The Effect of Field Training Officers on Police Use of Force

Chandon Adger, Matthew Ross, and CarlyWill Sloan *

March 27, 2025

Abstract

The influence of on-the-job training and supervisors, especially in high-stakes settings like policing, is poorly understood. Examining a central behavior in the debate surrounding police reform, we investigate the impact of a field training officer (FTO) on a recruit's use of force. Leveraging a setting with conditional as-good-as-random assignment, we demonstrate a causal link between FTO and recruit use of force. A one standard deviation increase in FTO force propensity leads to a 14 to 18 percent rise in recruit force, persisting for at least two years. This underscores field training's impact and reveals a promising avenue for reform.

*Correspondence: Adger, U.S. Office of Personnel Management (OPM), chandon.adger@gmail.com; Ross, Northeastern University, ma.ross@northeastern.edu; Sloan, United States Military at West Point, carlywill.sloan@westpoint.edu Acknowledgments: We are grateful for valuable feedback from Bocar Ba, Shawn Bushway, Jesse Bruhn, Jamein Cunningham, Jennifer Doleac, Jeffrey Grogger, Elizabeth Luh (twice!), John MacDonald, Steve Mello, Aurélie Ouss, Anne Piehl, Steve Billings, Stephen Ross, Roman Rivera, and the Economics faculty at West Point. We also appreciate insights from participants at various conferences since 2021. Sloan and Ross acknowledge financial support from the Russell Sage Foundation (G-2011-30016) for our project, Does More Training Mitigate Disparities in Police Use of Force? Sloan also thanks the Society for Economics Representation and Networking for supporting writing retreats. We are especially grateful to Felipe Goncalves and Emily Weisburst for sharing DPD overtime and shift data. Finally, we extend our thanks to the Dallas Police Department for their valuable discussions, critical insights, and for welcoming us to visit their facilities

1 Introduction

Despite significant investment in workplace training across various sectors (Black et al., 2023), our understanding is surprisingly limited in terms of how these experiences shape worker behavior. This knowledge gap is particularly critical in law enforcement, where training directly influences vital aspects of society like public safety and the administration of justice. Amid an ongoing and contentious national debate about police reform, both policymakers and law enforcement often highlight training as a central strategy for reducing high rates of police violence and improving community relations. This emphasis aligns with public opinion, as nearly 90 percent of Americans favor increased training in areas such as de-escalation techniques and violence prevention (Ipsos, 2021). Beyond classroom instruction and simulations at the academy, most new police recruits undergo lengthy and formal on-the-job field training. This crucial phase of real-world training, pairing a recruit with a veteran officer, is widely recognized as a critical period where new officers develop the practical skills and judgment needed to guide their future decision-making.

Despite the widespread public support for police reform through training, there is limited empirical evidence on the impact of training on officer behavior and enforcement outcomes. While recent studies indicate some positive effects from classroom training (e.g., Dube et al., 2023; Mello et al., 2024; McLean et al., 2020; Owens et al., 2018), there is a dearth of knowledge about the critical field training period, generally six months of on-the-job training. Several factors suggest that field training may have a more profound and lasting influence compared to classroom instruction. First, the law enforcement community itself emphasizes field training as the most pivotal stage in an officer's development (Warners, Ronald, 2020). Second, recent research indicates that on-the-job experiences significantly shape subsequent enforcement behavior (Holz et al., 2023; Rivera, 2025; West, 2019; Weisburst, 2022). Finally, evidence from other professions in both the public and private sector illustrates that supervisors can have a substantial impact on worker behavior (e.g., Bennedsen et al., 2020; Hoffman and Tadelis, 2021; Lazear et al., 2015; Fenizia, 2022). This highlights the potential

of field training, a standard practice in most U.S. police agencies, to more effectively transmit improved policing practices and foster positive cultural norms across generations of officers.

In this paper, we investigate a critical question with implications for police reform: does a recruit's exposure to a more aggressive field training officer (FTO) influence their subsequent use of force? To address this question, we leverage administrative data on calls for service from the Dallas Police Department covering 2013 to 2019. Our choice of this institutional setting was motivated by initial conversations with the Dallas Police Department, where officials described a conditionally random process of assigning recruits to FTOs. Our analysis confirms that recruit and FTO characteristics are uncorrelated, a key identifying assumption for our empirical strategy. Similar to research on judge effects where researchers leverage random case assignment to identify the impact of judicial stringency (even when correlated with other attributes), this as-good-as random variation allows us to identify the causal effect of being assigned to a high-force FTO on a recruit's future propensity to use force.

We characterize aggressive FTOs by constructing a measure of each officer's propensity to use force prior to being assigned a particular recruit. In constructing our measure, we account for the possibility that some FTOs may be assigned to, or gravitate towards, more dangerous calls or locations. In particular, we control for a number of factors (e.g. neighborhood, call characteristics, and date/time) in order to isolate an officer's tendency to use force relative to their peers when placed in similar situations. Our measure of FTO propensity to use force is strikingly stable over time, with a correlation of 0.7968 when comparing force before and after leading a spell of field training with a given recruit. While our measure clearly captures systematic differences in how an FTO responds to an incident, it's important to note that these differences may reflect additional factors beyond simply a propensity or preference for using force. Importantly, our measure does not inherently indicate whether force was justified or necessary in a given instance. Moreover, we remain agnostic about whether more or less force is socially desirable and instead focus our analysis on whether field training is a channel through which the propensity to use force is transmitted from veteran to rookie

officers.

A one standard deviation increase in an FTO's prior propensity to use force (124% relative to the sample mean of 1.16 per 1,000 calls) increases a recruit's likelihood of using force by 14–18% after training. This effect persists for up to two years and is driven entirely by the first FTO assignment, consistent with recruits spending the most time and receiving formal evaluations from their initial FTO. We propose that this effect primarily operates through knowledge transfer about appropriate force use during field training. Exposure to a high-force FTO in this formative stage appears to lower a recruit's threshold for using force. Supporting this, recruits trained by high-force FTOs engage in more discretionary uses of force, take more marginally dangerous calls, and make more arrests for low-level charges unlikely to lead to prosecution.

Our findings have significant policy implications: recruits assigned to high-force FTOs are more likely to use force throughout their careers. Understanding why is crucial for designing effective interventions. While some alternative explanations—such as FTO demographics, proactive policing, post-training assignments, recruit exposure to force during training, and reporting behaviors—may play a minor role, our analyses indicate they are not the primary drivers of this effect. Instead, the strongest evidence points to the transmission of norms and expectations about force as the key mechanism. Even if we cannot pinpoint the exact mechanism with certainty, our findings provide clear causal evidence that supervisors shape officer behavior. This underscores the critical role of FTOs in establishing enforcement norms, whether through direct instruction, implicit modeling, or other influences.

Simulations based on our estimates suggest that replacing the most forceful FTOs with the least forceful could reduce recruit use-of-force incidents by 10.6%, lowering the total from 1,395 to approximately 1,247—a decrease of about 68 incidents per year. This is likely a conservative estimate, as it focuses only on recruit force and does not account for the dynamic effects of training lower-force officers who eventually become FTOs themselves. This finding emphasizes the importance of careful FTO selection during on-the-job training,

as these individuals play a crucial role in shaping the next generation of officers and their understanding of acceptable police behavior

Beyond contributing to a growing literature on the economics of public safety and policing, our paper also makes important contributions to broader the field of labor economics. First, our paper contributes to a literature on the impacts of on-the-job training, a widespread practice that remains surprisingly understudied. Police field training, where recruits are formally paired with an experienced officer for approximately six months, is similar to training models common in many other professions. In fact, surveys of incumbent workers suggest that 25-29 percent of the employed workforce has participated in six months or more of on-the-job training.[1] This type of training is particularly prevalent in fields with high-stakes outcomes and complex decision-making, such as emergency services (firefighters, EMTs/paramedics), healthcare (medical residents, nurses), and education (K-12 teachers). Research on apprenticeships, a more formalized type of on-the-job training, demonstrates their prevalence internationally and that there are positive impacts on labor market outcomes (e.g., Wolter and Ryan, 2011; Buechtemann et al., 1993; Cooke, 2003; Fersterer et al., 2008). However, studies on apprenticeships in the U.S. context have primarily focused on their earnings impacts (Hollenbeck, 2008; Reed et al., 2012) rather than job performance or the mechanisms through which skills and behaviors are transmitted. This knowledge gap is particularly significant in public service professions like law enforcement, where wages are relatively homogeneous and on-the-job training (which resembles an apprenticeship) is very common.

Second, we contribute to a growing literature on the effect of supervisors on workplace culture and organizational behavior. Although the impact of peers has been widely documented across various settings (e.g., schools, workplaces, neighborhoods, and the military),

[1]According to surveys of workers across occupational categories, there were only a scant 4 percent of occupations where the majority of workers reported having received no form of on-the-job training. These and the in-text statistic were generated through the authors tabulation of O*Net survey data on "On-the-Job Training" and "On-Site or In-Plant Training".

the impact of hierarchical relationships and mentorship is a newer research area that is understudied overall but particularly in high-stakes settings like policing.[2] The gap in the existing literature largely stems from the lack of available data on individual workplace behavior as well as the difficulty of obtaining plausible causal identification in this context (Bertrand and Schoar, 2003; Bloom and Van Reenen, 2007). Recent studies leveraging detailed personnel records and quasi-random assignments have documented supervisor effects in both private firms (Bennedsen et al., 2020; Hoffman and Tadelis, 2021; Bandiera et al., 2020; Bloom et al., 2013; Lazear et al., 2015; Baltrunaite et al., 2020; Minni, 2023) and public institutions (Tsai et al., 2015; Bloom et al., 2015; Rasul and Rogger, 2018; Fenizia, 2022). However, these studies typically focus on routine tasks and measures of productivity, leaving open the question of how supervisors influence complex, discretionary decisions or decisions with severe consequences. Our examination of police field training provides a unique set of insights into how supervisory relationships shape high-stakes decision-making in a setting where judgment and discretion are paramount.

Finally, we build on prior work examining peer effects and training in policing. In settings closer to our own, Holz et al. (2023) and Rivera (2025) study peer influences on police use of force and arrests, respectively. We differ from Holz et al. (2023) in both mechanism and persistence: while they study how peer injury affects the propensity to use force, we examine the transmission of force behavior through training relationships, finding longer-lasting impacts. Similarly, while Rivera (2025) documents persistent effects of minority peers on the decision to make an arrest, our study uniquely shows how training officers influence use of force decisions for up to two years after training ends. Our work also builds on research in criminology by Getty et al. (2014), who find a descriptive correlation between FTO fixed effects and recruit misconduct complaints using data from the same police department over

[2]See, for example, Anelli and Peri, 2019; Bifulco et al., 2011; Carrell et al., 2013; Jackson and Bruegmann, 2009; Sacerdote, 2001; Whitmore, 2005 on peer effects in schools; Mas and Moretti, 2009 on peer effects at work; Billings et al., 2019; Billings and Schnepel, 2020; Glaeser et al., 1996 on peer effects in neighborhoods; and Murphy, 2019 on peer effects in the military.

a different period.[3] Our more formal quasi-experimental design allows us to obtain plausibly causal estimates to address a different conceptual question: whether an FTO's propensity to use force influences their recruit's propensity to use force. Building on these prior studies, our analysis offers a more nuanced understanding of how training relationships shape an officer's subsequent enforcement behavior through what we believe is skill and norm transmission.

Related more specifically to the national conversation on policing reform, our results suggest that targeted interventions in training, such as careful selection of field training officers, could meaningfully influence long-term officer behavior and contribute to reducing use of force. This finding has significant policy implications, particularly given the widespread public support for police reform through training and the potential for these effects to amplify over time as new officers become trainers themselves. By demonstrating the impact of training supervisors on a recruit's subsequent high-stakes decision-making, our findings underscore the importance of investing in evidence-based training and mentorship programs to promote effective and responsible policing.

2 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). Like most agencies, however, training in Dallas is divided into two distinct stages before recruits become full-fledged police officers, i.e., Academy training and field training.[4] We believe that the findings from our study are broadly generalizable because the training received by recruits in Dallas is representative of

[3] While Getty et al. (2014) also use data from the Dallas Police Department, they apply a descriptive hierarchical modeling approach and do not distinguish between complaints received before, during, or after field training. Restructuring our data to match this study's approach yields qualitatively similar results. However, applying our quasi-experimental design to these data do not. Namely, we find no relationship between recruit complaints and FTO complaints or between FTO force and recruit complaints.

[4] After probationary officer, Police Officer is the lowest rank in the Dallas Police Department.

how the vast majority of policing agencies across the country train new officers. Dallas field training, in particular, is based on the "San Jose Model" which is the most commonly used program across the US (President's Task Force on 21st Century Policing, 2015). All of that said, recruits in Dallas tend to receive more training (both at the Academy and in the field) than the average US agency and the requirements also generally exceed those mandated by the State of Texas. Although our paper focuses exclusively on the field training component of a recruit's preparation for becoming a police officer in Dallas, we provide a brief but comprehensive discussion of the full training process.

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

After completing basic training at the Academy, recruits enter a second stage of training referred to as "field training". As noted above, this second stage of training is the focus of our study and has largely been overlooked by the existing empirical literature on policing. During field training, recruits ride with a more experienced officer (i.e., their FTO) in an

apprenticeship-style model where they are gradually afforded more autonomy. FTOs have a dual responsibility of providing service in their patrol area 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 a full-fledged police officer. 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 consisting of extensive on-the-job training was first developed by the San Jose Police Department in the 1960s and has since become a nearly universal standard for how law enforcement agencies in the United States approach the training of new recruits (President's Task Force on 21st Century Policing, 2015).

In Dallas, the field training process takes a total of six months to complete and consists of four phases. In the first and fourth phase, a recruit is paired with the same FTO. In the second and third phases, the FTO is different. The first three training phases of field training are each seven weeks long. The final evaluation phase lasts three weeks and is conducted by the initial FTO that a recruit was assigned during the first phase. When field training begins, recruits are instructed to take on a more observational role. As training progresses, they are given more autonomy and become active participants in responding to calls for service. For example, in the early weeks of field training, a recruit may simply watch their FTO respond to a call for service. However, the recruit quickly begins assuming a greater role in 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 used by the FTO to provide the recruit with extensive feedback on their performance but not read by command staff. When a recruit returns to their FTO for their final three-week evaluation, the FTO issues a final recommendation that determines whether a recruit is graduated through field training.

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

enforcement behavior. We decided to focus on the first FTO for three reasons. First, in our conversations with Dallas police officers, they communicated that the first field training shapes officers' policing "style" much more than their training at the Academy. Command staff in Dallas also emphasized that this phase of training is the most critical part of a recruit's development and that all peace officers remember the lessons learned during field training for the rest of their careers. Second, police officers and command staff in Dallas noted that the first phase of field training is the most significant because it is a recruit's first exposure to providing service. Third, recruits often return to their initial FTO for their final training and evaluation phase. Despite potential differences in FTO behavior across the three phases, we feel strongly that the first FTO is most influential because their assessment dictates a recruit's graduation from field training. [5]

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 where they do not have a permanent partner. Figure A.1 shows an example timeline for a recruit who began field training on December 7th, 2015. Beyond the fact that we believe FTOs represent a more influential relationship than these temporary partners, we also focus on the first FTO because they are conditionally as-good-as randomly assigned.

Key to the empirical design of our study, recruits in Dallas have no discretion in choosing their initial FTO and command staff do not strategically make these pairings. First, recruits are assigned to one of seven divisions for field training based on the staffing needs of the department. While we were told that district assignments are occasionally made to fill

[5]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.9). Our parameter estimate on the first FTO's force rate remains stable to the inclusion of these additional measures. We interpret this as evidence that is consistent with our conversations with Dallas officers about the first FTO being the most influential.

minimum staffing requirements for female and Spanish-speaking officers, they are not based on skills, preferences, or performance at the Academy. Within each division, division command staff then assign recruits to FTOs along with their associated shift and patrol area. According to command staff at Dallas PD, these assignments are made without knowledge of or regard to recruit characteristics or Academy performance. Importantly, we provide empirical evidence supporting the claim that, conditional on division, the initial assignment of recruits to Phase 1 FTOs is as-good-as random. At later stages of field training when command staff gain familiarity with recruits, we are less confident that the assignments are as credibly random. Our concern about potential selection at later stages of field training stems from the fact that an FTO assignment is implicitly tied to a shift assignment.[6]

In our study, we document FTO and recruit behavior using 911 calls for service. When a civilian calls 911 in Dallas, they are first connected to a 911 operator. The operator will then record essential characteristics of the call such as location, description of events, and time in the Computer Aided Dispatch System (CAD). The operator will also place the call into a standardized category, such as "domestic disturbance." Finally, the operator also records their perception of the urgency and severity of a call. This is referred to as the priority of the call, with each assigned a value from 1 to 4, with 1 being the highest priority. The information recorded in the CAD system is then provided to police dispatchers whose job it is to assign calls to police officers. Dispatchers assign calls to officers based on priority level (relative to other calls in the queue), proximity, and availability. Unlike in other jurisdictions, officers in Dallas do not have assigned beats. It is very common to observe an officer in many different beats during their shift. If there are more active calls than available officers, lower-priority calls are postponed until higher-priority calls are resolved. Dispatchers also decide

[6]To formally assess this concern, we conducted balance tests separately for Phases 2 and 3. While balance holds in Phase 1 and in the combined sample of all three phases, it also only holds in Phase 2 but marginally fails in Phase 3, with notable imbalances in FTO age and experience. Given that shift assignments for non-probationary officers (i.e., FTOs) are based on a bidding system that considers rank and experience, and that FTOs are relatively homogeneous in rank, we believe these balancing failures are consistent with selection into more favorable shifts as recruits progress through training.

the number of officers to assign to a call based on Dallas PD's standards. For example, more serious incidents (such as shootings and mental health calls) may require a dispatcher to assign multiple officers. It is also possible for officers to observe an incident while on patrol or "dispatch" their car to a particular call, i.e., self-select into calls. Unlike some jurisdictions where officers cannot self-select calls, Dallas allows officers some discretion about which calls to answer. After responding to a call, officers have significant discretion in how they handle the incident, including the decision to make an arrest or use of force.

To measure officers' use of force, we link 911 calls to force reports.[7] 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 requires all force incidents to be documented as a "Response to Resistance" entry in a proprietary database called BlueTeams licensed from a software company called IAPro.[8] All force incidents are reviewed by a supervisor (Dallas Police Department, 2021) and the internal affairs division has access to the entire database. According to the Dallas Police "The physical control techniques used may range from the use of handcuffs in an arrest, strikes with an impact weapon, or the use of a firearm" (Dallas Police Department, 2021). According to police officers and command staff at Dallas PD, the penalty for not correctly reporting a force incident is extremely severe and compliance is virtually universal. In written correspondence with the authors, Stephen Bishopp Major of Police at DPD Research Division, stated to us that "Compliance with use-of-force policies is strictly enforced and widely monitored, thus under-reporting of force events is rare and not been historically

[7]Dallas refers to force as a "response to resistance".

[8]Entry of a Response to Resistance event is required for any force that is "Soft Empty Hand Control" or above on the Response Continuum, with the exception of "Compliant Handcuffing". These include, but are not limited to, the following types of force: 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.

problematic."[9]

To further refine our analysis and illuminate the specific mechanism under study, we examine three categories of force that typically receive heightened scrutiny from internal investigations units and federal consent decree monitors during use-of-force policy audits (U.S. Department of Justice, 2021). These categories focus on incidents where the justification for force may be less clear-cut, allowing us to better isolate the influence of FTOs on recruit behavior. Specifically, we consider: (1) force incidents without a subsequent arrest, as arrests typically follows when force is used for self-defense, defense of others, or response to resistance; (2) force incidents where the narrative describes relatively innocuous civilian actions (e.g., "furtive gesture," "not following commands," "null," or "off safety-no resistance"); and (3) force incidents resulting in civilian injury, which may indicate a lack of proportionality. While these metrics cannot perfectly identify unnecessary force, they collectively provide a consistent framework for identifying potentially excessive or discretionary force applications where officer proportionality and de-escalation were lacking.

In addition to examining 911 calls that result in force, we also examine the likelihood of a call ending in an arrest. Arrest reports, linked to specific calls for service, include the criminal charge(s) as well as a flag categorizing these charges as a felony, misdemeanor, or n-class (i.e., warrant). To further explore officer discretion, we develop two additional arrest measures. First, we disaggregate arrests into"filed" (accepted for prosecution) and "unfiled" (dismissed) categories which we only observe for misdemeanor or felony arrests. Unfiled arrests potentially reflect lower-quality police work, as dismissal typically occurs when the

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

evidentiary threshold for prosecution is not met. Second, we distinguish between felony arrests (e.g., robbery, domestic violence, aggravated assault, homicide) and misdemeanor arrests (e.g., disorderly conduct, narcotics possession, petty theft, simple assault). Felony arrests typically necessitate an arrest, while misdemeanor arrests often involve greater officer discretion due to the ambiguous nature of the offenses, allowing officers more latitude in deciding whether to make an arrest based on their interpretation of the situation, the severity of the offense, and the suspect's demeanor. We also construct a measure of officer-specific complaints, including both internal and external complaints filed against individual Dallas police officers. While these complaints cannot be linked to specific calls for service, they provide a broader view of officer behavior and potential misconduct.

Lastly, the Dallas Police Department did not have a dedicated traffic division or conduct many vehicle or pedestrian stops during our sample period. We also lack sufficient information to meaningfully analyze behavior related to other types of officer-initiated events, Terry stops, or searches. In fact, using the definition in Weisburst (2022), we observe that less than 1% of calls are officer-initiated. Although discretionary traffic and Terry stops are important in many other settings, these measures are not feasible in our context but could represent another way in which an FTO influences a recruit's future enforcement decisions. Importantly, because officer-initiated events constitute a very small share of incidents in our setting, selection into different types of calls for service is another channel through which FTOs could hypothetically shape recruits behavior. We will refer to this particular channel at later points as an example of "proactive policing" — i.e., where officers choose to respond to a higher volume or more dangerous calls.

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.[10] [11] 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 that a recruit arrives to a call with a senior police officer.[12] 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

[10] In linking the force and arrest records with calls for service, we do so based on the incident identifier but not the officer badge number. We have taken a conceptual stance that it is more correct to associate an incident resulting in force with every officer on the scene. This is because one officer may influence another officer's behavior even if they're not explicitly listed as applying force. We also restrict force incidents to those we are confident (based on the timestamp) occurred at the scene of an incident as opposed to those occurring after a suspect is in custody and being transported to jail. The likelihood of a call for service to result in force in our sample is comparable to that reported in Weisburst, 2022, Hoekstra and Sloan, 2022, and calculated by the authors in other publicly available data. For use of force, if we also match on officer name we lose about 10 percent of our force observations. However, our results are qualitatively similar using this alternative data restriction (coefficients of 0.000208*** & 0.000205*** & 0.000161*** for the specification shown in Table 3).

[11] 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 with a 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.

[12] 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.

inverse number of senior officers on a given call. The intuition of this weighting scheme is that the calls where a recruit arrives with only one other officer (weight = 1) provide more information about the identity of their FTO relative to calls where there are many senior officers (weight = $\frac{1}{n}$) on the scene. In our sample, we have a total of 521 recruits and we identify a total of 251 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.[13] 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 515 of the officer-recruit pairs that we have identified as FTO-recruit pairings also appear in the overtime pay dataset while 6 (approximately 1 percent) are not. Our results are robust to dropping these pairings where we fail to find the FTO in the overtime data.[14] 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

[13]Our primary method for identifying each recruit's FTO in each field training phase is to use the calls for service. Giving more weight to calls with with fewer officers, we determine the senior officer with whom the recruit most frequently appears on the scene of an incident. However, we can also link our inferred FTO-recruit assignments from the calls data to overtime pay records. These records identify when an FTO receives overtime pay or comp time due to the requirement that they complete a required a daily recruit evaluation at the end of each shift. When we do this, we find that our approach correctly identifies 95.22 percent of actual FTOs. Our results are also robust to creating the recruit-FTO pairs in a more restrictive fashion that relies primarily on these administrative overtime pay records. While both methods yield qualitatively similar results, we feel that our preferred method better represents the treatment that we actually care about, i.e. exposure to a senior police officer, even if that particular officer isn't always necessarily their "official" FTO.

[14]We estimate coefficients that are quantitatively similar in magnitude and precision.

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 persist for three years should expect a promotion.[15] Although command staff emphasized that there is still some selection in terms of who is allowed to become an FTO, it was not necessarily a position reserved for only highly experienced or exceptionally talented officers. According to our data, the average age of an FTO is 42. FTOs were also generally representative of the whole police force in terms of demographics, but less diverse. Specifically, 15 percent of FTOs were Hispanic, 18 percent were Black, and 67 percent were White, compared to 28 percent, 22 percent, and 45 percent in the recruit sample, respectively.

3.2 Force Rate Calculation

Next, we assign each of the 521 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 before being assigned a particular 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.*[16] In estimating this fixed effect, we also control for important call characteristics such as the

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

[16]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. We will subsequently address the issue of varying levels of sample size and precision using Bayesian shrinkage.

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.[17] 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. While our preferred specification includes call characteristics and implicitly assumes no selection into different call types, this assumption is conservative. Since officers in Dallas (unlike other agencies) have some discretion about which calls they respond, part of the underlying mechanism could also be driven by selection into different types of calls. Therefore, we also calculate an alternative force rate without these controls. While selection into more dangerous calls appears to play a minor role, it does not explain most of the observed effect that we will eventually estimate (see Table A.1).

We assess the stability of our FTO force over the sample period by first examining whether an FTO's use of force prior to training a particular recruit predicts their use of force after the recruit has completed training. Comparing the likelihood of an FTO to use force pre and post-training, we find a raw correlation of 0.80. Regressing post-training on pre-training force, we get a coefficient estimate of 1.04 with a standard error (clustered on FTO) of 0.05. We believe that the high correlation and near-unit coefficient estimate demonstrate that use of force represents a stable behavioral tendency and not a factor that is primarily driven by random variation or idiosyncratic situational factors.

Formally, we estimate

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

where $force_{o(r),c}$ is a binary variable equal to one if call c answered by FTO o ends in force

[17] There are 48 priority-by-type fixed effects. The type of call is categorized by the call taker. An example type of call is burglary.

and zero otherwise.[18] 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 an FTO has historically been more aggressive (i.e., uses force more frequently) while lower values of indicate an FTO has been less aggressive (i.e., uses less force). Since a given training officer can appear paired with several recruits over the sample period, we cluster our standard errors by FTO.

As discussed, our analysis focuses on recruit cohorts joining the Dallas Police Department between July 2014 and July 2018, and our data spans January 2013 to July 2019. Thus, the FTO force measure varies in the number of underlying pre-training calls for service depending on timing of a particular FTO-recruit pairing relative to the sample window. For instance, 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 July 2018 will rely on (at most) 6 years of pre-period data. In addition to the issue of left truncation, force is also a relatively rare outcome 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 fewer pre-period calls (due to left-truncation), observe answering fewer calls for service (for any other reason), or who just have a larger within-officer variance in their propensity to use

[18] 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.

force. The intuition of applying Empirical Bayes is that the resulting FTO force measure primarily reflects variation across the estimates in which we have the highest confidence.[19]

Formally, we estimate the across officer variance in FTO force, σ^2_A, and a within-officer variance, σ^2_W [20] Next, we use our two variance measures and the number of observations per officer to estimate a shrinkage factor $\frac{\sigma^2_A}{\sigma^2_A + \frac{\sigma^2_W}{N_{o(r)}}}$. Finally, we construct our final shrunken force rates as

$$\Lambda_{o(r)} = \frac{\sigma^2_A}{\sigma^2_A + \frac{\sigma^2_W}{N_{o(r)}}} * \lambda_{o(r)} \tag{2}$$

where we multiply our shrinkage factor by our original fixed effects. We plot the distribution of police officer force rates for all 521 FTO-recruit pairs in Figure 1a 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 calculated by subtracting the mean and dividing by the standard deviation of our force measure) for ease of interpretation.[21]

The distribution of standardized effects is shown in Figure 1b. One standard deviation increase in FTO effects is a 0.144 percentage point (124 percent compared to the sample mean of 0.116 in Table 1), increasing from 1.50 to 3.37 use of force incidents per year[1]. Moving from the FTO that used the least to the most amount of force represents an approximate 6.5 standard deviation increase, resulting in about 15 additional use of force incidents per

[19] We can also condition the data such that we construct FTO force using the same volume of calls (i.e. 2,000) before a recruit is assigned. Our main results (Table 3) are similar in magnitude to those obtained when we impose this additional restriction and re-estimate alternative FTO force rates. The limitation of this method is that we must restrict our sample of FTOs in a relatively non-random and adhoc manner.

[20] We calculate the within officer residual variance as $\sigma^2_W = E(\epsilon^2_{o(r)})$.

[21] We note that our main results are robust to using the unshrunken estimates as well as a number of alternative specifications (see Table A.1).

year per officer (1,000 percent increase). Replacing an FTO at the 10th percentile for one at the 90th percentile corresponds to just over a 2.5 standard deviation change, leading to approximately 7 additional use of force incidents per year per officer (467 percent increase).[22]

Finally, we compare FTO force rates to the force rates of other patrol officers in Dallas. To do so, we first construct a force rate for each officer using our entire sample of calls for service. Next, we shrink and standardize the force rates as described above. Our results are shown in Figure A.2.[23] On average, FTOs use force more, about 0.08 standard deviations on average, than the typical non-FTO officer, and more than the average senior Corporal or Sergeant (the ranks most likely to be FTOs). Despite these differences, our main takeaway is that there is significant overlap between the two distributions.

3.3 Summary Statistics

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 1. In our sample, roughly 4 in 100 calls end in any arrest, 2 in 100 end in a misdemeanor arrest and only 1 in 1,000 calls end in a use of force. We characterize a call as having involved force or arrest regardless of the specific officer who used force or made the arrest. This conceptual decision was motivated by our concern about possible endogeneity in terms of the specific officers on the scene of an incident and who actually ends up using force.[24] 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 (beat), date, and time. Dallas uses four priority categories, and we observe 234

[22] Base rate $R_0 = 0.116\%$, one standard deviation increase $SD = 0.144$ percentage points. Average calls per officer per year: $1,296 \times (R_0 + SD) = 3.37$ incidents, compared to baseline $1,296 \times R_0 = 1.50$ incidents.

[23] There are a few (3 percent) very extreme force users in our sample that we drop to create a figure that is easier to visually inspect.

[24] 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 to occur after a suspect is arrested and being transported to jail.

beats and 20 types of calls. As a descriptive exercise that serves as a preview of our main results, we also split the sample into calls where the recruit had a high-force FTO (force measure greater than average FTO) and recruits who have a low-force FTO (force measure less than average FTO) in Table 1. Consistent with our main result, we find that recruits assigned to a high-force FTO are more likely to use force during a call for service after they complete training.

We also present summary statistics at the recruit level in Table A.2. As noted, there are 521 recruits in our sample. The average recruit is younger than the average FTO (26 vs. 40).[25] Most recruits are White (45 percent), 22 percent are Black, and 28 percent are Hispanic. Thirty-eight percent have a bachelor's degree or higher, 8 percent had prior law enforcement experience (jailer, 911 operator, etc.), and 4 percent had training in de-escalation or crisis intervention prior to field training. Given the conditional random assignment of recruits to their first FTO, we should expect that recruit characteristics do not differ across the type of FTO. While our formal balance tests control for cohort-by-division fixed effects, raw summary statistics suggest similar demographics across high- and low-force FTOs, though educational attainment and prior law enforcement experience differ.[26]

4 Empirical Methods

4.1 Estimation Model

The conditional random assignment of recruits to Phase 1 FTOs provides an ideal setting for investigating how field training shapes a recruit's subsequent policing behavior. We formally

[25] We use age at training, though this variable may contain some measurement error. Notably, a few recruits exceed Dallas's stated age range for recruits. To account for potential inaccuracies, we include a binary variable that flags these observations as likely erroneous in all specifications that use age.

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

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 before their pairing. As discussed, we shrink this measure using Empirical Bayes and standardize it (subtract mean and divide by standard deviation) for ease of interpretation. Thus, our coefficient of interest β_1 can be interpreted as the difference in a recruit's likelihood of using force on a given call caused by a one standard deviation increase in their FTO's prior propensity to use force. We control for possible variation across recruits in their initial assignment over time by including θ_r representing a set of 152 Academy cohort by division fixed effects.[27] To control for variation across calls, we also include X_c representing a vector of call and recruit attributes. In our fully saturated model, this vector includes FTO characteristics and recruit characteristics (age gender, race, education, prior law experience, prior training), geographic fixed effects (234 beats), call characteristics (priority, call type), number of officers dispatched, as well as year-by-month and day of the week-by night fixed effects. We implement two-way clustering of standard errors on FTO and recruit.[28] By applying Bayesian shrinkage to our primary explanatory variable (FTO force), we account for estimation error in this derived variable, similar to inverse variance weighting, which we find produces similar results.

The model's identifying assumption is that FTO characteristics, primarily prior propen-

[27] Most recruits are assigned to the 6 largest divisions, so it is probably more reasonable to think of this as 132 groups. Our main results (e.g., Table 3) are very similar if we control for division and cohort separately. We choose division-by-cohort fixed effects because they reflect the natural experiment we leverage: FTOs are assigned to recruits within each police class and division.

[28] We cluster on FTO because they can be paired with multiple recruits and we use these pairings to develop our primary independent variable. We cluster on recruit because the unit of observation is a call for service answered by a recruit in the period following training. We are also robust (i.e. our estimates are statistically significant at the 5 percent level or less) to one-way clustering by FTO.

sity to use force, are not correlated with recruit characteristics after conditioning on division by cohort. Therefore, identification relies on the conditional random assignment of recruits to FTOs within a given division by cohort. Without random assignment of recruits to FTOs, our estimates of the impact of an FTO's prior force propensity could be biased, as this measure might be correlated with other recruit characteristics that influence call outcomes. 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. In our subsequent estimates of the effect of FTO prior propensity to use force on a recruit's subsequent use of force, we will also demonstrate parameter stability to including FTO and recruit characteristics as control variables.

4.2 Research Design

While we have previously discussed that our conversations with Dallas command staff suggest that Phase 1 FTO assignments are as-good-as random, we will now provide empirical evidence to support this claim. We start by demonstrating that the characteristics of FTOs are not correlated with those of their assigned recruits, a hypothetical source of bias in our main estimates. We regress FTO characteristics against recruit characteristics, treating each recruit-FTO pairing as an observation.[29] Given that FTOs are randomly assigned only within divisions and during each cohort, we include fixed effects consistent with this assignment mechanism. We analyze correlations between FTO age, race, and use of force rate with recruit age, race, gender, and hire date (in years since 1960).[30]

The results of testing for balance are contained in the first six columns of Table 2 where we find that none of the 60 coefficient estimates are statistically significant above conventional

[29] A recruit-FTO pair means one observation per recruit, though each FTO may be assigned to multiple recruits throughout the sample period.

[30] Hire date is denoted in years since 1960.

levels (i.e. exceeding a 95 percent confidence level) and only three coefficient estimates are marginally significant (i.e. exceeding a 90 percent confidence).[31] Moreover, joint F-tests also yield no significant p-values as does our application of an omnibus test of joint significance across all pairwise combinations of dependent and independent variables. [32] We also conduct t-tests for each combination of FTO and recruit characteristics within each academy cohort—yielding 717 distinct t-tests.After correcting for multiple hypothesis testing using Simes (1986) method, only 12 tests are significant at the 5 percent level or below.[33] Figure A.3 also illustrates the distribution of FTO force rates by recruit characteristics. Given our conditional random assignment, we would expect these distributions to be similar, which is what we observe. A Kolmogorov-Smirnov test also fails to find statistically significant differences between these distributions.

Additionally, we demonstrate that recruit characteristics themselves predict recruit use of force, as shown in the last two columns of Table 2. We calculate recruit use of force both with and without accounting for call controls as outcomes. The results show that certain recruit characteristics—specifically prior law enforcement experience, de-escalation training, and age (40s)—are statistically significant predictors of recruit force use, particularly when call controls are not included (column "Recruit Force NC"). This finding is important because even if these recruit characteristics do not predict FTO assignment, they are meaningful characteristics that would reveal selection bias if it were present. The fact that these characteristics predict recruit behavior but not FTO assignment further strengthens our claim of random FTO assignment within divisions and cohorts.

[31] Here, we present standard errors clustered at the FTO level but note that we find quantitatively similar levels of precision using robust standard errors. See Table A.3.

[32] To test for systematic differences in recruit characteristics across FTOs in a single well-powered test, we conduct a pairwise omnibus test. We stack the data into 60 FTO-recruit characteristic pairs (6 FTO variables $\times$ 10 recruit characteristics) and regress the dependent variable on the independent variable interacted with one of 60 pair-specific fixed effects. We include cohort year by division by pair fixed effects and cluster standard errors on FTO by pair. A joint significance test on all 60 coefficient estimates yields a p-value of 0.7953, indicating no statistically significant differences.

[33] The expected number of t-tests exceeds 717, but a lack of variation in rarer recruit or FTO characteristics in some cohorts reduces the total.

5 Empirical Analysis

5.1 Evidence from the Raw Data

To begin our analysis, we provide a set of figures to illustrate the underlying data patterns. While our formal analysis accounts for cohort-by-division (i.e., the institutional structure where random assignment happens), we believe examining the raw data sheds light on the relationship between an FTO's tendency to use force and their recruit's subsequent behavior. In the top panel of Figure 2a, we plot the local average recruit use of force across different FTO force rates. Observations are binned to ensure each point represents an equal number of calls, and a linear regression line is fitted across all force rates. This figure offers two key insights. First, we see both high and low-force FTOs being dispatched to calls. Second, there's a clear positive relationship between an FTO's force rate and subsequent use of force by their recruits. The slope of the linear fit indicates that a one-standard-deviation increase in the FTO's propensity to use force corresponds to a 0.025 percentage point rise in recruit use of force.

In the bottom panel of Figure 2b, we investigate whether recruits assigned to high-force FTOs are more likely to take on calls that are inherently more dangerous and more likely to end in force. If true, this would suggest that the observed relationship between FTO force and recruit use of force might result from selective assignment to more hazardous calls, rather than a general propensity for force. Alternatively, if there is a strong positive relationship, it could suggest that high-force FTOs lead to more predictable and perhaps more appropriate uses of force, rather than possibly erratic or excessive force.

To assess this possibility, we first regress recruit use of force on cohort and initial assignment fixed effects, then examine the residuals against various covariates related to each call. These covariates include fixed effects for the number of officers present, the beat, the type of call (prioritized by type), year-by-month, and day-by-night. We use this model to predict the likelihood of force for each call, adding the average use of force rate to obtain our

estimate.[34] Although there is a slight positive correlation between predicted force and FTO force rates, it's much weaker than the top panel's relationship. Given these results, and to account for any confounding variables, we estimate our main results with call controls. This approach allows us to minimize the influence of selection into calls and focus on the causal relationship between FTOs and recruit behavior.

Overall, these figures suggest that FTOs play a significant role in shaping recruits' use of force. However, there's potential for confounding due to recruits potentially sorting across cohorts and divisions. Thus, we proceed to our main analysis which exploit conditional random assignment by incorporating cohort 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 3. Each specification includes cohort 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.[35] 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. Our results show that recruits with FTOs that use force one standard deviation more are 0.0206 percentage points or 18 percent more likely to use force.[36]

[34] This creates a linear combination of call characteristics, with weights chosen to predict the probability of force being used.

[35] Namely, all columns are statistically significant at the 5 percent level or less.

[36] In Table A.1, we present results using alternative measures of force rates. Specifically, our findings remain robust when using the unstandardized measure, the inverse hyperbolic sine transformation of our shrunken force measure, and the unshrunk force measure. The results are similar and statistically significant at conventional levels. We also report estimates using our force measure calculated without call controls. As expected, the coefficients are slightly larger, but the key findings remain unchanged.

In column 2, we add controls for recruit characteristics (age, gender, race, education, prior law and training experience). Given our conditional random assignment and the results in Table 2 and Figure A.3, we would not expect recruit characteristics to alter our estimate meaningfully. Column 2 indicates this is the case because our estimates only decline by about 1 percent.

In column 3, we add controls for call characteristics which are the same used to predict FTO force. Even if recruits are indeed randomly assigned to FTOs, the inclusion of call controls changes the interpretation of our treatment effect. This is because assignment to a high-force FTO could cause recruits to work in areas where calls tend to be more severe or to select into more dangerous calls for service. Another way to interpret column 3 is to think of the inclusion of call controls as accounting for systematic differences in the types of calls recruits respond to, which could include self-selection or engaging in more proactive forms of policing. Given the small positive correlation in Figure 2, we should expect the magnitude of our estimates to attenuate slightly. Indeed, our estimate in column 3 is about 23 percent smaller than the prior column but remains economically meaningful and highly statistically significant. This decline suggests that proactive policing may play a role in explaining some of the observed effect— about a fifth of the estimate— but does not fully account for the observed relationship. In a later section, we will introduce further FTO attributes that we believe capture different aspects of proactive enforcement into our baseline model. These attributes include things like response times, time on calls, complaint rates, and arrest rates. Importantly, even after controlling for these factors and holding constant these alternative channels fixed, we find a one standard deviation increase (124 percent) in FTO force increases recruit force by 0.01657 percentage points or 14 percent- see Table 6. While we discuss these results in more detail later, we note here that they are further evidence in support of our preferred mechanism: that FTOs transfer information to recruits about the appropriate use of force rather than merely increasing their likelihood of engaging in more proactive forms of policing.

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 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 by their FTO's prior propensity to use force. Notably, coefficients for officers with prior training or law enforcement background are meaningfully larger, though we feel that some caution is warranted given the relatively small sample sizes of these groups.

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 Hispanic 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.

5.3 Persistent Effects and Potential Attrition

Understanding the importance of FTOs in terms of force behavior requires considering the duration of our treatment effect. This is particularly vital given that Holz et al. (2023) documents only short-term effects for the same outcome but a different treatment, i.e., the effect of peer injury on the use of force.

To examine the evolution of our main effects, we estimate a model of the form:

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

where *biannual* is a binary variable that equals 1 in t 6-month periods after the end of training. We also introduce separate fixed effects for *biannual*. All other terms remain unchanged from Equation 1, and column 3 controls are included (i.e., call and recruit characteristics). Results are shown in Figure 4. The coefficient is positive for each of the first four time periods. However, after two years, the effects appear to attenuate. We note that our sample becomes sparse as we examine further time horizons beyond two years, likely relying on less variation in FTO force propensity. Nonetheless, we feel that it is worth noting that most officers are promoted to senior corporal after three years of service and may themselves then 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. To address these concerns, firstly, we note that in our sample, 35 (7 percent) recruits leave before one year of service and 67 (13 percent) before two years. While there is some attrition, we do not feel that it substantively impacts our estimates. In particular, we explore the relationship between recruit attrition and FTO and recruit characteristics in Table A.6. Here we regress characteristics of FTOs or recruits on a binary variable for whether a recruit leaves the sample. The table shows recruit and FTO characteristics are not good predictors of leaving. In fact, only one characteristic (prior law enforcement experience) is statistically significant (at the 10 percent level)[37]

We alternatively address this potential concern by limiting our estimation sample to the calls for service data occurring in the first year or two years after training, when most of the sample is still working for DPD. Using these alternative samples in Table A.7, we find very similar results to Table 3. Finally, we investigate whether our main treatment effects

[37]Recall that we measure the hire date in terms of days since 1960.

differ across recruits who stay and leave. To do so, we estimate heterogeneous treatment effects for recruits that do and do not leave in Table A.8. The coefficient on the interaction term indicates there is no statistical difference in the effect of high-force FTOs for recruits that stay or leave. Given these results, we believe that we can confidently set 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 3. In our main estimates, we follow a conventional approach by two-way clustering our standard errors at the recruit and FTO level (Bertrand et al., 2004). However, the concern motivating this robustness test is that our outcome variable (force by call) is a very rare event in our sample, occurring in only about 1 percent of calls. In cases where an outcome variable is a rare event, standard asymptotic assumptions related to the distribution of point estimates and associated standard errors may be inappropriate. Here, we use randomization inference to construct an empirical distribution of point estimates and reassess the validity of the hypothesis tests conducted for our primary set of estimates.

As discussed in Efron (2004), randomization inference is most appropriate to non-experimental settings when researchers are able to replicate the data generating process of the observed data. In our institutional setting, recruits from a given academy cohort are randomly assigned to FTOs by command staff within their respective division. As discussed previously, our balancing tests support that this source of variation is as-good-as random. Thus, our randomization procedure attempts to replicate this variation in constructing an empirical distribution of point estimates and associated standard errors. For each recruit in our sample, we randomly draw an FTO from the set of eligible officers we observe working in the

recruit's respective division. [38] 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 3. In order to obtain p-values for a two-sided hypothesis test, we replicated this procedure 1000 times and calculate the share of the simulations when the t-statistic exceeds the absolute value of the t-statistics from Table 3.

For the models corresponding to columns 1 and 2 of table 3, we obtain p-values for a two-sided hypothesis test using randomization inference of 0.013 and 0.008 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 3. In our randomization procedure, we find that only 37 of the simulations resulted in a t-statistic more extreme than our baseline estimate, corresponding to a two-sided test p-value of 0.037. Of these, 23 simulations had t-statistics strictly greater than our baseline estimate, corresponding to a one-sided p-value of 0.023. The dashed line in Figure 5 marks our original t-statistic. We interpret these results as providing additional evidence indicating that the inference in our main results is not entirely driven by potential issues associated with incorrect asymptotic assumptions or rare-event bias.

5.5 Analysis of Discretionary Force

To further investigate our preferred story about the main effect being driven by the transmission of norms for applying force, we examine three subsets of force incidents where it

[38] 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.

is more likely that the justification is marginal and its application is discretionary. Although we lack data to definitively assess whether specific force incidents were determined to be within policy, even if we had that data, we would suspect that the determination is endogenous. Instead, we consider: (1) force incidents without a subsequent arrest, which may indicate unnecessary escalation; (2) force incidents resulting in civilian injury, suggesting disproportionate force; and (3) less justifiable force incidents, where civilian resistance was minimal (e.g., "furtive gesture," "not following commands," "null," or "off safety-no resistance"). These categories, which often receive heightened scrutiny from internal investigations and federal consent decree monitors during use-of-force audits (U.S. Department of Justice, 2021), help identify incidents where force may have been unwarranted. Results are shown in Table 4.[39]

Across all three outcomes, the assignment of a high-force FTO corresponds to an increase in each category of force, with the strongest effect observed for force incidents without a subsequent arrest. A one standard deviation increase in FTO force propensity translates to a 12-16 percent increase in force without an accompanying arrest, comparable in magnitude to our main estimates. While the results for less justifiable force and force with civilian injury are not always statistically significant at conventional levels, they suggest that high-force FTOs may lead to increases in more discretionary applications of force. These findings are consistent with, but do not definitively establish, the explanation that high-force FTOs increase recruits' propensity to use force in less warranted circumstances.

5.6 Analysis of Decision to Arrest

Having shown that recruits assigned to high-force FTOs are more likely to use force later in their careers, including in more discretionary situations, we now examine whether this effect extends to another important outcome: the decision to make a discretionary arrest.

[39]It is also possible that force incidents resulting in civilian injury are more likely to be reported, even if the injury is documented by police officers. One could interpret these results as consistent with the argument that reporting does not drive our findings.

Importantly, an FTO's propensity to use force has no impact on a recruit's overall likelihood of making arrests (see Table A.5). Instead, we focus on an officer's discretion in making different types of arrests by developing four additional measures.

First, we distinguish between arrests that were "filed" (accepted for prosecution) and those that were "unfiled" (not accepted for prosecution). Notably, we observe these classifications only for misdemeanor and felony arrests. Unfiled arrests may indicate lower-quality police work, as dismissals typically occur when the evidentiary threshold for prosecution is not met. Second, we distinguish between felony arrests (e.g., robbery, domestic violence, aggravated assault, homicide) and misdemeanor arrests (e.g., disorderly conduct, narcotics possession, petty theft, simple assault). Felony arrests often necessitate an arrest, whereas misdemeanor arrests involve greater officer discretion. Officers have more latitude in deciding whether to arrest based on their interpretation of the situation, the severity of the offense, and the suspect's demeanor.

Table 9 presents results for our additional arrest-related outcomes. Panel A shows results for more discretionary arrests, while Panel B focuses on more serious, less discretionary arrests. Our analysis indicates that FTO force rates primarily affect lower-level enforcement actions, though the effects are modest. In Panel A, a one standard deviation increase in an FTO's force rate is associated with a roughly 2 percent rise in misdemeanor arrests and a statistically significant 2.8 percent increase in unfiled arrests (those declined by prosecutors, often due to insufficient evidence). Conversely, Panel B shows minimal impact on more serious enforcement. The estimates for felony and filed arrests are small, inconsistent in sign, and statistically insignificant. For instance, the fully controlled model for felony arrests (column 3) shows a near-zero effect of -0.06 percent, while filed arrests show only a 0.19 percent effect.

These findings align with our broader results, which show that high-force FTOs influence recruits to engage in more discretionary and less justifiable uses of force rather than fundamentally altering their overall enforcement activity. Although discretionary arrests can

be a tool for proactive policing, the absence of corresponding increases in overall and felony arrest volumes suggests they are not the primary driver of our main results. The fact that we observe no increase in felony arrests or total arrests suggests that the primary mechanism at play is more likely a transfer of behavioral norms regarding how recruits exercise discretion during an incident. We believe that these results are more consistent with our preferred explanation that FTOs shape a recruit's on-scene decision-making, rather than selection into particular types of incidents or proactive enforcement.

6 Evidence Against Alternative Mechanisms

Our findings thus far suggest that high-force FTOs influence recruits to adopt a lower threshold for using force, potentially through the transmission of norms regarding acceptable police behavior. This proposed mechanism implies that FTOs convey information about appropriate levels of aggression in different situations. Consistent with this mechanism, we have documented that recruits with high-force FTOs select into marginally more dangerous calls for service, engage in more discretionary uses of force, and make more arrests for lower-level charges that often do not lead to prosecution. To further support our preferred explanation, we now rule out several plausible alternative mechanisms. Specifically, we address: (1) the possibility that omitted FTO characteristics are correlated with FTO force and a recruit's subsequent force; (2) whether our main findings are driven by more active forms of policing; (3) whether recruits are more likely to be dispatched to calls for service with their FTO even after completing training; (4) whether recruits paired with a high-force FTO are indeed witnessing force during the training period; (5) differential behavior in terms of reporting force; and (6) whether our results are driven by under-reporting of less severe force incidents. As we will demonstrate, these additional analyses provide little support for alternative mechanisms which further reinforces the notion that knowledge transfer about the appropriate use of force is occurring during a critical phase of on-the-job training.

First, we explore whether FTO demographics are predictive of force rates. In Figure A.4, we plot the distribution of force rates by FTO characteristics. On average White FTOs have higher force rates than Hispanic and Black FTOs but none of the distributions are statistically different from each other.[40] Figure A.4b shows that female FTOs are about one-third of a standard deviation more likely to use force than males and these two distributions are statistically different from each other (K-Smirnov p-value $= 0.002$).[41] Figure A.4a also shows that younger FTOs (i.e., less than the mean age) use force about 0.13 standard deviations more frequently than older FTOs and the two distributions are statistically different from each other). Finally, we regress our FTO force measure on FTO race, age, and gender. This regression yields an R-squared of 0.014 and an F-test p-value of 0.56 , suggesting these officer-level attributes play a limited role in explaining force usage patterns. 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 3 (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. [42]

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

[40] We also note that there are few Black and Hispanic FTOs constituting only 96 and 76 recruit-FTO pairs respectively.

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

[42] An alternative approach could involve replacing the FTO force rate with other FTO behavioral measures as the key independent variable. We instead prioritize a framework that directly isolates the effect of force transmission by controlling for other FTO rates, some of which are correlated with recruit force. This choice reflects a preference for focusing on the central research question of whether exposure to a high-force FTO alters a recruit's threshold for using force, independent of other factors. By controlling for these other FTO characteristics, we ensure that our main estimates are not confounded by other omitted variables, thereby strengthening the argument that force-related norms—not other FTO attributes—drive our findings.

as response times and time spent on a call.[43] 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 leads to a 0.46 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, time spent on a call and force incidents with an officer injury. 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. A 1 SD increase in the FTO force rate corresponds to a 1/3 SD increase in officer injury. This correlation is partly mechanical, as an increase in force naturally raises the likelihood of injury. In our later analysis, we will examine whether our results are entirely driven by FTO injury. We also consider FTO complaints. FTO force rate is not correlated with overall levels of complaints. However, there is a correlation between use of force and internal complaints and FTO use of force (see Table 5). While there is likely to be a mechanical relationship between the volume of force incidents and complaints associated with force, we cannot rule out other potential stories and so include this measure as a control in some of our models. Since we observe that FTO force is correlated with other measures that capture a more active form of policing, we now formally control for arrest rates, misdemeanor arrest rates, response time, time on call, complaints, and injury in rows 5-10 of 6. Overall, these estimates are similar in size and significance to our main results shown again in row 1.

We include all FTO characteristics (rows 2-4) as well as other measures of FTO proactive policing (rows 5-10 of Table 6. Our estimate is very similar 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

[43]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).

other FTO characteristics or proactive policing, is the most likely explanation for our results. Row 10 reports results after controlling for the FTO Force and Injury rate. Since force is a prerequisite for injury, we expect this control to affect our estimate. Nevertheless, even after accounting for Force and Injury rate, our estimate remains statistically significant, suggesting that injury exposure alone does not fully explain our findings.

Third, we examine whether our results are driven by who a recruit is dispatched with after training. First, we calculate the officer a recruit is most likely to be dispatched with after training. Sixteen recruits are most likely to be dispatched with their FTO. If we drop those calls from our sample, our Table 3 column 3 (full controls) estimate is a similar magnitude and is statistically significant at the 1 percent level. Thus, we it is unlikely that our results are driven by of recruits continuing to respond to calls with their FTOs after completing field training.

Next, we analyze the influence of a recruit's temporary partner during the six-month probationary period following field training, known as "little T," in Table 7. Unlike FTO assignments, these partnerships are not determined by command staff but instead chosen by probationary officers. We define a recruit's partner as the officer they most frequently respond with during this period, though it is possible that recruits continue to select into calls with this officer beyond probation. In the odd-numbered columns, we assess the recruit's own use of force, while the even-numbered columns examine whether FTO force rates influence the force rates of a recruit's partners. Column (2) suggests a marginally significant relationship, with a coefficient of 0.0726 ($p<0.1$), implying that recruits trained by high-force FTOs may subsequently partner with officers who also have higher force rates. However, this relationship weakens and becomes statistically insignificant when recruit characteristics (column 4) and call controls (column 6) are included. Importantly, despite potential selection into a temporary partner, our main results on recruit force remain robust. Compared to Table 3, controlling for these temporary partners reduces the coefficient from 0.000206 to 0.000185, a decrease of approximately 10 percent, indicating that partner selection accounts

for only about one-tenth of the main effect. Adding recruit characteristics further reduces the coefficient to 0.000182, and including call controls decreases it to 0.000132. Nevertheless, the coefficient remains statistically significant at the 5 percent level, underscoring the persistence of the FTO effect. While our identification strategy does not extend to partner characteristics, and their corresponding force rate is endogenous, Table 7 demonstrates that partners do not fully explain our findings. The enduring effect of FTOs on recruit force use, even when accounting for temporary partner selection, suggests that training exposure has a lasting impact beyond merely choosing a more aggressive partner.

Fourth, we ask whether recruits paired with more aggressive FTOs might be more likely to experience force during their training periods. Said another way, do recruits and FTOs experience a correlated shock (a force incident) that might explain our results. This early exposure to a force event could drive our results. One hundred and twelve recruits (21 percent) experienced a force incident during training. To investigate this explanation, we allow our effect to vary by whether a recruit experienced force during their training. Figure A.5 shows that results hold no matter if a recruit experienced force during their training period. Within each specification (following specifications in Table 3), coefficients are not statistically different from each other.

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.[44]

To explore differences in report-writing behavior, we develop three measures based on the section of incident reports completed by the responding officer. [45] Following the procedure described in 3.2, we calculate measures to 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 8, we repeat the three columns presented in Table 3 but with the addition of our reporting controls. Across each column each of the estimates are similar in magnitude and significance to Table 3.[46]

The Dallas Police department also introduced bodyworn cameras during our study period and implemented a policy of random audits. If Dallas police were systematically under-reporting force, we should expect to observe that the total volume of force would increase following the introduction of bodyworn cameras and the random auditing procedure. Figure A.6 plots the average daily number of force incidents over time.[47] We note three important dates which include the purchase of police cameras and additional training for police cameras that could affect reporting. The figure demonstrates that there is no clear change in the volume of force incidents reports across these dates. This finding is consistent with Dallas PD's statements that under-reporting is not an issue within their department.

Overall, our analyses provide little support for alternative mechanisms such as selection into specific calls, prolonged FTO influence after training, or broader proactive policing

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

[45]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.

[46]There are 5 FTOs that we cannot link to incident reports. We assign them the average rate.

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

tendencies. While we find some correlations between FTO force rates and measures of policing style—such as arrest rates and time spent on calls—controlling for these factors does not meaningfully alter our main results. Similarly, we find no evidence that continued exposure to an FTO after training, early exposure to force incidents, or differential reporting practices explain our findings. Although we cannot rule out the possibility that some degree of proactive policing or call selection is captured within our treatment effect, the lack of an increase in overall arrests or felony arrests suggests that the primary mechanism at play is not just proactive enforcement. Instead, we argue that the evidence is consistent with our preferred explanation of knowledge transfer during field training, where recruits adopt their FTO's threshold for applying force in discretionary situations. Even if we cannot pinpoint the exact mechanism with certainty, our findings provide clear causal evidence that supervisors shape officer behavior. This underscores the critical role of FTOs in establishing enforcement norms, whether through direct instruction, implicit modeling, or other influences.

7 Conclusion

This study provides new evidence that field training officers (FTOs) play a crucial and persistent role in shaping how new police officers make decisions about when to apply force. In particular, our analysis of administrative data from the Dallas Police Department demonstrates that being assigned to a more aggressive FTO leads to a significant increase in a recruit's subsequent likelihood of using force which persists for about two years. The magnitude of these effects is large - a one standard deviation increase in FTO force propensity (124 percent relative to mean) leads to a 14 to 18 percent increase in recruit force use. Our coefficient estimates translate to an elasticity of between 0.08 to 0.15 which meets or exceeds previous estimates of peer effects in policing by Holz et al. (2023) (0.09) and Rivera (2025) (0.12), suggesting that hierarchical training relationships may be particularly important in

transmitting behaviors.[48] This finding underscores the powerful influence of supervisors and on-the-job training on future high-stakes decision-making, especially in contexts requiring considerable discretion.

Importantly, the effects of FTO assignment persist for approximately two years before dissipating. The two-year window of influence carries particular significance as policing agencies face a national policing shortage (Police Executive Research Forum, 2022) and have confronted increasing turnover rates since 2011 (Grunwald, 2024). As new officers enter the force and eventually transition into training roles, there is a critical and low-cost opportunity for policing agencies to reshape their culture through strategic FTO selection. Our results suggest that targeted interventions in the FTO selection process could be a promising avenue for police reform. By strategically assigning recruits to less forceful FTOs, departments could potentially achieve meaningful reductions in use-of-force incidents. For instance, simulations of our data and estimates indicate that replacing the top tercile of FTOs with those from the lowest tercile could lead to a 10.6 percent reduction in overall force (from 1,395 to about 1,247 force incidents or about 68 incidents per year) among recruits.[49] However, it is important to note that these back-of-the-envelope calculations are likely too conservative. In particular, we focus only on aggregate recruit force and do not account for the dynamic effects of training lower-force officers who eventually become FTOs. Nonetheless, we believe that this illustrate the potential for even modest interventions related to FTO assignments to generate substantial long-term impacts.

[48] Our estimates also align closely with seminal studies of workplace peer effects that find elasticities between 0.14-0.15 (Mas and Moretti, 2009; Falk and Ichino, 2006), indicating that FTOs may be as influential as traditional peers in shaping worker behavior. Mas and Moretti (2009) report a 10 percent increase in coworker permanent productivity is associated with a 1.5 percent increase in reference worker productivity which translates to an elasticity of 0.15. Falk and Ichino (2006) report a 10 percent increase in peer output results in a 1.4 percent increase in individual productivity which translates to an elasticity of 0.14.

[49] We arrive at these estimates by conducting 1,000 simulations where we replace FTOs from the highest tercile of forcefulness with those from the lowest tercile. We predict each recruit's use of force in the period after training by apportioning it using the coefficient (0.000157) from specification 3 of Table 3 multiplied by the difference between their real and simulated FTO force relative to the global mean (0.00116).

As we discussed at length in the introduction, this study contributes to a growing body of literature examining the influence of supervisors and on-the-job training on worker behavior, particularly in high-stakes environments. By focusing on police use of force, we provide novel evidence on how training relationships shape complex, discretionary decisions with potentially severe consequences. We also contribute to an existing literature on on-the-job training as well as apprenticeship-style training models by providing causal evidence of persistent impacts in a particularly consequential segment of the public sector. We strongly believe that our findings underscore the need for more research on the topic of on-the-job training more generally as well as within policing. Within policing, we believe future research should explore the impact of field training and supervisors on other important outcomes like police misconduct and disparate treatment. Related to on-the-job training more generally, further investigation is needed on whether these findings generalize to other public service professions with similar high-stakes environments and training models, including healthcare and education. Lastly, we emphasize that our study underscores the dire need for greater investment in evidence-based training and mentorship models that promote effective and responsible 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.

Baltrunaite, A., G. Bovini, and S. Mocetti (2020). Opening the black-box of managerial talent. *Journal of Corporate Finance*.

Bandiera, O., A. Prat, S. Hansen, and R. Sadun (2020). Ceo behavior and firm performance. *Journal of Political Economy 128*(4), 1325–1369.

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

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

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

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

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

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

Black, D. A., L. Skipper, and J. A. Smith (2023). Firm training. *Handbook of the Economics of Education 7*, 287–468.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does management matter? evidence from india. *The Quarterly journal of economics 128*(1), 1–51.

Bloom, N., R. Lemos, R. Sadun, and J. Van Reenen (2015). Does management matter in schools? *The Economic Journal 125*(584), 647–674.

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.

Buechtemann, C. F., J. Schupp, and D. J. Soloff (1993). Determinants of employment stability in germany: Labor market and policy context. *Industrial Relations Journal 24*(2), 97–111.

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.

Cooke, L. P. (2003). A comparison of initial and early life course earnings of the german secondary education and training system. *Economics of Education Review 22*(1), 79–88.

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

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

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

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

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

Fersterer, J., J.-S. Pischke, and R. Winter-Ebmer (2008). Returns to apprenticeship training in austria: Evidence from failed firms. *Scandinavian Journal of Economics 110*(4), 733–753.

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.

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

Grunwald, B. (2024, June). A large-scale study of the police retention crisis. Working Paper.

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

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

Hollenbeck, K. (2008). State use of workforce system net impact estimates and rates of return. *Technical report, Upjohn Institute*.

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

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

Jackson, C. K. and E. Bruegmann (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics 1*(4), 1–27.

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.

Mello, S., M. Ross, S. Ross, and H. Johnson (2024). Diversity training and employee behavior: Evidence from the police. *Working Paper*.

Minni, V. (2023). Making the invisible hand visible: Managers and the allocation of workers

to jobs. *Working Paper*.

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.

Police Executive Research Forum (2022, March). Perf survey shows steady staffing decrease over the past two years. Research report, Police Executive Research Forum, 1120 Connecticut Ave. NW Suite 930 Washington, DC 20036.

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

Rasul, I. and D. Rogger (2018). Management of bureaucrats and public service delivery: Evidence from the nigerian civil service. *The Economic Journal 128*(608), 413–446.

Reed, D., A. Y.-H. Liu, R. Kleinman, A. Mastri, D. Reed, S. Sattar, and J. Ziegler (2012). An effectiveness assessment and cost-benefit analysis of registered apprenticeship in 10 states. Technical report, Mathematica Policy Research.

Rivera, R. (2025). Do peers matter in the police academy? *American Economic Journal: Applied Economics (Forthcoming)*.

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

Simes, R. J. (1986). An improved bonferroni procedure for multiple tests of significance. *Biometrika 73*(3), 751–754.

Tsai, T. C., A. K. Jha, A. A. Gawande, R. S. Huckman, N. Bloom, and R. Sadun (2015). Hospital board and management practices are strongly related to hospital performance on clinical quality metrics. *Health Affairs 34*(8), 1304–1311.

U.S. Department of Justice (2021). 2021 Bias-Free Policing Annual Report: New Orleans

Police Department Consent Decree.

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. (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*.

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

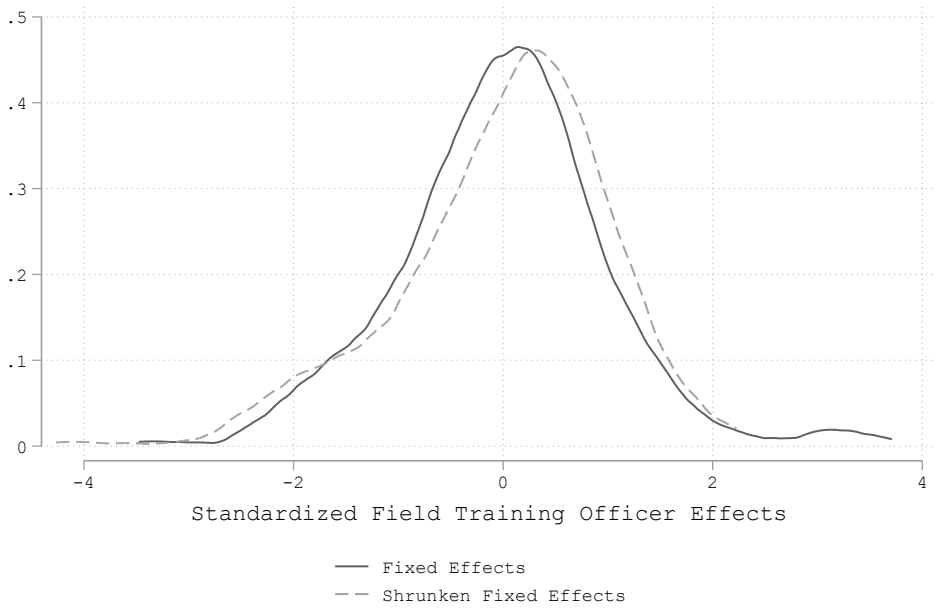
Wolter, S. C. and P. Ryan (2011). Apprenticeship. In *Handbook of the Economics of Education*, Volume 3, pp. 521–576. Elsevier.

Figures and Tables

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

(a) Field Training Officer Effects

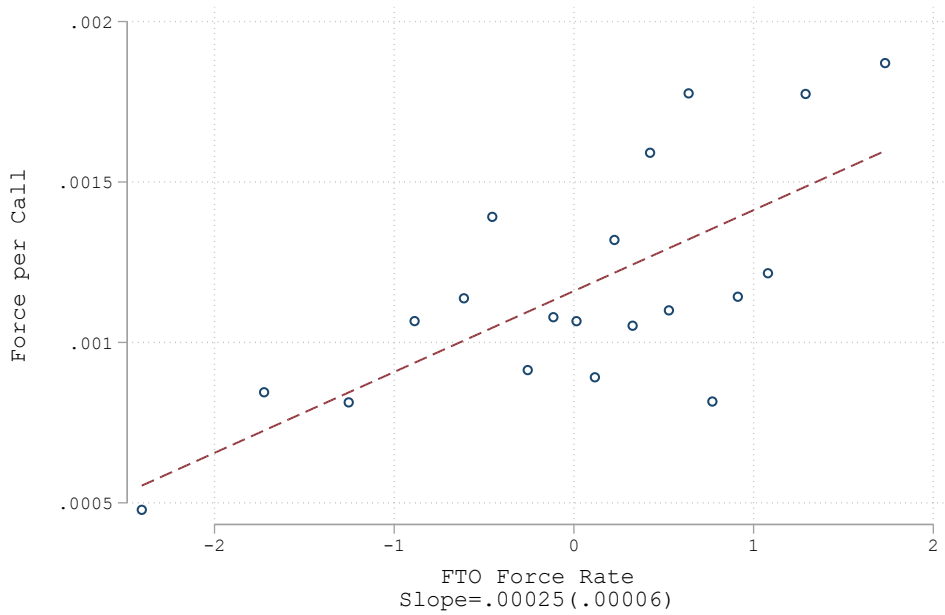
(b) Standardized Field Training Officer Effects



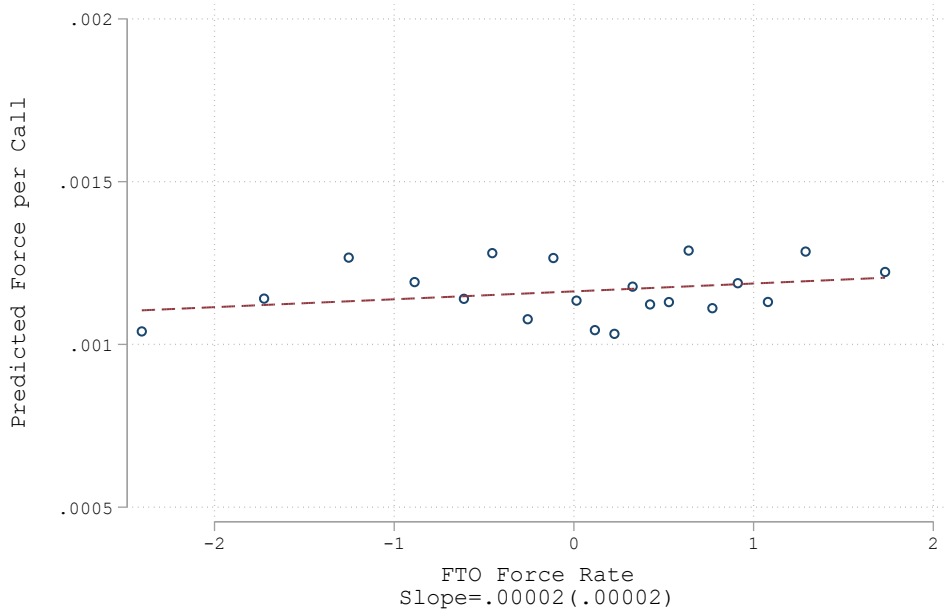
Notes: These figures display the distribution of Field Training Officers' (FTOs) propensity to use force across 521 FTO-recruit pairs. The solid line shows raw fixed effects while the dashed line shows shrinkage-adjusted estimates using Empirical Bayes methods. Fixed effects control for call characteristics: number of officers present, police beat, type of call (48 priority by type groups), year-month, and day-of-week-by-night. Values above zero indicate FTOs with above-average force usage, while negative values indicate below-average force usage. The standardized effects show that a one standard deviation increase in FTO force propensity corresponds to a 124 percent increase in force incidents relative to the mean.

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

(a) Use of Force

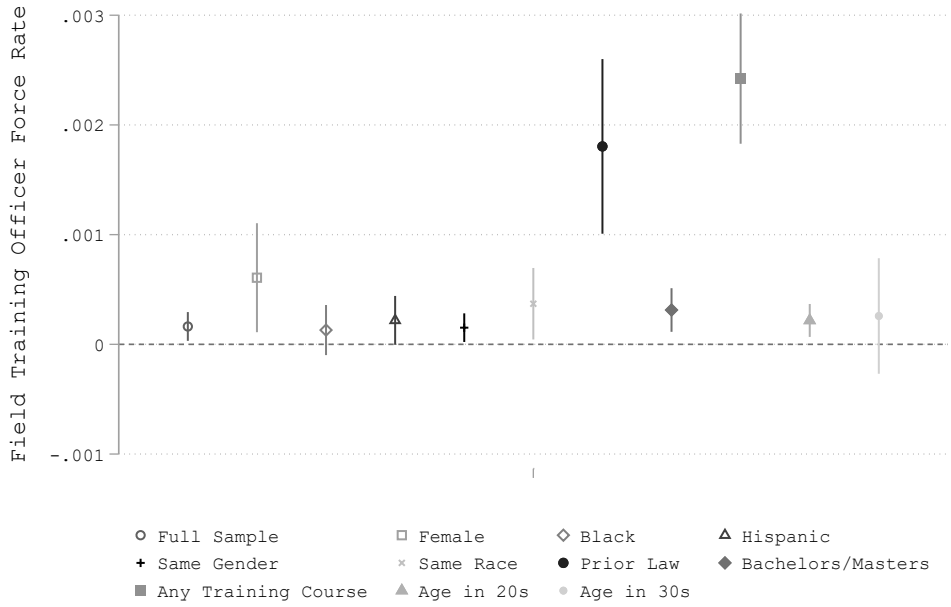


(b) Predicted Use of Force

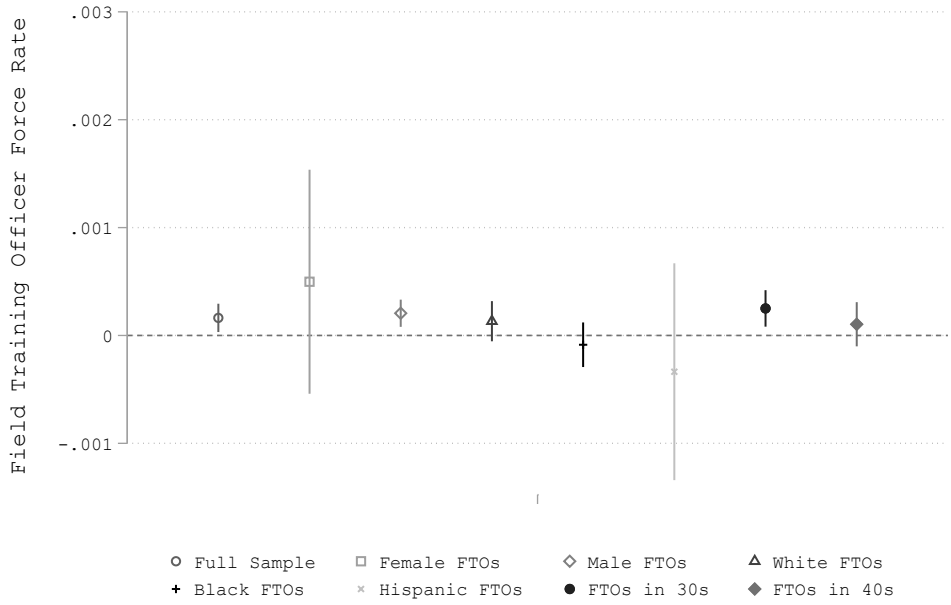


Notes: These figures show the relationship between FTOs' propensity to use force and their recruits' subsequent behavior. Panel (a) plots the actual force used by recruits against their FTO's force rate. Panel (b) plots predicted force, which is calculated in three steps: (1) residualizing recruit use of force on cohort and assignment fixed effects, (2) regressing these residuals on call characteristics - which include controls for number of officers present, police beat, type of call (48 priority by type groups), year-month, and day-of-week-by-night, and (3) using this model to predict force likelihood for each call, adding back the average force rate. Each point represents an equal-sized bin of observations. The red dashed line shows the linear fit.

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



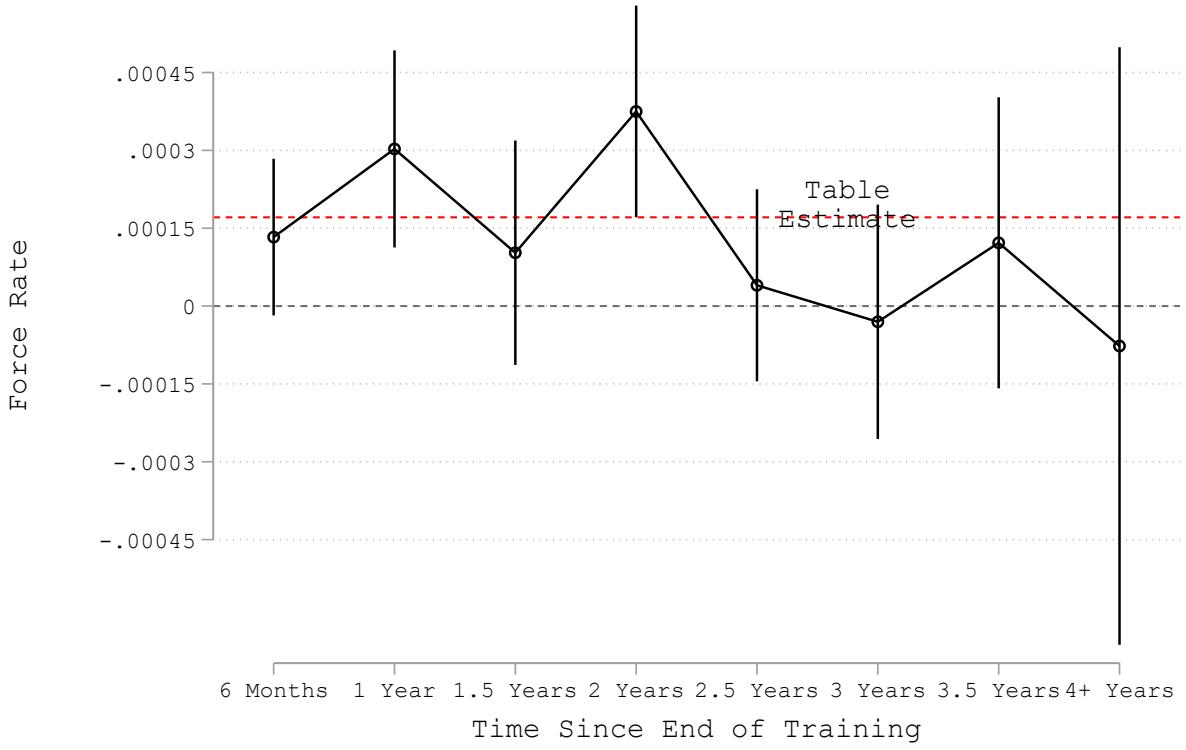
(a) Recruit Subgroups



(b) FTO Subgroups

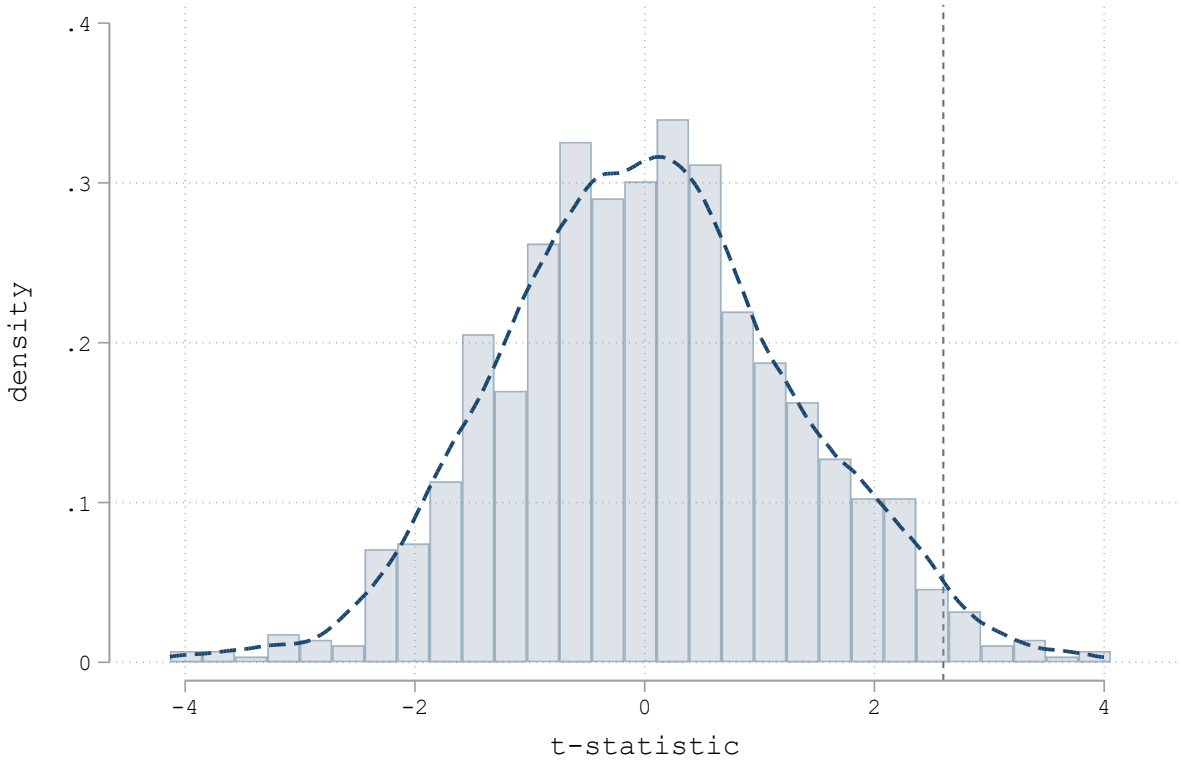
Notes: These figures show how Field Training Officers' impact on use of force varies across different subgroups. Panel (a) shows effects across recruit characteristics: recruits' gender (Female), race (Black, Hispanic), pairing characteristics (Same Gender, Same Race), and other characteristics (Prior Law Enforcement, Bachelors/Masters Degree, Any Training Course, Age in 20s, Age in 30s). Panel (b) shows effects across FTO characteristics: gender, race, and age. Each point represents a coefficient from a separate regression, with vertical lines showing standard errors clustered at the recruit and FTO level. Fixed effects control for number of officers present, police beat, type of call (48 priority by type groups), year-month, and day-of-week-by-night.

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



Notes: This figure shows how the effect of Field Training Officers' use of force on recruits evolves over time since the end of training. Each point represents the effect for a 6-month period, from 6 months up to 4+ years after training completion. The red dashed line marks the baseline estimate from the main analysis. Vertical lines show standard errors clustered at the recruit and FTO level. The analysis includes controls for call characteristics (number of officers present, police beat, type of call, year-month, and day-of-week-by-night) and recruit characteristics. The decreasing coefficients after two years coincide with officers typically becoming eligible for promotion to senior corporal.

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 3 by using 1000 randomized simulations of the data generating process. Only 37 of the simulations resulted in a t-statistic more extreme than our baseline estimate, i.e. two-sided test p-value of 0.037. Of these, 23 simulations had t-statistics strictly greater than our baseline estimate, corresponding to a one-sided p-value of 0.023. The dashed line marks our original t-statistic

Table 1: Call Level Summary Statistics (Outcomes)

	(1) Entire Sample	(2) High Force Trainer	(3) Low Force Trainer
Force	0.00116	0.00132	0.000943
	(0.0341)	(0.0364)	(0.0307)
Force with No Arrest	0.000511	0.000588	0.000405
	(0.0226)	(0.0242)	(0.0201)
Force Unjustified	0.0000860	0.000104	0.0000618
	(0.00928)	(0.0102)	(0.00786)
Force & Civilian Injured	0.000158	0.000188	0.000116
	(0.0126)	(0.0137)	(0.0108)
Any Arrest	0.0363	0.0366	0.0358
	(0.187)	(0.188)	(0.186)
Misd. Arrest	0.0212	0.0219	0.0202
	(0.144)	(0.146)	(0.141)
Felony Arrest	0.00731	0.00726	0.00739
	(0.0852)	(0.0849)	(0.0857)
Filed Arrest	0.0144	0.0143	0.0146
	(0.119)	(0.119)	(0.120)
Unfiled Arrest	0.0141	0.0149	0.0130
	(0.118)	(0.121)	(0.113)
Observations	1198564	681957	516607
Additional Statistics			
Average Force per Recruit per Year	1.5		
Average Force per Recruit	3.25		

Notes: This table reports call-level summary statistics (means and standard deviations). High Force FTOs have higher force rates than the average FTO, and Low Force FTOs have lower force rates than the average FTO. Force Unjustified refers to force used when civilian resistance was minimal, specifically categorized as "furtive gesture," "not following commands," "null," or "off safety-no resistance." There are three types of arrests in our dataset: felony, misdemeanor, or n-class. Most n-class arrests are for outstanding warrants. We also categorize felony and misdemeanor arrests as filed or unfiled. If an arrest is unfiled, the district attorney decided not to move forward with the case, and the defendant will not be charged with a crime.

Table 2: Balance Test

Recruit Chars	Age	Female	Black	Hispanic	Hire Date	FTO Force	Recruit Force	Recruit Force NC
Female	-0.4661	0.0137	0.0620	0.0342	0.3789	0.0000	-0.0000	-0.0000
	(1.3053)	(0.0445)	(0.0529)	(0.0481)	(1.1767)	(0.0001)	(0.0001)	(0.0001)
Black	-0.5571	0.0634	0.0867	-0.0770*	0.1093	-0.0001	-0.0001	-0.0002**
	(1.4275)	(0.0477)	(0.0585)	(0.0460)	(1.2146)	(0.0001)	(0.0001)	(0.0001)
Hispanic	0.5837	0.0544	-0.0140	0.0328	-0.7922	-0.0000	-0.0000	-0.0001
	(1.1288)	(0.0436)	(0.0522)	(0.0439)	(0.9791)	(0.0001)	(0.0001)	(0.0001)
Prior Law	0.7052	-0.0278	-0.0892	-0.0915	-0.7689	0.0002	0.0003**	0.0003**
	(2.2380)	(0.0811)	(0.0991)	(0.0724)	(1.8674)	(0.0001)	(0.0001)	(0.0001)
Bachelors	-1.4008	0.0213	0.0257	0.0131	0.6346	-0.0000	-0.0001	-0.0001
	(1.0387)	(0.0329)	(0.0472)	(0.0362)	(0.8823)	(0.0001)	(0.0001)	(0.0001)
Masters	-4.0827	-0.0469	0.0536	0.0912	2.5177	-0.0001	0.0001	0.0000
	(2.7617)	(0.0868)	(0.1226)	(0.1259)	(3.1131)	(0.0002)	(0.0002)	(0.0002)
Descalation Training	1.5722	-0.0430	-0.1082	0.2807*	2.5551	-0.0003	-0.0006**	-0.0005**
	(4.6316)	(0.1501)	(0.1284)	(0.1545)	(4.2796)	(0.0002)	(0.0003)	(0.0003)
Crisis Intervention Training	0.6236	0.1982	0.1823	-0.1812*	-1.6883	0.0000	-0.0003	-0.0004
	(3.7214)	(0.1853)	(0.1264)	(0.1084)	(4.0741)	(0.0002)	(0.0003)	(0.0003)
Age 30s	-0.1551	0.0361	0.0117	-0.0192	0.9462	-0.0000	0.0000	-0.0001
	(1.1352)	(0.0436)	(0.0483)	(0.0430)	(1.0836)	(0.0001)	(0.0001)	(0.0001)
Age 40s	1.0537	0.0870	0.1627	0.1003	-0.1839	0.0002	-0.0001	-0.0002**
	(3.2497)	(0.1276)	(0.1657)	(0.1519)	(2.7487)	(0.0002)	(0.0001)	(0.0001)
Div-x-Cohort FE	X	X	X	X	X	X	X	X
Observations	521	521	521	521	521	521	521	521
Outcome Mean	42.1382	0.1248	0.1843	0.1459	41.6620	0.0001	-0.0000	0.0001
F-Test P-Value	0.885	0.883	0.269	0.157	0.959	0.240	0.318	0.028
F-Test P-Values for Alternative Specifications								
Pairwise Omnibus Test	0.7953							
T-test Exercise	Reject 12/717 t-tests							

Notes: Columns report coefficients from regressing each FTO characteristic on all recruit characteristics (with FTO-clustered SEs and division-by-cohort fixed effects). Column F-tests assess joint significance. The bottom panel shows alternative specifications (robust and division-by-cohort clustered SEs). The pairwise omnibus test combines all 60 FTO-recruit characteristic pairs in a stacked regression with pair and division-by-cohort-by-pair fixed effects, and the t-test exercise conducts cohort-specific tests (n=620) using Simes (1986) correction. The outcome "Recruit Force NC" is the recruit force rate excluding call controls.

Table 3: The Effect of High Force Field Training Officers on Recruit Use of Force

	(1) Force	(2) Force	(3) Force
FTO Force Rate	0.000206***	0.000203***	0.000157***
	(0.0000658)	(0.0000588)	(0.0000604)
Observations	1198564	1198564	1198564
Outcome Mean	0.00116	0.00116	0.00116
Assigned Div by Cohort FE	Y	Y	Y
Recruit Characteristics	-	Y	Y
Call Controls	-	-	Y

Standard errors in parentheses

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

Notes: This table presents the effect of FTO force rate on recruit use of force. Our primary regression specification is given by $\text{Force}_{r,c} = \theta_r + \beta_1 \Lambda_{o(r)} + \beta_2 X_c + \epsilon_{r,c}$ (Equation 3), where β_1 represents the effect of a one standard deviation increase in FTO force rate. Standard errors are multiway clustered at the field training officer and recruit levels. Column 2 includes controls for recruit characteristics (age, gender, race, prior law enforcement experience, education, and training course participation), while Column 3 further adds call characteristic fixed effects (number of officers on the scene, beat, call type—priority-by-type, year-by-month, and day-of-week-by-night interactions).

Table 4: The Effect of High Force Field Training Officers on Recruit Use of Force (by Type)

	Panel A: Force with No Arrest		
FTO Force Rate	0.0000836***	0.0000816***	0.0000624**
	(0.0000317)	(0.0000288)	(0.0000298)
Observations	1198564	1198564	1198564
Outcome Mean	0.000511	0.000511	0.000511
	Panel B: Force with Citizen Injury		
FTO Force Rate	0.0000302	0.0000312*	0.0000240
	(0.0000190)	(0.0000165)	(0.0000164)
Observations	1198564	1198564	1198564
Outcome Mean	0.000158	0.000158	0.000158
	Panel C: Less Justifiable Force		
FTO Force Rate	0.0000229*	0.0000217*	0.0000188*
	(0.0000125)	(0.0000111)	(0.0000108)
Observations	1198564	1198564	1198564
Outcome Mean	0.0000860	0.0000860	0.0000860
Assigned Div by Cohort FE	Y	Y	Y
Recruit Characteristics	-	Y	Y
Call Controls	-	-	Y

Notes: This table presents the effect of being assigned a higher force FTO on different types of recruit use of force. Standard errors are multiway clustered at the field training officer and recruit levels. Column 2 includes controls for recruit characteristics, such as age, gender, race, prior law enforcement experience, education, and training course participation. Column 3 further adds call characteristic fixed effects, including the number of officers on the scene, beat, call type (priority-by-type), year-by-month, and day-of-week-by-night interactions. Less Justifiable Force is classified by the authors based on recorded civilian resistance i.e., when civilian resistance was minimal, specifically categorized as "furtive gesture," "not following commands," "null," or "off safety-no resistance."

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	0.457***	0.440***	0.338***	0.341***	0.496***
	(0.0678)	(0.0722)	(0.0602)	(0.0644)	(0.0722)
Observations	521	521	521	521	521

	Response Rate	Time on Call	Force and Injury Rate
Panel B:			
FTO Timing & Injury			
Force Rate	-0.0427	-0.172***	0.368***
	(0.0756)	(0.0507)	(0.0559)
Observations	521	521	521

	Complaints	UOF Complaints	Internal Complaints
Panel C:			
FTO Complaints			
Force Rate	0.0274	0.132**	0.120**
	(0.0501)	(0.0535)	(0.0486)
Observations	521	521	521

	Num. Characters Rate	Num. Words Rate	Write Nothing Rate
Panel D:			
FTO Reporting			
Force Rate	-0.0996	-0.0885	0.00487
	(0.0714)	(0.0704)	(0.0707)
Observations	514	514	514
Div-X-Cohort FE	Y	Y	Y

Standard errors in parentheses

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

Notes: This table presents correlations between FTO force rate and other characteristics, estimated via regression of FTO characteristics on FTO force race. All variables are standardized, and standard errors are clustered at the field training officer level. Response time is the hours between arrival and assignment. Time on call is the hours from enroute to call clearance. We cannot link complaints to calls for service, so we compute complaints per call. Filed arrests were officially submitted to the District Attorney's office, whereas unfiled arrests refer to incidents where an arrest was made but not filed with the DA's office.

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

Control Variable	Coefficient (FTO Force Rate)	Standard Error	p-value
Original Estimate	0.000157	(0.0000604)	0.009***
FTO Demographics			
FTO Gender	0.000165	(0.0000615)	0.007***
FTO Race	0.000164	(0.0000613)	0.007***
FTO Age	0.000167	(0.0000591)	0.005***
Other FTO Rates			
FTO Arrest Rate	0.000196	(0.0000628)	0.002***
FTO Misd Arrest Rate	0.000176	(0.0000616)	0.003***
FTO Response Time Rate	0.000156	(0.0000606)	0.012**
FTO Time on Call Rate	0.000151	(0.0000599)	0.012**
FTO Complaint Rate	0.000129	(0.0000600)	0.032**
FTO Force & Injury Rate	0.000120	(0.0000718)	0.075*
Recruit On-Call Actions			
Recruit Arrest	0.000152	(0.0000588)	0.010**
Recruit Arrest Type	0.000151	(0.0000583)	0.010**
Kitchen Sink			
Kitchen Sink Model	0.0001657	(0.0000586)	0.005***
Outcome Mean	0.00116		
Observations	1,198,564		
Assigned Div by Cohort FE	Y		
Recruit Characteristics	Y		
Call Controls	Y		

Standard errors in parentheses

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

Notes: This table presents the effect of the FTO force rate on recruit use of force. Standard errors are clustered at both the recruit and FTO levels. Each specification includes controls for recruit characteristics and fixed effects for call characteristics, including the number of officers on the scene, beat, call type (priority-by-type), and time variables (year-by-month and day-of-week-by-night). Row 1 replicates Row 3 of Table 3. Rows 2–4 incorporate additional FTO characteristics, while Rows 5–10 introduce controls for other FTO rates (see Appendix B for details). Rows 11 and 12 account for whether a recruit made an arrest during the call and the type of arrest.

Table 7: **Partners**: The Effect of High Force Field Training Officers on Recruit Partners

	(1) Force	(2) Partner Force Rate	(3) Force	(4) Partner Force Rate	(5) Force	(6) Partner Force Rate
FTO Force Rate	0.000185***	0.0726*	0.000182***	0.0663	0.000132**	0.0650
	(0.0000609)	(0.0427)	(0.0000564)	(0.0443)	(0.0000561)	(0.0440)
Observations	1194016	1194016	1194016	1194016	1194016	1194016
Outcome Mean	0.00116	0.0259	0.00116	0.0259	0.00116	0.0259
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

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

Notes: This table presents the effect of the Field Training Officer (FTO) force rate on the recruit's partner's force rate and the recruit's own use of force. Standard errors are multiway clustered at the recruit and FTO levels. Odd-numbered columns use recruit force as the outcome, while even-numbered columns use the recruit's partner's force rate. Only odd-numbered columns control for the recruit's most frequent partner post-training (the officer observed with them most in little t), including partner gender, race, and force rate. Columns (1) and (2) provide baseline estimates, controlling only for assigned division by cohort fixed effects. Columns (3) and (4) add recruit characteristics (age, gender, race, education, prior law enforcement experience, and de-escalation training). Columns (5) and (6) further introduce call controls (number of officers on the scene, beat, type of call, year-by-month, and day-of-week-by-night fixed effects).

Table 8: **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	0.000207***	0.000204***	0.000159***	0.000208***	0.000205***	0.000160***	0.000206***	0.000204***	0.000156**
	(0.0000658)	(0.0000590)	(0.0000605)	(0.0000656)	(0.0000588)	(0.0000604)	(0.0000661)	(0.0000588)	(0.0000605)
Observations	1198564	1198564	1198564	1198564	1198564	1198564	1198564	1198564	1198564
Outcome Mean	0.00116	0.00116	0.00116	0.00116	0.00116	0.00116	0.00116	0.00116	0.00116
Div-Cohort FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Recruit Chars	-	Y	Y	-	Y	Y	-	Y	Y
Call Controls	-	-	Y	-	-	Y	-	-	Y
FTO # Characters	Y	Y	Y	-	-	-	-	-	-
FTO # Words	-	-	-	Y	Y	Y	-	-	-
FTO Write Nothing	-	-	-	-	-	-	Y	Y	Y

Standard errors in parentheses

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

Notes: This table presents the effect of FTO force rate on recruit use of force. Standard errors are clustered at both the recruit and FTO levels. We also control for an FTO's report writing behavior—measured by the number of characters/words used and the frequency of leaving the section blank—using the "Modus Operandi (MO)" field from the incident report data. This open-ended field is completed by the reporting officer, and the resulting MO rates are calculated similarly to the force rates. Further details are provided in Appendix B.

Table 9: **Other Outcomes**: The Effect of High Force Field Training Officers on Recruit Arrests

	(1) Misd. Arrest	(2) Misd. Arrest	(3) Misd. Arrest	(4) Unfiled Arrest	(5) Unfiled Arrest	(6) Unfiled Arrest
Panel A: Less serious, More discretionary						
FTO Force Rate	0.000585	0.000596	0.000415	0.000524*	0.000609*	0.000395*
	(0.000367)	(0.000385)	(0.000297)	(0.000304)	(0.000318)	(0.000223)
Observations	1198564	1198564	1198564	1198564	1198564	1198564
Outcome Mean	0.0212	0.0212	0.0212	0.0141	0.0141	0.0141
	(1) Felony Arrest	(2) Felony Arrest	(3) Felony Arrest	(4) Filed Arrest	(5) Filed Arrest	(6) Filed Arrest
Panel B: More serious, Less discretionary						
FTO Force Rate	0.0000583	0.0000592	-0.00000455	0.000131	0.0000595	0.0000273
	(0.000148)	(0.000142)	(0.000122)	(0.000247)	(0.000244)	(0.000218)
Observations	1198564	1198564	1198564	1198564	1198564	1198564
Outcome Mean	0.00731	0.00731	0.00731	0.0144	0.0144	0.0144
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: This table presents the effect of FTO force rate on recruit arrests. Standard errors are clustered at the recruit and FTO level. Columns 2 and 5 add controls for recruit characteristics (age, gender, race, education, prior law enforcement experience, and training). We add call characteristics fixed effects (number of officers on the scene, beat, type of call—priority-by-type,year-by-month, and day of the week-by night) in columns 3 and 6. Unfiled arrests are arrests that are not filed with the Dallas District Attorney's Office.

Appendix Tables and Figures

Figure A.1: Recruit and FTO Training Timeline

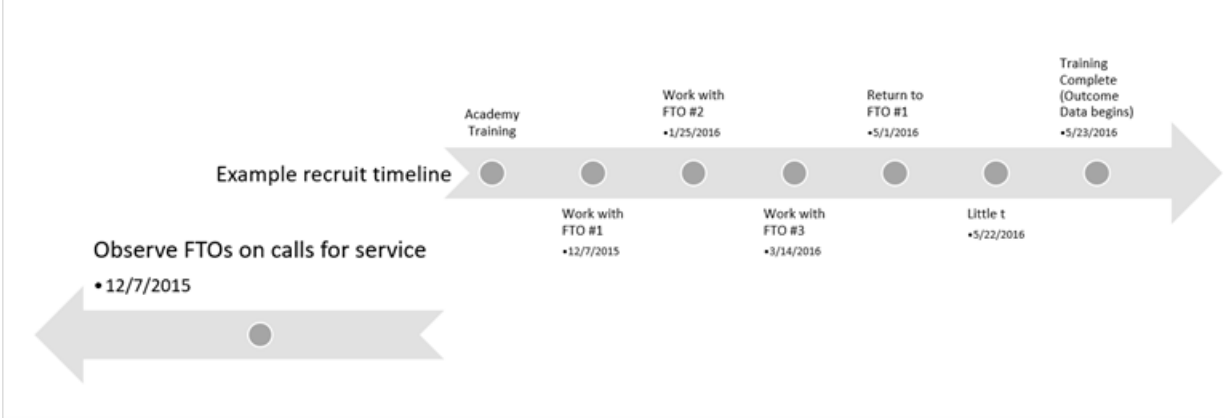
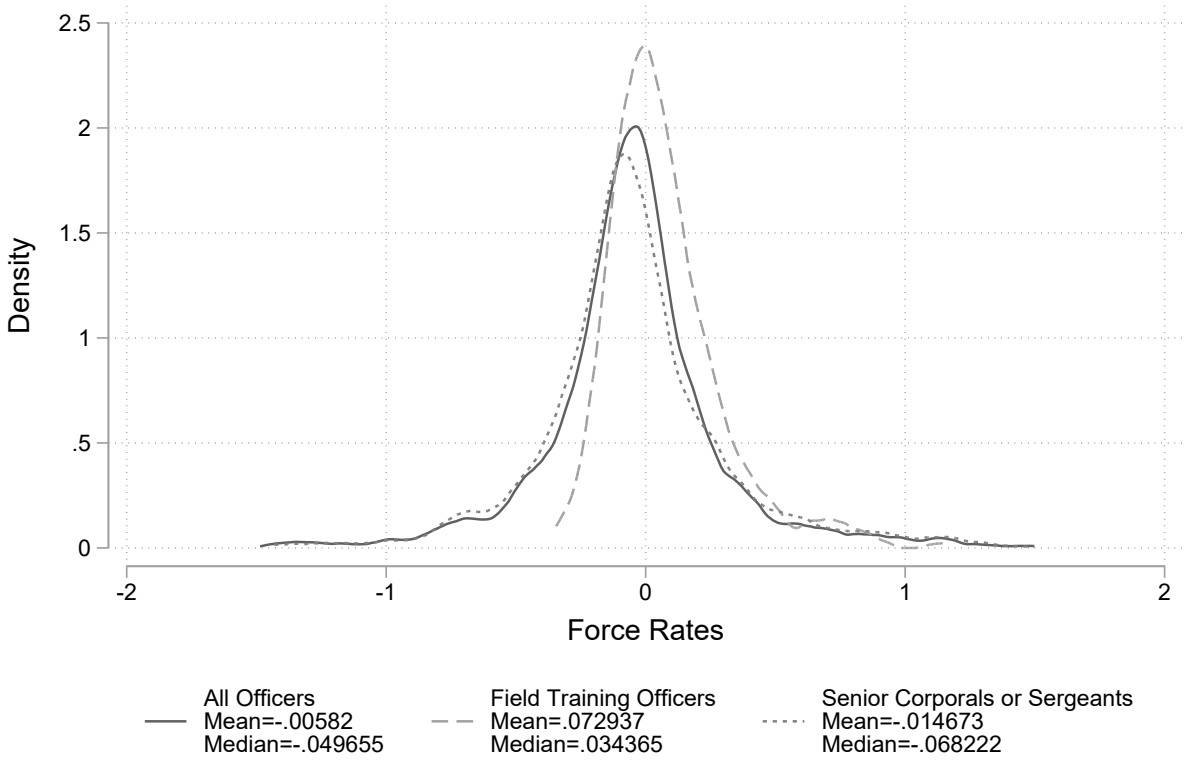
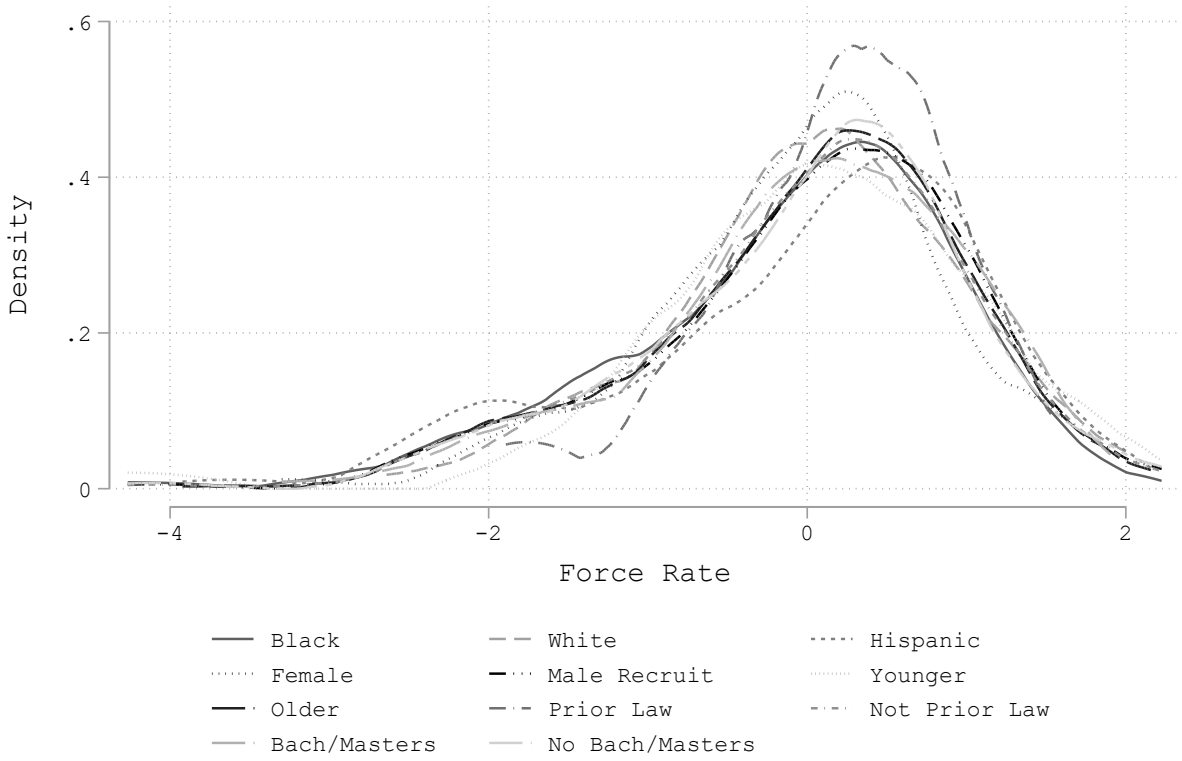


Figure A.2: Density of Officer Propensity to Use Force for All Officers



Notes: The figure shows density plots comparing force rates among three groups: All Officers, Field Training Officers, and Senior Corporals or Sergeants. Field Training Officers demonstrate slightly higher force rates compared to the overall average, while Senior Corporals and Sergeants show lower rates. Senior Corporals and Sergeants represent the first two possible promotion ranks and are the most common rank for Field Training Officers. The distributions are trimmed to exclude the top 3 percent of extreme force users for better visualization. Force rates are centered near zero across all groups, showing that most officers use similar levels of force in comparable situations

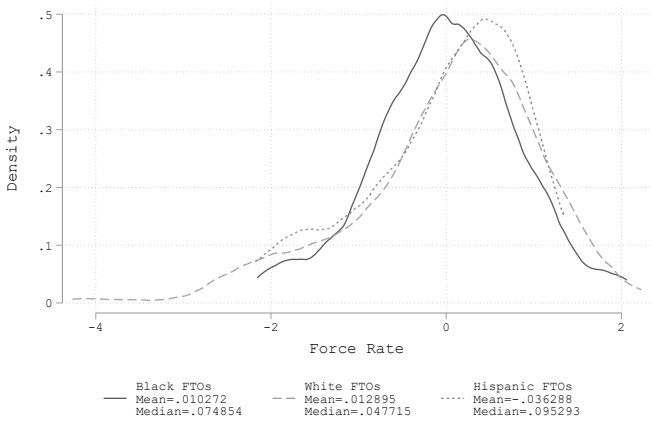
Figure A.3: Density of Field Training Officer Propensity to Use Force by Recruit Characteristics



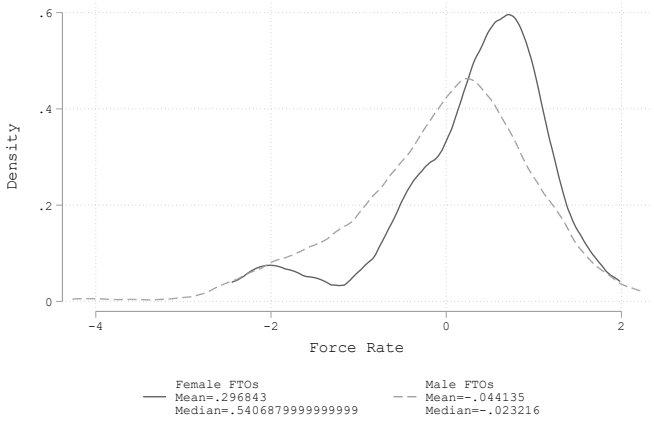
Notes: This figure presents the density of FTO force rates, grouped by recruit demographics, including race (Black, White, Hispanic), gender (Male, Female), age (Older, Younger), prior law enforcement experience, and educational attainment (Bachelor's/Master's vs. No Bachelor's/Master's). Older recruits are classified as those above the sample mean age. The figure illustrates the variation in FTOs' historical use of force propensity across different recruit groups, providing insight into whether recruits with different characteristics are systematically exposed to higher or lower-force FTOs.

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

(a) Field Training Officer Race

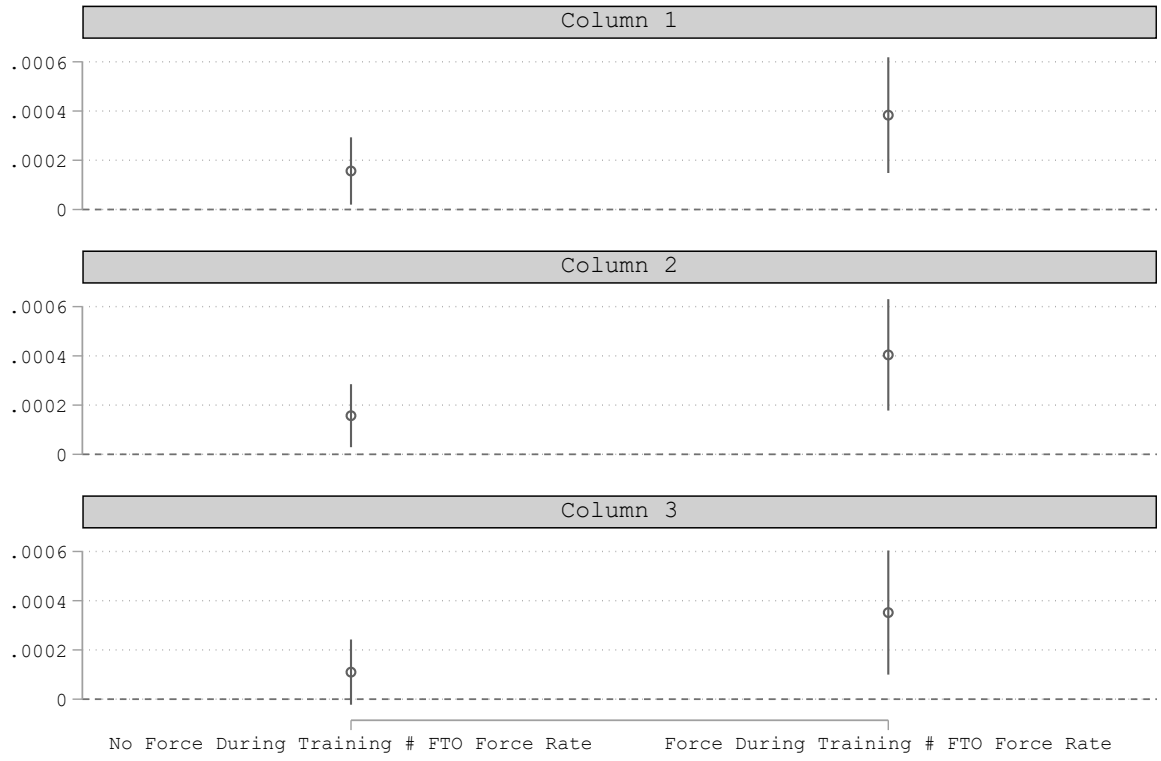


(b) Field Training Officer Gender



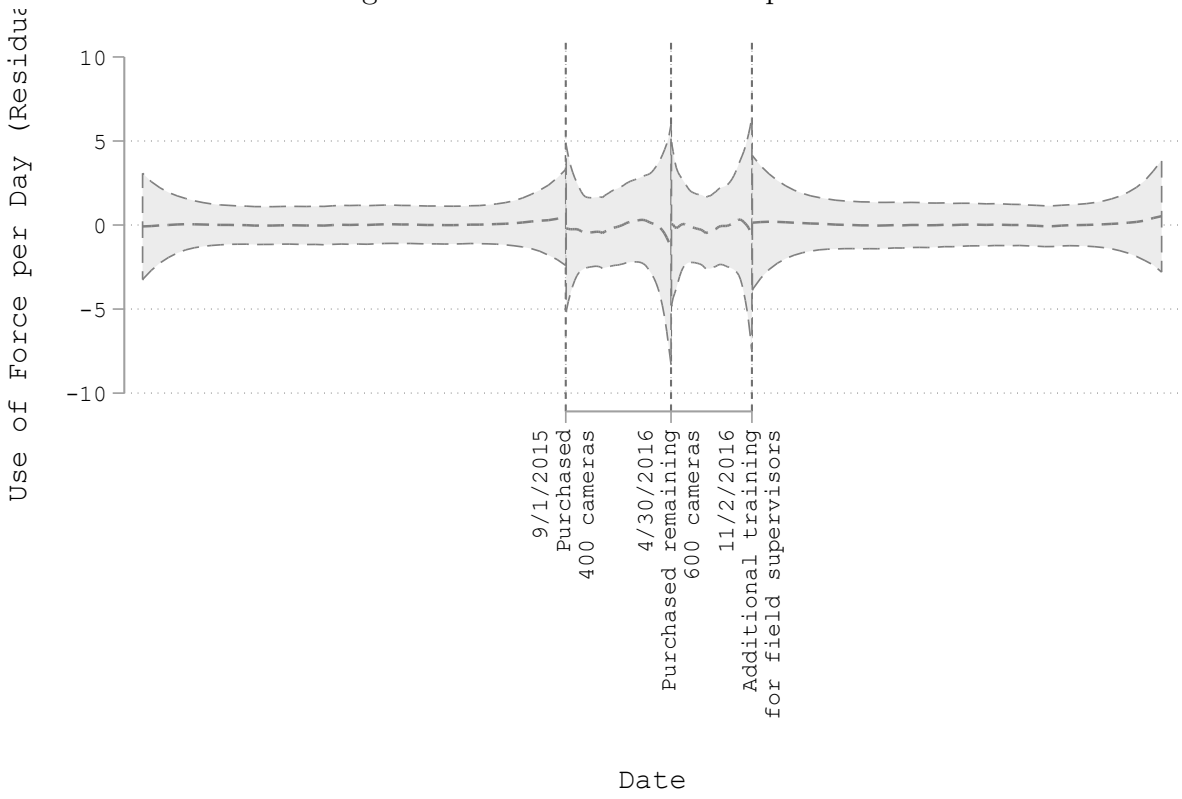
(c) Field Training Officer Age

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



This figure presents estimates of the effect of field training officer (FTO) force rates on recruit use of force, separately for recruits who did and did not experience force incidents during their training period. The plotted points represent coefficient estimates of β_1 from our primary specification $\text{Force}_{r,c} = \theta_r + \beta_1 \Lambda_{o(r)} + \beta_2 X_c + \epsilon_{r,c}$, where β_1 captures the effect of a one standard deviation increase in FTO force rate. Vertical lines show confidence intervals with standard errors multiway clustered at the FTO and recruit levels. Column 1 shows the baseline specification with recruit fixed effects. Column 2 adds controls for recruit characteristics (age, gender, race, prior law enforcement experience, education, training). Column 3 further includes call characteristic fixed effects (number of officers, beat, call priority-by-type, and temporal controls).

Figure A.6: Police Camera Adoption



Notes:This figure plots the daily average use of force over time, after removing year-by-month fixed effects. The solid line represents a local linear polynomial fit, with the gray shaded area showing 95 percent confidence intervals. Vertical dashed lines indicate three key dates in the body-worn camera program implementation: the initial purchase of 400 cameras (September 1, 2015), the subsequent purchase of 600 additional cameras (April 30, 2016), and the completion of supplementary training for field supervisors (November 2, 2016). The y-axis shows the residualized number of force incidents per day, controlling for seasonal patterns through year-by-month fixed effects.

Table A.1: **Robustness to Different Force Measures**: The Effect of Field Training Officer Force Rate on Recruit Use of Force

	(1) Force	(2) Force	(3) Force	(4) Force	(5) Force	(6) Force	(7) Force	(8) Force	(9) Force	(10) Force	(11) Force	(12) Force
Theta Shrunk	0.141***	0.139***	0.107***									
	(0.0450)	(0.0402)	(0.0412)									
Inverse Hyperbolic Sine				0.141***	0.139***	0.107***						
				(0.0450)	(0.0402)	(0.0412)						
Unshrunken Force Rate							0.122***	0.121***	0.0941***			
							(0.0380)	(0.0338)	(0.0343)			
Theta Shrunk No Call Controls										0.197***	0.199***	0.158**
										(0.0726)	(0.0726)	(0.0796)
Observations	1198564	1198564	1198564	1198564	1198564	1198564	1198564	1198564	1198564	1198564	1198564	1198564
Outcome Mean	0.00116	0.00116	0.00116	0.00116	0.00116	0.00116	0.00116	0.00116	0.00116	0.00116	0.00116	0.00116
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
* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: This table presents estimates of field training officer (FTO) effects on recruit use of force using three different measures of FTO force rates. Our primary regression specification is $\text{Force}_{r,c} = \theta_r + \beta_1 \Lambda_{o(r)} + \beta_2 X_c + \epsilon_{r,c}$. Columns (1)–(3) use the empirical Bayes-shrunken force measure, columns (4)–(6) apply its inverse hyperbolic sine transformation, and columns (7)–(9) use the raw, unshrunken force rates. Columns (10)–(12) estimate effects using our primary measure while excluding call controls. For each measure, the baseline specification includes assigned division-by-cohort fixed effects. The second specification (columns 2, 5, 8, and 11) adds recruit characteristics, including age, gender, race, prior law enforcement experience, education, and training. The third specification (columns 3, 6, 9, and 12) further incorporates call characteristic fixed effects, such as the number of officers, beat, call priority-by-type, and temporal controls. Standard errors, shown in parentheses, are multiway clustered at the FTO and recruit levels.

Table A.2: Recruit Level Summary Statistics

	(1) Entire Sample	(2) High Force Trainer	(3) Low Force Trainer
White	0.453	0.440	0.469
	(0.498)	(0.497)	(0.500)
Black	0.219	0.215	0.224
	(0.414)	(0.412)	(0.418)
Hispanic	0.278	0.287	0.268
	(0.449)	(0.453)	(0.444)
Female	0.182	0.181	0.184
	(0.386)	(0.386)	(0.389)
Age	32.01	32.13	31.85
	(5.147)	(5.207)	(5.076)
Prior Law Enforcement	0.0787	0.0887	0.0658
	(0.270)	(0.285)	(0.248)
Bachelors	0.351	0.341	0.364
	(0.478)	(0.475)	(0.482)
Masters	0.0250	0.0273	0.0219
	(0.156)	(0.163)	(0.147)
Defensive Training	0.0211	0.0273	0.0132
	(0.144)	(0.163)	(0.114)
Crisis Intervention Training	0.0192	0.0137	0.0263
	(0.137)	(0.116)	(0.160)
Observations	521	293	228

Notes: This table presents recruit-level summary statistics (means with standard deviations) comparing demographics and training across the sample and by FTO force usage. High Force FTOs show above-average force rates, while Low Force FTOs exhibit below-average rates compared to departmental means. Defensive Training and Crisis Intervention Training occur before field training begins. Age reflects recruit age when training commenced.

Table A.3: **Robust to Alternative Standard Errors (Robust SEs)** Balance Test: Correlation between Recruit and Field Training Officer Characteristics

Recruit Chars	Age	Female	Black	Hispanic	Hire Date	FTO Force	Recruit Force	Recruit Force NC
Female	-0.4661	0.0137	0.0620	0.0342	0.3789	0.0000	-0.0000	-0.0000
	(1.2947)	(0.0460)	(0.0525)	(0.0527)	(1.2128)	(0.0001)	(0.0001)	(0.0001)
Black	-0.5571	0.0634	0.0867	-0.0770*	0.1093	-0.0001	-0.0001	-0.0002**
	(1.3818)	(0.0459)	(0.0590)	(0.0458)	(1.1653)	(0.0001)	(0.0001)	(0.0001)
Hispanic	0.5837	0.0544	-0.0140	0.0328	-0.7922	-0.0000	-0.0000	-0.0001
	(1.1267)	(0.0449)	(0.0498)	(0.0448)	(0.9934)	(0.0001)	(0.0001)	(0.0001)
Prior Law	0.7052	-0.0278	-0.0892	-0.0915	-0.7689	0.0002*	0.0003**	0.0003**
	(2.2115)	(0.0818)	(0.0929)	(0.0636)	(1.8971)	(0.0001)	(0.0001)	(0.0001)
Bachelors	-1.4008	0.0213	0.0257	0.0131	0.6346	-0.0000	-0.0001	-0.0001
	(1.0440)	(0.0375)	(0.0463)	(0.0398)	(0.9034)	(0.0001)	(0.0001)	(0.0001)
Masters	-4.0827	-0.0469	0.0536	0.0912	2.5177	-0.0001	0.0001	0.0000
	(2.7644)	(0.1077)	(0.1233)	(0.1408)	(3.1270)	(0.0002)	(0.0002)	(0.0002)
Descalation Training	1.5722	-0.0430	-0.1082	0.2807*	2.5551	-0.0003	-0.0006**	-0.0005**
	(4.5776)	(0.1522)	(0.1280)	(0.1521)	(4.2585)	(0.0002)	(0.0003)	(0.0003)
Crisis Intervention Training	0.6236	0.1982	0.1823	-0.1812*	-1.6883	0.0000	-0.0003	-0.0004
	(3.7082)	(0.1843)	(0.1251)	(0.1071)	(4.0911)	(0.0002)	(0.0003)	(0.0003)
Age 30s	-0.1551	0.0361	0.0117	-0.0192	0.9462	-0.0000	0.0000	-0.0001
	(1.1927)	(0.0430)	(0.0477)	(0.0456)	(1.0875)	(0.0001)	(0.0001)	(0.0001)
Age 40s	1.0537	0.0870	0.1627	0.1003	-0.1839	0.0002	-0.0001	-0.0002**
	(3.3371)	(0.1265)	(0.1699)	(0.1547)	(2.8227)	(0.0002)	(0.0001)	(0.0001)
Div-x-Cohort FE	X	X	X	X	X	X	X	X
Observations	521	521	521	521	521	521	521	521
Outcome Mean	42.1382	0.1248	0.1843	0.1459	41.6620	0.0001	-0.0000	0.0001
F-Test P-Value	0.866	0.819	0.318	0.157	0.967	0.201	0.438	0.031
F-Test P-Values for Alternative Specifications								
Pairwise Omnibus Test					0.7880			

Notes: Columns report coefficients from regressing each FTO characteristic on all recruit characteristics (with robust SEs and division-by-cohort fixed effects–i.e, replicating Table 2 with robust SEs). Column F-tests assess joint significance. The pairwise omnibus test combines all 60 FTO-recruit characteristic pairs in a stacked regression with pair and division-by-cohort-by-pair fixed effects. The outcome "Recruit Force NC" is the recruit force rate excluding call controls.

Table A.4: Balance Test for All Three Field Training Officers: Correlation between Recruit and Field Training Officer Characteristics

Recruit Chars	Age	Female	Black	Hispanic	Hire Date	FTO Force	Recruit Force	Recruit Force NC
Female	-0.0998	-0.0014	0.0198	0.0087	0.0183	-0.0000	-0.0000	-0.0000
	(0.4861)	(0.0199)	(0.0211)	(0.0209)	(0.4395)	(0.0000)	(0.0001)	(0.0001)
Black	0.2741	-0.0052	0.0265	-0.0222	-0.1188	0.0000	-0.0001	-0.0002***
	(0.5838)	(0.0198)	(0.0265)	(0.0191)	(0.5118)	(0.0000)	(0.0001)	(0.0001)
Hispanic	0.0656	-0.0050	-0.0198	0.0149	-0.3013	0.0000	-0.0000	-0.0001
	(0.5177)	(0.0157)	(0.0226)	(0.0207)	(0.4250)	(0.0000)	(0.0000)	(0.0000)
Prior Law	0.7050	0.0096	-0.0102	-0.0093	-1.1754	0.0001	0.0003***	0.0003***
	(0.9183)	(0.0397)	(0.0428)	(0.0368)	(0.8198)	(0.0001)	(0.0001)	(0.0001)
Bachelors	0.0483	0.0067	-0.0039	-0.0025	-0.1650	0.0000	-0.0001**	-0.0001*
	(0.4024)	(0.0135)	(0.0166)	(0.0160)	(0.3640)	(0.0000)	(0.0000)	(0.0000)
Masters	0.6460	-0.0003	-0.0276	0.0762	-1.4769	-0.0000	0.0000	-0.0000
	(1.3585)	(0.0473)	(0.0504)	(0.0476)	(1.3864)	(0.0001)	(0.0001)	(0.0001)
Descalation Training	0.1594	0.0037	-0.0960	0.0438	1.2192	-0.0001	-0.0006***	-0.0005***
	(1.7804)	(0.0734)	(0.0743)	(0.0717)	(1.8310)	(0.0001)	(0.0002)	(0.0002)
Crisis Intervention Training	-0.5844	0.0638	0.0432	0.0052	1.2164	0.0001	-0.0003	-0.0004*
	(1.3399)	(0.0675)	(0.0558)	(0.0761)	(1.3781)	(0.0001)	(0.0002)	(0.0002)
Age 30s	-0.2075	0.0136	-0.0044	0.0045	0.2211	-0.0000	-0.0000	-0.0001
	(0.4978)	(0.0183)	(0.0231)	(0.0249)	(0.4388)	(0.0000)	(0.0000)	(0.0000)
Age 40s	-2.6722**	-0.0134	0.1467**	-0.0622	2.0179	0.0000	-0.0001	-0.0003***
	(1.3449)	(0.0647)	(0.0622)	(0.0695)	(1.2409)	(0.0001)	(0.0001)	(0.0001)
Div-x-Cohort FE	X	X	X	X	X	X	X	X
Observations	1558	1558	1558	1558	1558	1558	1558	1558
Outcome Mean	42.0334	0.1316	0.1849	0.1521	41.8202	0.0001	-0.0000	0.0001
F-Test P-Value	0.849	0.993	0.380	0.750	0.784	0.749	0.000	0.000

Notes: Columns report coefficients from regressing each FTO characteristic on all recruit characteristics using FTO-phase SRs and division-by-cohort fixed effects (i.e., replicating Table 2 for all three FTOs). Column F-tests assess joint significance. The outcome "Recruit Force NC" is the recruit force rate excluding call controls.

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

	(1) Arrest	(2) Arrest	(3) Arrest
FTO Force Rate	0.000594	0.000647	0.000342
	(0.000553)	(0.000551)	(0.000444)
Observations	1198564	1198564	1198564
Outcome Mean	0.0363	0.0363	0.0363
Assigned Div by Cohort FE	Y	Y	Y
Recruit Characteristics	-	Y	Y
Call Controls	-	-	Y

Standard errors in parentheses

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

Notes: This table presents the effect of FTO force rate on recruit arrests. Our primary regression specification is given by $\text{Arrest}_{r,c} = \theta_r + \beta_1 \Lambda_{o(r)} + \beta_2 X_c + \epsilon_{r,c}$ (Equation 3), where β_1 represents the effect of a one standard deviation increase in FTO force rate. Standard errors are multiway clustered at the field training officer and recruit levels. Column 2 includes controls for recruit characteristics (age, gender, race, prior law enforcement experience, education, and training), while Column 3 further adds call characteristic fixed effects (number of officers on the scene, beat, call type—priority-by-type, year-by-month, and day-of-week-by-night interactions).

Table A.6: **Robustness Attrition**: Correlation between FTO, Recruit Characteristics and Attrition

	Coefficient on Leaver	Standard Error	Outcome Mean
FTO Characteristics			
Age	0.1578	(1.2622)	46.7447
Female	0.0120	(0.0524)	0.1248
Black	0.0207	(0.0534)	0.1843
Hispanic	0.0129	(0.0401)	0.1459
Hire Date	350.1441	(435.2672)	1.52e+04
Force Rate	-0.0573	(0.1267)	-0.0000
Recruit Characteristics			
Female	0.0305	(0.0509)	0.1823
Black	-0.0332	(0.0531)	0.2188
Hispanic	-0.0114	(0.0639)	0.2783
Prior Law Enforcement	-0.0626*	(0.0356)	0.0787
Bachelors	-0.0216	(0.0658)	0.3512
Masters	0.0282	(0.0302)	0.0250
Defensive Training	-0.0225	(0.0186)	0.0211
Crisis Intervention Training	-0.0063	(0.0177)	0.0192
Age 30s	-0.0833	(0.0577)	0.2169
Age 40s	0.0024	(0.0262)	0.0326
Observations	521		
Div-x-Cohort FE	X		

Standard errors in parentheses clustered at FTO level

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

Notes: This table examines whether recruits who leave the department after training differ from those who remain in terms of their assigned FTOs' characteristics or their own personal attributes. Each row presents results from a separate regression of the characteristic on an indicator for leaving, controlling for division-by-cohort fixed effects. Standard errors are clustered at the FTO level. The table reports coefficients, standard errors, and outcome means for each characteristic.

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

	Two Years After Training			**One Year After Training**		
	(1)	(2)	(3)	(4)	(5)	(6)
	Force	Force	Force	Force	Force	Force
FTO Force Rate	0.000258***	0.000245***	0.000202***	0.000250***	0.000243***	0.000216***
	(0.0000620)	(0.0000556)	(0.0000576)	(0.0000643)	(0.0000614)	(0.0000629)
Observations	895551	895551	895551	553394	553394	553394
Outcome Mean	0.00119	0.00119	0.00119	0.00118	0.00118	0.00118
Div by Cohort FE	Y	Y	Y	Y	Y	Y
Recruit Chars	-	Y	Y	-	Y	Y
Call Controls	-	-	Y	-	-	Y

Notes: This table presents the effect of Field Training Officer (FTO) force rate on recruit use of force. Each column shows regression results with Force as the dependent variable. Columns 1-3 analyze calls for service within two years after training completion, while columns 4-6 examine calls within one year. Models progressively add controls: all specifications include division by cohort fixed effects, columns 2-3 and 5-6 add recruit characteristics, and columns 3 and 6 include additional call-level controls. The coefficients represent the effect of a one-unit increase in FTO force rate on recruit force probability. Thirty-five recruits leave the department before completing one year of service, and sixty-seven leave before completing two years.

Table A.8: **Robustness Attrition**: The Effect of High Force Field Training Officers on Recruit Use of Force by Retention Status

	(1) Force	(2) Force	(3) Force
FTO Force Rate	0.000232***	0.000222***	0.000176***
	(0.0000626)	(0.0000602)	(0.0000635)
Leaver	-0.0000912	-0.000133	-0.000177
	(0.000133)	(0.000125)	(0.000135)
Force Rate*Leaver	-0.000145	-0.000112	-0.000118
	(0.000153)	(0.000134)	(0.000147)
Observations	1198564	1198564	1198564
Outcome Mean	0.00116	0.00116	0.00116
Assigned Div by Cohort FE	Y	Y	Y
Recruit Characteristics	-	Y	Y
Call Controls	-	-	Y

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

Notes: This table presents the effect of FTO force rate on recruit use of force, comparing recruits who remain with the department ("stayers") versus those who leave ("leavers"). The interaction term (Force Rate*Leaver) tests whether FTO influence differs for officers who eventually quit. Column 1 presents the baseline specification with assigned division by cohort fixed effects. Column 2 adds controls for recruit characteristics (age, gender, race, prior law enforcement experience, education, and training course participation). Column 3 further incorporates call characteristics fixed effects (number of officers on scene, beat, call priority-by-type, year-by-month, and day of week-by-night). Standard errors in parentheses

Table A.9: The Effect of High Force Field Training Officers for All Field Training Officers

	(1) Force	(2) Force	(3) Force
FTO 1 Force Race	0.000199***	0.000206***	0.000148**
	(0.0000680)	(0.0000662)	(0.0000696)
FTO 2 Force Race	-0.0000491	-0.0000106	-0.0000448
	(0.0000701)	(0.0000695)	(0.0000729)
FTO 3 Force Race	0.0000307	0.0000207	0.0000192
	(0.0000642)	(0.0000616)	(0.0000645)
Observations	1194521	1194521	1194521
Outcome Mean	0.00116	0.00116	0.00116
Assigned Div by Cohort FE	Y	Y	Y
Recruit Characteristics	-	Y	Y
Call Controls	-	-	Y

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

Notes: This table presents the effect of individual field training officers (FTOs) on recruit use of force, examining the distinct influence of each of the three FTOs in a recruit's training rotation. Our specification is $\text{Force}_{r,c} = \theta_r + \beta_1 \Lambda_{o_1(r)} + \beta_2 \Lambda_{o_2(r)} + \beta_3 \Lambda_{o_3(r)} + \beta_4 X_c + \epsilon$, where the coefficients represent the effect of a one standard deviation increase in each FTO's force rate on subsequent recruit behavior. Column 1 presents the baseline specification with assigned division by cohort fixed effects. Column 2 adds controls for recruit characteristics (age, gender, race, prior law enforcement experience, education, and training). Column 3 incorporates call characteristic fixed effects (number of officers on scene, beat, call priority-by-type, year-by-month, and day of week-by-night). Note that the sample size is slightly smaller than other analyses because we are missing FTOs for some recruits.

Appendix B: Other Field Training Officer Rates

To better understand the mechanism behind our results, and to rule out explanations such as reporting, we calculate field training officer propensity to make arrests, respond to calls in a timely manner, and write up informative reports in a manner similar to our force rate calculations. Namely, we estimate Equation 1 using arrest, misdemeanor arrests, felony arrests, filed arrests, unfiled arrests, response time, and time spent on a call as our outcome. We then shrink our FTO-recruit pair estimates of $\lambda_{o(r)}$ according to Equation 2. To address whether our results are driven by officer reporting, we also estimate field training officers 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 507 of our 514 field training officer-recruit pairs. We attempt to estimate our $\lambda_{o(r)}$'s in a very similar manner.

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

The results of these calculations are show in B.1. Figures B.1a , B.1b, B.1c, B.1d, B.1e show the distribution for our unshrunken and shrunken measures for a field training officer's propensity to make different types of arrests. Both distributions have a longer right tail, indicating that there are some officers with much higher arrest rates than the average field training officer. Further, there is substantial variation in our arrests rates. A one standard deviation increase in officer effects corresponds to a 32 percent (0.012/.037)and 33 percent (0.007/ 0.0215) increase in arrest or misdemeanor arrest rates.

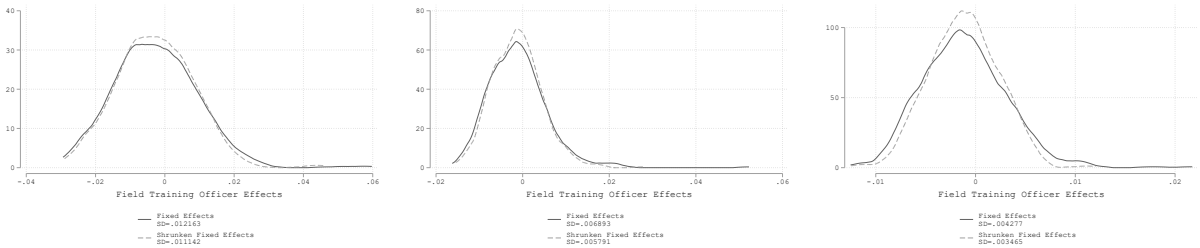
Figures B.1c and B.1d show results for our measures of time use (measured in hours). A one standard deviation increase in response time is 1 minute (0.02 hours) or 17 percent (.02/.12 hours) increase. A one standard deviation increase in time on a call is 0.359 hours or a 150 percent (.359/.24) increase.

Next, we consider how many words an officer uses when writing up an incident in Figures B.1e, B.1f, and B.1g . Unsurprisingly, the two distributions for number of words and characters look similar. The average number of characters used in an incident report is 43 and there are a few officers that are very wordy. A one standard deviation increase in number of characters used is an increase of 8.5 characters or 19 percent. 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 percent (1.4/7.31) increase. Finally, we consider an officers propensity to write nothing. On average 10 percent of incidents don't have a description. A one standard deviation increase in writing nothing corresponds to a 81 percent (0.08/0.099) increase in writing nothing. Together these figures show substantial variation in other officer behaviors.

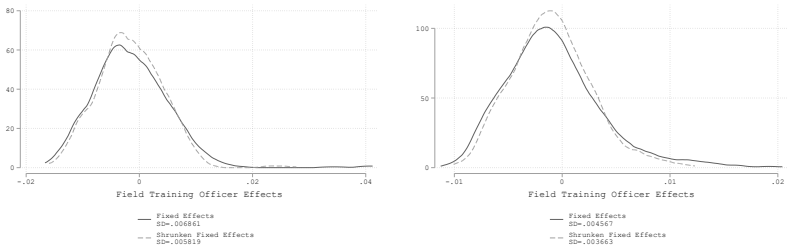
Finally, we consider officer complaints. Complaints are fairly common for FTOs in our sample; Over 80 percent of FTOs have at least one complaint. We also consider internal complaints (complaints filed by a DPD employee) and use of force complaints separately. FTOs receive a complaint about once for every 500 calls.

Figure B.1: Other Field Training Officer Rates

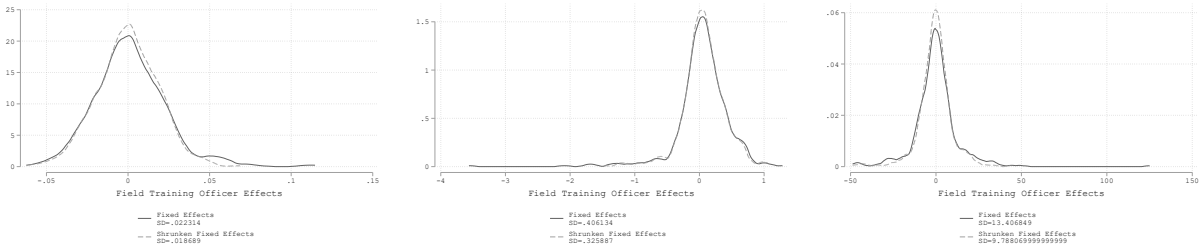
(a) Arrest (b) Misdemeanor Arrests (c) Felony Arrests



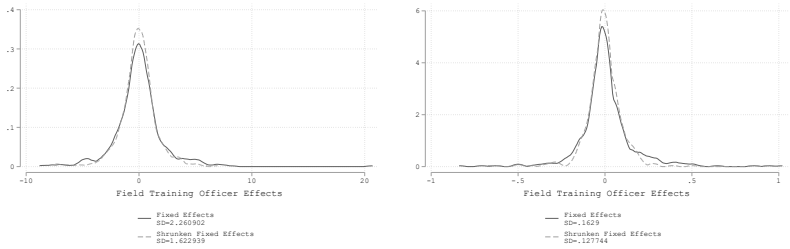
(d) Filed Arrests (e) Unfiled Arrests



(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, response time, time on a call and measures of wordiness. Response time is the number of hours between arrival time and assigned time. Time on call is the number of hours between the time an officer was enroute and when the call was cleared.