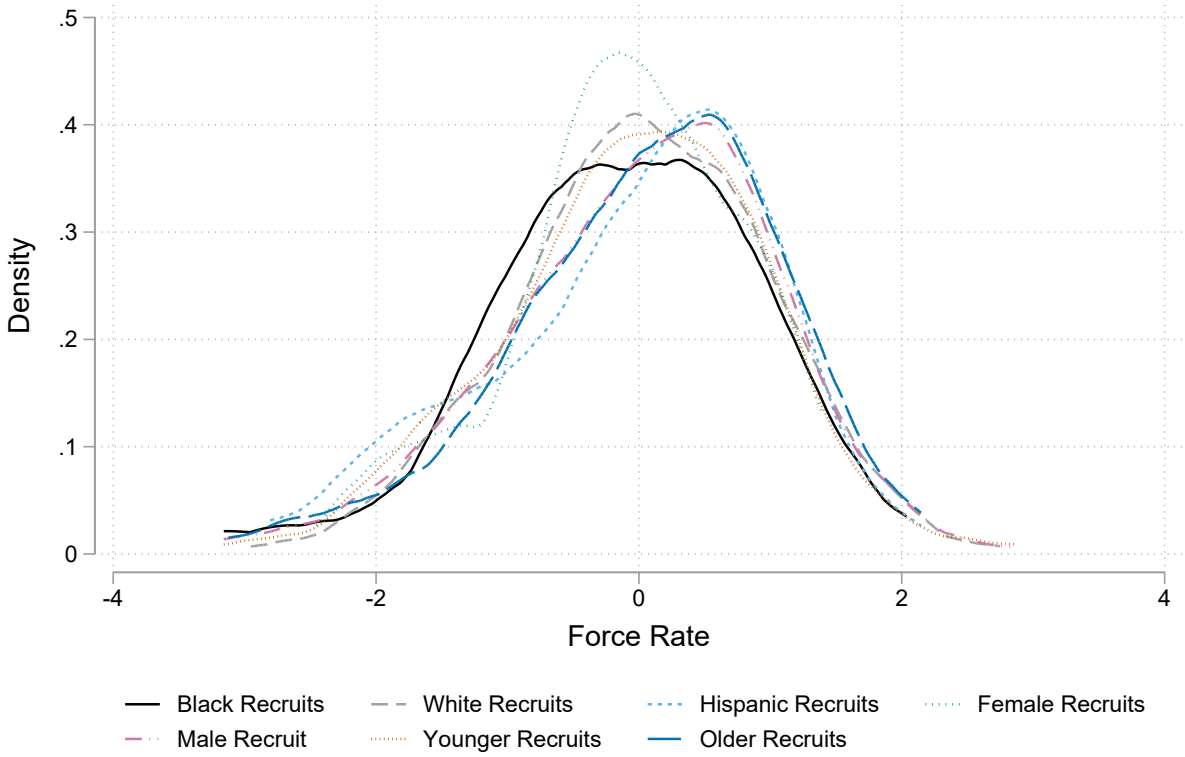
Notes: This figure plots the distribution of field training officer effects by field training officer characteristics.

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: This figure plots the distribution of field training officer effects by recruit characteristics. Older recruits are recruits older than the average age (36 years old).

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	0.000199***	0.000207***	0.000154**
	(0.0000673)	(0.0000666)	(0.0000711)
FTO 2 Force Race	0.0000562	0.000112*	0.000101
	(0.0000627)	(0.0000643)	(0.0000666)
FTO 3 Force Race	-0.0000229	0.00000277	0.0000151
	(0.0000673)	(0.0000716)	(0.0000713)
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: Standard errors are clustered recruit level.

Table A.2: Balance Test: Correlation between Recruit and Field Training Officer Characteristics

	(1) Age	(2) Female	(3) White	(4) Black	(5) Hispanic	(6) Hire Date	(7) Force Rate
White	-1.0720	-0.1124	-0.1555	-0.0052	0.0952	281.5651	-0.3147
	(1.8206)	(0.0836)	(0.1025)	(0.0756)	(0.0731)	(912.6772)	(0.2048)
Black	-1.4251	-0.0514	-0.0800	0.0090	0.0061	334.1880	-0.4693**
	(1.9620)	(0.0824)	(0.1159)	(0.0749)	(0.0769)	(955.4714)	(0.2186)
Hispanic	-1.3049	-0.0561	-0.1512	0.0158	0.0730	418.5266	-0.3657
	(1.8905)	(0.0914)	(0.1074)	(0.0810)	(0.0752)	(996.3690)	(0.2424)
Female	-0.1773	0.0293	-0.0923	0.0502	0.0560	289.7770	-0.1121
	(1.1544)	(0.0455)	(0.0597)	(0.0577)	(0.0516)	(415.3940)	(0.1691)
Age	0.0502	-0.0000	0.0027	0.0003	0.0004	10.6934	0.0051
	(0.0953)	(0.0032)	(0.0053)	(0.0042)	(0.0033)	(31.8118)	(0.0112)
Observations	411	411	411	411	411	411	411
Div-x-Cohort FE	X	X	X	X	X	X	X
Outcome Mean	49.45	0.129	0.628	0.170	0.165	15298	-1.51e-09
F-Test P-Value	0.970	0.257	0.0818	0.955	0.331	0.945	0.431

Notes: Standard errors are clustered at the initial assignment-by-cohort year level.

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

	(1) Age	(2) Female	(3) White	(4) Black	(5) Hispanic	(6) Hire Date	(7) Force Rate
White	-1.0720	-0.1124	-0.1555	-0.0052	0.0952	281.5651	-0.3147
	(1.7427)	(0.0845)	(0.0997)	(0.0688)	(0.0705)	(722.8539)	(0.1942)
Black	-1.4251	-0.0514	-0.0800	0.0090	0.0061	334.1880	-0.4693*
	(1.9739)	(0.0939)	(0.1163)	(0.0951)	(0.0801)	(809.8303)	(0.2394)
Hispanic	-1.3049	-0.0561	-0.1512	0.0158	0.0730	418.5266	-0.3657
	(1.8763)	(0.0874)	(0.1094)	(0.0921)	(0.0806)	(760.2191)	(0.2420)
Female	-0.1773	0.0293	-0.0923	0.0502	0.0560	289.7770	-0.1121
	(0.9826)	(0.0433)	(0.0639)	(0.0524)	(0.0538)	(328.7996)	(0.1070)
Age	0.0502	-0.0000	0.0027	0.0003	0.0004	10.6934	0.0051
	(0.0798)	(0.0033)	(0.0049)	(0.0038)	(0.0042)	(29.2205)	(0.0101)
Observations	411	411	411	411	411	411	411
Div-x-Cohort FE	X	X	X	X	X	X	X
Outcome Mean	49.45	0.129	0.628	0.170	0.165	15298	-1.51e-09
F-Test P-Value	0.961	0.515	0.338	0.928	0.389	0.877	0.324

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

Table A.4: Robustness Different Force Measures: 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	(10) Force	(11) Force	(12) Force
Theta Shrunk	0.113***	0.116***	0.0846**									
	(0.0375)	(0.0368)	(0.0397)									
Above Avg Theta Shrunk				0.000221	0.000262*	0.000179						
				(0.000137)	(0.000139)	(0.000141)						
IHS Theta Shrunk							0.113***	0.116***	0.0846**			
							(0.0375)	(0.0368)	(0.0397)			
Unshrunken Force Rate										0.0987***	0.0986***	0.0728**
										(0.0322)	(0.0315)	(0.0333)
Observations	1085020	1085020	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	0.00123	0.00123
Assigned Div by Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Recruit Characteristics	-	Y	Y	-	Y	Y	-	Y	Y	-	Y	Y
Call Controls	-	-	Y	-	-	Y	-	-	Y	-	-	Y

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

Table A.5: **Robustness Attrition**: The Effect of High Force Field Training Officers on Recruit Use of Force

	(1) Force	(2) Force	(3) Force
FTO Force Rate	0.000226***	0.000230***	0.000186**
	(0.0000700)	(0.0000695)	(0.0000732)
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: 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. The sample is only calls within two years of training.

Table A.6: **Robustness More Reportable Force**: The Effect of High Force Field Training Officers on Recruit Use of Force

	(1) Reportable Force	(2) Reportable Force	(3) Reportable Force
FTO Force Rate	0.0000136**	0.0000115*	0.0000106*
	(0.00000624)	(0.00000595)	(0.00000599)
Observations	1083076	1083076	1082876
Outcome Mean	0.0000507	0.0000507	0.0000508
Assigned Div by Cohort FE	Y	Y	Y
Recruit Characteristics	-	Y	Y
Call Controls	-	-	Y

Standard errors in parentheses

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

B.2 Appendix: 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.

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 35% (0.01306/.037)and 31% (0.0066/ 0.0215) increase in arrest or misdemeanor arrest rates.

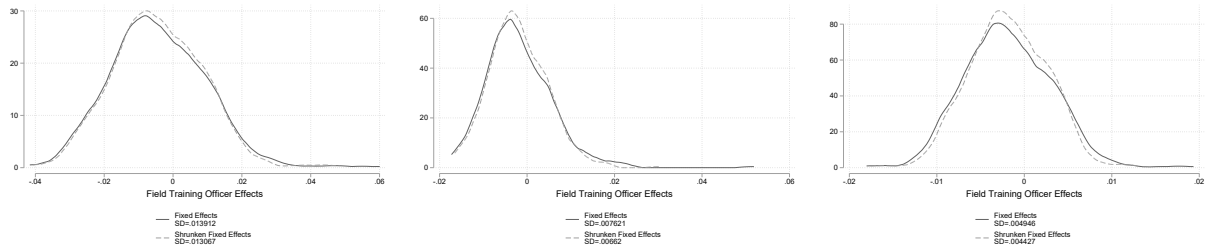
Figures B.1c and B.1d show results for our measures of time use (measured in hours). A one standard deviation increase in response time is 1 minute (0.02 hours) or 17% (.02/.12

hours) increase. A one standard deviation increase in time on a call is 0.359 hours or a 150% (.359/.24) increase.

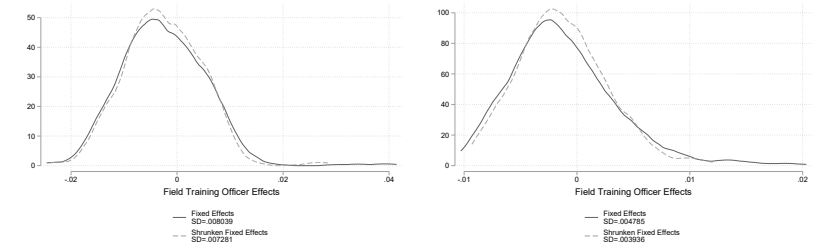
Finally, we consider how many words an officer uses when writing up an incident in Figures B.1e, B.1f, and B.1j . Unsurprisingly, the two distributions for number of words and characters look similar. The average number of characters used in an incident report is 43.19976 and there are a few officers that are very wordy. A one standard deviation increase in number of characters used is an increase of 8.39 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% (1.39/7.31) 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.

Figure B.1: Other Field Training Officer Rates

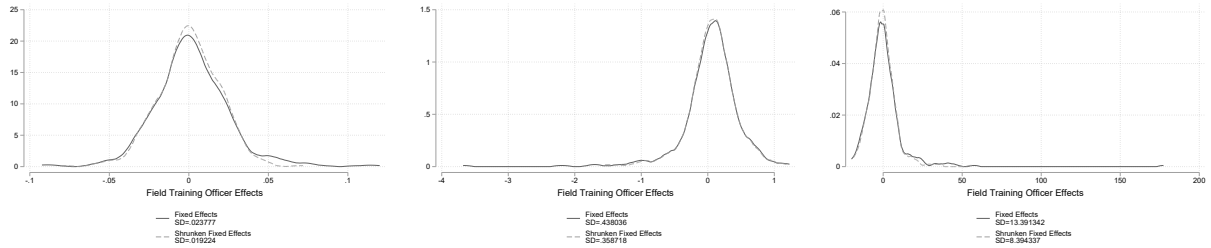
(a) Arrest (b) Misdemeanor Arrests (c) Felony Arrests



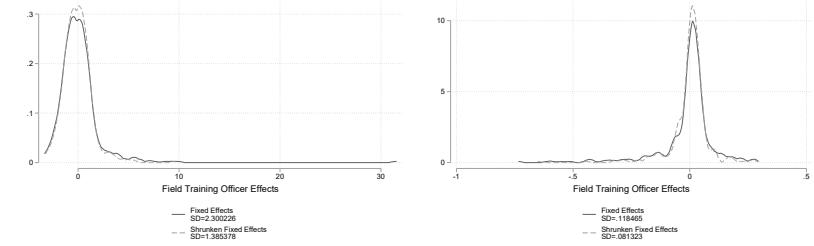
(d) Filed Arrests (e) Unfiled Arrests



(f) Response Time (g) Time on Call (h) Number of Characters



(i) Number of Words (j) Write Nothing



Notes: This figure plots the distribution of field training officer effects for arrests, misdemeanor 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