Do Peers Matter in the Police Academy?

Roman Rivera*
UC Berkeley

May 26, 2024
Version Date:

Abstract

Increasing underrepresented groups' representation in police departments is a common proposal to reduce aggressive policing. This paper documents the effects of peer composition in the Chicago police academy on officers' future arrests by exploiting the lottery system, which provides exogenous variation in cohort composition. I find that higher shares of peers from groups that police less aggressively, such as female and older officers, reduce all officers' future low-level arrests. Peer race matters by amplifying the effects of gender and age. Overall, the results are most consistent with peers' preferences for less aggressive policing shifting officers' preferences and changing future behavior.

*I am grateful to Sandra Black, Bentley MacLeod, and Simon Lee for guidance and advice. For feedback and comments, I thank Amani Abou Harb, Douglas Almond, Bocar Ba, Pat Bayer, Michael Best, Felipe Goncalves, Sakshi Gupta, Michelle Jiang, Jenny Jiao, Dean Knox, Taeho Kim, Bob LaLonde, Claire Montialoux, Jonathan Mummolo, Brendan O'Flaherty, José Luis Montiel Olea, Nayoung Rim, Evan Rose, Rajiv Sethi, Miguel Urquiola, and Emily Weisburst, as well as the colloquium participants at Columbia University and the Texas Economics of Crime Workshop. I would like to thank Mohammad Abou Harb, Emma Herman, the Invisible Institute, and Sam Stecklow for their contributions to this data set and Rachel Ryley for sharing assignment data. For detailed explanation of the Cook County court system, I thank Ali Ammoura. This paper supersedes a prior version circulated as "The Effect of Minority Peers on Future Arrest Quantity and Quality". See Rivera (2024) for replication data and code. IRB approval for this study and data collection was received from Columbia University (#AAAT6426). Email: rgrivera@berkeley.edu

1 Introduction

Aggressive policing, such as the excessive use of force and over-policing of low-level crimes, has numerous social and economic costs and disproportionately affects minority communities. Increased racial and gender diversity in policing is one of the most common policy proposals to address aggressive policing, as well as to build community trust and improve police legitimacy (DOJ (2016)). These policies address the disproportionately white and male composition of police departments relative to the communities they police. Existing research finds that minority and female officers police less aggressively (Ba, Knox, et al. (2021), Goncalves and Mello (2021), Hoekstra and Sloan (2022)). However, changing the composition of new officers may also result in spillover effects onto their fellow officers. For example, officers may adopt their peers' preferences for policing behavior, such as aggressive policing, and hiring more minority officers could lead to more interracial friendships and reduce racial bias.

In this paper, I provide novel evidence of the influence of peer composition in the police academy on officers' future arrest outcomes. Such evidence has previously proved elusive for two main reasons. First, it requires peer groups to be randomly assigned to avoid self-selection. Second, it requires data on individual police officer demographics, arrests, and peer groups, which are difficult to obtain. I overcome both obstacles using detailed data on police officers in the Chicago Police Department (CPD) who were randomly assigned to peer groups in the form of police academy cohorts based on lottery numbers. The police academy is a highly policy-relevant environment: increased departmental diversity requires training increasingly diverse cohorts; the academy forms recruits' first major experiences and relationships in policing; and the CPD academy involves training with one's cohort for six months (900 hours), making the composition of one's peer group a relatively intense intervention. While the CPD is a single department, it provides a particularly useful setting because it is a sufficiently large department in a diverse city, with a long-standing lottery system for the academy, and maintains high-quality data.

I first document that an officer's randomly assigned police academy cohort influences their future arrest outcomes. After removing the influence of temporal and individual officer characteristics and correcting for sampling error, cohort-specific effects account for a significant share of the variation in future officer arrests. A 1 standard deviation (SD) increase in cohort effects increases officers' future low-level arrests by 2.6 arrests per 100 shifts per officer. This is equivalent to around 20% of the mean of low level arrests and similar in magnitude to the difference between effect of deploying a Black officer compared with a white officer on arrests in Ba, Knox, et al. (2021). However, cohort effects generally explain little, if any, variation in officers' future serious arrests.

Given the influence of cohorts on future low-level arrests, I study the role of an officer's cohort's composition in explaining differences in future arrest outcomes. I find that higher shares of female and older peers, and to a lesser extent Black peers, have economically significant and negative effects on future low-level arrests. Peer age proves to be a significant factor with the largest effects produced when distinguishing between the shares of younger (less than 27 years old, around the 35th percentile) and older peers. In the main specification, a 5pp increase in the share of peers older than 27 or in the share of female peers both reduce future low-level arrests by 0.7 per 100 shifts, around 6% of the mean. While controlling for peer age and gender, the effect of Black peers is not economically or statistically significant.

I explore two potential channels through which cohort composition could influence future arrests: officer behavior and their assignments. Using data on millions of daily shifts and assignments, I model officers' arrests as the sum of the influence of their behavior, as a fixed individual propensity to make arrests, and the influence of their assignment (e.g., where, when, and in what role they work). While the relationship between peer composition and an officer's future assignments is inconsistent with the arrest results, I find that the relationship between peer composition and officers' arrest propensities is consistent with peers' effects on average arrests. This suggests that peers in the academy influence arrests through changes in officer's future behavior.

Understanding the mechanisms driving these effects is crucial for developing appropriate policy. I propose and test a variety of mechanisms through which cohort composition could influence officers' future behavior, including by influencing selection into trainers or future co-workers. However, I find the most robust evidence for a preference-spillovers mechanism, wherein peers influence each others' future behavior through forming group cultures, influencing social identity, and altering each others' preferences and personalities (Akerlof and Kranton (2000), Fryer and Torelli (2010), Anwar, Bayer, and Hjalmarsson (2019), Golsteyn, Non, and Zölitz (2021)).

At the officer level, female and Black officers, and to a lesser extent Hispanic and other minority officers, tend to make fewer low-level arrests and have lower arrest propensities than their male and white counterparts, consistent with prior research. Furthermore, officer age significantly influences how aggressively officers, particularly male officers across all race groups, police low-level crimes. For example, though white officers police more aggressively than minorities, I find that white male officers who enter the academy at 32 (the 75th percentile) or older make significantly fewer arrests for low-level offenses, around 5 fewer per 100 shifts which is around one-half a standard deviation, than their counterparts who enter the academy younger than 27 years old.

At the peer level, I find that officers exposed to higher shares of female and older peers of all racial groups make fewer low-level arrests, with the effects of older Black peers, followed by older non-Black minority peers, being the largest. Under the interpretation of officer-level differences as differences in preferences, these patterns are highly consistent with preference spillovers, as having higher shares of low-aggression peers results in less aggressive policing in the future. Among other tests, I use officer-level characteristics to predict the average arrests and arrest propensities of officers and their peers. Consistent with having peers who individually prefer less aggressive policing causing the officer to prefer less aggressive policing themselves, peers' predicted arrests are strongly related to future arrests for low-level offenses while I find no such effect for serious arrests. I also find some evidence consistent with a smaller secondary effect of interracial socialization beyond preference effects as white officers are more strongly influenced by minority peers.

Notably, the reductions in low-level arrests are matched with smaller but less robust increases in serious arrests (e.g., property crime), suggesting officers are not reducing effort overall. While I find mixed effects on arrest quality as measured by court outcomes, the decline in low-level arrests and minor positive effects on serious arrests are unlikely to be welfare-reducing in the context of mounting evidence that arrests and prosecution of low-level crimes can significantly harm individuals and lead to future criminality (Agan, Doleac, and Harvey (2023)) and that reductions in low-level arrests do not necessarily lead to increases in crime (Cho, Goncalves, and Weisburst (2021)).[1] Furthermore, these effects on low-level arrests endure even four years after the academy has ended, in contrast with more short-lived effects of police trainings (Owens et al. (2018)) and peer interventions in other environments (Dahl, Kotsadam, and Rooth (2021)).

While the results leverage variation in composition across cohorts, the most relevant policy implications apply to altering the entire pool of recruits if peers influence each others' future behavior, as opposed to the optimal assignment of officers to cohorts. As the results are most consistent with behavioral changes, the main implications of these findings are that increased racial and gender diversity can have positive spillovers through peer effects in the police academy, leading to less aggressive policing of low-level crimes without reducing effort toward serious offenses. However, policies that aim to reduce aggressive policing through increased diversity should take additional factors into account, such as the age of recruits, as the enforcement preferences of officers not only have implications for how they individually police but also for their spillover effects onto their peers. This reinforces the importance of research on better recruiting public employees (Linos (2018)).

[1] See also Aizer and Doyle (2015), Gupta, Hansman, and Frenchman (2016), Stevenson (2018), and Dobbie, Goldin, and Yang (2018) for work on courts and pretrial detention. Weisburst (2022) also uses court outcomes as a metric for arrest quality, and Ater, Givati, and Rigbi (2014) discusses arrest quality using whether an arrest led to a charge as a quality measure.

This paper builds on the literature on diversity and policing. Prior research finds that minority and female representation influences city-level policing outcomes (McCrary (2007), Miller and Segal (2018)), and individual officer race and gender are associated with differential policing behavior (Goncalves and Mello (2021), Ba, Knox, et al. (2021), Hoekstra and Sloan (2022)).[2] This paper extends this research by identifying peer composition as a determinant of an officer's future arrest outcomes. In contrast with the general focus on officer race, I find that officers' starting age is a highly important factor in both individual officer behavior, consistent with Ridgeway (2020), and their effect on peers.

By providing evidence for peer preference spillovers as the primary mechanism, I also contribute to the literature on the effects of social identity on behavior.[3] Specifically, I contribute to the literature studying how peers influence outcomes through shifts in preferences and social identity (Akerlof and Kranton (2000), Akerlof and Kranton (2010), Austen-Smith and Fryer (2005), Benjamin, Choi, and Strickland (2010), Golsteyn, Non, and Zölitz (2021), Holden, Keane, and Lilley (2021)). For example, Anwar, Bayer, and Hjalmarsson (2019) find that jurors' political alignment influences trial outcomes through changing peer opinions. I provide multiple findings consistent with the effect of peer preferences on officer behavior, most centrally that exposure to peers from groups who police less aggressively causes both white and minority officers to police less aggressively in the future. This is consistent with the findings of Adger, Ross, and Sloan (2022), which shows that police officers use more force after being assigned to more aggressive training officers. I also find some evidence consistent with minority peers more strongly influencing white officers, consistent with prior work on peer diversity.[4]

This paper proceeds as follows. In Section 2, I describe background information, and I discuss data and summary statistics in Section 3. I discuss the empirical strategy in Section 4, and Section 5 presents the main results. Section 6 explores mechanisms and policy implications. Section 7 concludes.

[2] For the effect of representation on more macro-level outcomes, see Donohue and Levitt (2001), Garner, Harvey, and Johnson (2019), Harvey and Mattia (2024), and Cox, Cunningham, and Ortega (2021). For more on the relationship between officer race and policing outcomes, West (2018) and Hoekstra and Sloan (2022). See also Nicholson-Crotty, Nicholson-Crotty, and Fernandez (2017).

[3] I employ a common identification strategy (random assignment of students to classrooms) in the educational peer effects literature (Sacerdote (2011)) but in a new setting. See Hoxby (2000) and Sacerdote (2001). Angrist (2014) discusses various studies in the educational peer effects literature. Holz, Rivera, and Ba (2023) studies police academy cohorts as well, but their identification hinges on a difference-in-differences design, similar to Ager et al. (2021) which studies pilots. Gould, Lavy, and Paserman (2009), Lavy and Schlosser (2011), Black, Devereux, and Salvanes (2013), Carrell, Hoekstra, and Kuka (2018), and Brenoe and Zölitz (2020) provide evidence for peer composition in educational environments influencing future outcomes.

[4] For the literature on how minorities influence whites' biases and beliefs, see Boisjoly et al. (2006), Anwar, Bayer, and Hjalmarsson (2012), Carrell, Hoekstra, and West (2019), and Schindler and Westcott (2021).

2 Background on CPD and Recruitment

2.1 Application to CPD and the Academy

Comprised of over 10,000 officers, the Chicago Police Department (CPD) is the second largest police force in the US. It polices the nation's third largest city, which is racially diverse and economically segregated. To recruit new officers, the CPD issues a call for officers, and applicants take a written exam, which they must pass to enter the academy. As a CPD Frequently Asked Question (FAQ) form, CPD (2017), explains:

> All applicants who pass the exam are placed on an eligibility list based on a randomly assigned lottery number. You will be referred to the Chicago Police Department in lottery order as vacancies become available.

After an applicant's number is called and if they pass required physical and psychological tests, they can start at the police academy (see Appendix A.1 for more discussion). Academy start dates, known as "appointed dates," correspond to officers beginning their time at the police academy. I define a cohort as a group of officers with the same appointed date.

During the academy, officers must complete 900 hours (about six months) of training in multiple areas, such as "firearms, control tactics, physical training, [and] classroom training" (CPD (2020)). After the academy, the recruits in a cohort enter an on-the-job training period for one year as "probationary police officers" during which they are split up, work in multiple areas of the city, and are evaluated under the supervision of Field Training Officers. After recruits meet the various requirements, complete their time as a probationary officers, and become "field qualified" (CPD (2018)), they exit their probationary period and become full (sworn) Chicago police officers. New sworn officers are then assigned to more permanent units.[5]

2.2 Sample and Requirements

This paper focuses on the cohorts with start dates between 2009 and 2016. These cohorts can be divided into three periods based on the year officers took the entrance exam, i.e. the level at which they were randomly assigned lottery numbers. In 2006, three exams were given in rapid succession, each attended by a relatively small number of applicants (all between 800 and 1,500 passing applicants). I collectively refer to these exams as Exam 2006. The next test was issued in 2010 (Exam 2010), with almost 8,000 passing applicants. In 2013, the final

[5]The 900 hours of training encompasses and surpasses the training required to pass the Illinois State Peace Officer's Certification Exam. In larger cohorts, officers are further subdivided into "homerooms" which take most of their trainings together– while data on homerooms could not be obtained, I use detailed training data to approximate these groups as discussed in Section 6.2 and Appendix A.4.

exam used in the sample was issued (Exam 2013) with over 12,000 passing applicants and an important policy change: the minimum age of entry was reduced from 23 to 21 (Pritchard (2013)) — the maximum age is 40 years old for all cohorts in the sample.

The CPD is massively over-subscribed: fewer than 3,000 applicants were called into the academy between 2009 and 2016, while over 20,000 applicants passed the exams. This is because CPD jobs are highly desirable by individuals from across the country, and applicants to law enforcement are highly passionate about joining a police force. While the CPD did not provide information on which officers belong to which test, Appendix A.1 discussed matching cohorts with exams, and Figure B.1 provides additional exam information. In this sample, the CPD began to call individuals into the academy years after they took the exam: the last batch of Exam 2006 recruits were called in October 2011, and the first cohorts of Exam 2010 and Exam 2013 recruits started in April 2012 and August 2014, respectively. Based on internet discussions on a popular policing forum, passing applicants with high lottery numbers (far into the queue) were advised that the CPD would likely test again before they would be called (feredeathpsn (2017)).

2.3 Units and Daily Assignments

All CPD officers work within units. New officers are generally assigned to the patrol units which correspond to geographical districts in Chicago, totaling 25 districts/patrol units prior to 2012 and 22 starting in 2013. These units occupy most CPD officers and relate to what is commonly considered police work. Other units for more specialized work are not studied in this paper and contain far fewer and more experienced officers, such as training and detective units. A seniority-based bidding process determines transfers between units and is only available to non-probationary sworn officers, meaning new officers have little to no choice in where and when they work (CPD (2011)).

Within units, officers also bid for shifts/watches (generally four with three corresponding to eight hour time periods), their 'day off group', and furlough days at the end of the preceding year — this is also seniority based. On any specific day, whether or not an officer is assigned to work depends on their rotating schedule, which is generally four days on followed by two days off during the period of study, and is predetermined by their day off group based on the CPD operations calendar (see Figure B.2 and Ba, Knox, et al. (2021) for more details). The exact days an officer works are effectively predetermined, not up to the officer's discretion on that day, and rotate over the days of the week.[6]

[6]For example, in one week, an officer works Tuesday through Friday, and the next week they work Monday through Thursday. So, the exact composition of the officers working in a unit and watch on a specific day will be different the following week as officers of different day off groups and furlough schedules will be working

3 Data and Summary Statistics

3.1 Data

The data for this study come from the Chicago Police Department, Chicago's Department of Human Resources, and the Circuit Court of Cook County. Combining data sets on CPD officers, I construct a detailed panel data set of officer assignments, arrests, and arrest outcomes in court between 2010 and 2018 — see Rivera (2024) for replication data and code. This contains officers' demographic information (race, gender, age), start dates, when officers exited the training unit (after the academy and probationary period), and other administrative information. Daily assignment and attendance data include daily records on officer assignments and time on duty for the geographic units. Additional data sets contain information on trainings, officer education, military status, and language ability. Collectively, these data permit highly granular analysis of an officer's working environment and peer groups. I restrict my analysis to observations of police officers (the lowest and most common rank, e.g., not detectives, sergeants, etc.) working on shift (watch) numbers 1-4, and assigned to regular assignments (e.g., not administrative, lockup, desk duty, etc.).

To recover individual officer's arrest metrics, I use arrest data and court data. The arrest data contain all arrests of adults by Chicago police officers including arrest date and time, crime description, arrestee race and primary arresting officers (generally two). By linking the arrest data to court records, I determine whether or not an arrest had a guilty finding (including pleas) associated with it. For more discussion of the data, see Appendix C.

3.2 Sample Selection

A total of 2,745 officers joined the CPD between July 2009 and December 2016. As defined above, an academy cohort is all the recruits who started at the CPD academy on the same date, resulting in 68 cohorts during this period. All officers in the sample were subject to a series of filters.

Before calculating initial cohort composition, I excluded very small cohorts that started during the sample period but had cohorts with fewer than 7 recruits which removed a total of 37– the majority were in single officer cohorts. These small 'cohorts' are likely errors as the next smallest cohort size is 25. I also dropped 1 recruit who reported starting too young, and another 9 officers who likely had erroneous start dates were also removed.

Attrition from the initial cohorts to the final sample can occur for multiple reasons: I drop recruits in cohorts who were not matched in the assignment data, recruits with invalid

together. Officers do not frequently work in different watches throughout a year.

durations in the academy or probationary period, and recruits not matched in the salary and unit assignment data; I also drop a few recruits that had fewer than 15 observations in the assignment panel. If attrition is related to cohort composition, it may contaminate the results. But, as I show in Appendix A.2, cohort composition has no significant relationship with attrition. After filters, the analysis sample contains 2,296 officers in 42 cohorts with 1,041,166 total officer-shift observations over 93 months.

3.3 Cohort Composition

Table 1 displays the demographic composition of each exam period (Exam 2006, 2010, 2013) in Columns (1)-(6), with even columns containing pooled means and odd columns containing means over cohort compositions before attrition. Column (7) contains the pooled demographics of all officers in the panel data for reference. By comparing pooled and cohort columns, it is apparent that the samples of recruits are very similar to that of their average cohorts, which is expected under random assignment of recruits to cohorts, and that attrition after entering the academy did not significantly alter the demographic composition of the pool of officers. Overall, recruits are around 80% male, 50% white, with some change in the composition over time, while Exam 2006 recruits are around 23% Black and 26% Hispanic, Exam 2010 and Exam 2013 recruits are 13% Black and over 30% Hispanic, and Asian/Native Americans comprise around 4% of recruits across Exams. Notably, between Exam 2010 and Exam 2013, the reduction in start age requirement from 23 to 21 was associated with a two year decline in the average start age of recruits. As shown in Panel A of Figure B.4 in the Exam 2010 and Exam 2006 cohorts, only 26.6% of recruits started before they were 27, while in the Exam 2013 cohorts 49% did, and Panel B shows that while 45% of white officers started before 27, 31% of Black and 31% of non-Black minorities started before 27.

For a visualization of cohort compositions see Figure B.3, and Table B.1 contains summary statistics for the ranges of cohort compositions. There is relatively wide support across cohort compositions, with share minority ranging from 35% to 62%, share Black between 2% and 28%, share female between 11% and 38%, and average start ages from 26.7 years to 31 years. Importantly, characteristics such as age, race, and gender are correlated, as shown in Table B.2. For example, Black officers are more likely to be female and start at an older age, while white officers start younger and are more likely to be male.

3.4 Policing Outcomes

Arrests are a common metric when studying individual officer and departmental performance and, in the light of concerns about over-policing — excessive and detrimental interactions between law enforcement and civilians — arrests are the main metric I use to measure officer enforcement activity. Opportunities for officers to make arrests depend on the crime rates where and when they work, influencing their quantity, quality, and kind of arrests. To distinguish between the seriousness of arrests, I divide them based on crime types: serious arrests, which I define as arrests for official FBI index crimes, as well as additional forms of homicide, fraud, domestic violence, sexual assault, and simple assault and battery; and low-level crimes are all other arrests—e.g., warrant, traffic, or drug crimes. I am also able to classify arrests based on arrestee race/ethnicity. Using Cook County court data, I determine if the arrest is associated with at least one guilty finding and interpret this as one measure of arrest quality, whereas 'non-guilty' arrests mean no charges were brought or all charges were either dropped, dismissed, or found not guilty. See Appendix C for more details.

Table 2 displays arrests per shift and observations in the daily panel data for all sample recruits by exam period. The average number of arrests per 100 officer-shifts declines over time, from around 18 for Exam 2006 to 15 for Exam 2013 arrests, consistent with general declines in crime in Chicago over the period. The composition of arrests across periods remains relatively constant. Black civilians make up the vast majority of arrests, over 75%, despite making up 30% of Chicago's population, followed by Hispanic arrests being far less common at around 15%. About 70% of arrests are for low-level crimes, and around 20% of all arrests eventually result in a guilty finding; however, Exam 2013 cohorts had a noticeably lower guilty rates at around 13% due to more of their arrests having less time to be completed — the median Exam 2013 officer exits their probationary period in late 2017.

Officers differ substantially in their arresting outcomes for low-level arrests, which are significantly more discretionary, across race, gender, and age. Figure 1 displays coefficients recovered from regressing average serious and low-level arrests on the interaction between officer gender (Male, Female), start age groupings — < 27, $[27, 32]$, > 32, with cuts corresponding to the 37th and 76th percentiles — and race groupings — Black, White, and non-Black minority (around 90% Hispanic and 10% other minorities) — all coefficients being relative to young (< 27), white, male officers in the sample. While female officers tend to make fewer serious arrests relative to their male counterparts, these differences are always less than 0.9 arrests per 100 shifts, and there are minor differences across age and race. For low-level arrests, Black officers consistently make fewer arrests than their non-Black counterparts across age and gender groups, with Black female officers (across all age groups) making roughly 10 fewer low-level arrests per 100 shifts compared to young white male officers.

Young male officers make the most low-level arrests, though older male officers make significantly fewer across all races. For example: young Black male officers make 4 fewer arrests per 100 shifts compared with young white males, while older Black male officers make

7 fewer per 100 shifts. There is a stronger relationship between age and arrests for non-Black officers: white and non-Black minority officers average around 6 and 4 fewer arrests per 100 shifts when starting at > 32 years old compared with starting at < 27 years old. However, the decline in low-level arrests between the young and mid-aged officers compared with mid-aged and older officers is more pronounced for Black and non-Black minorities than for white officers which display more similarly sized declines across groupings. These differences may be the result of differences in assignments or behavior on the job, reflecting differences in personal preferences for aggressive policing or biases, for example.

4 Empirical Strategy and Internal Validity

4.1 Empirical Strategy

The focus of this study is to determine if cohort diversity has a significant impact on officers' future outcomes. I leverage the institutional feature that recruits are assigned to cohorts through random lottery numbers. This random assignment to a specific cohort, conditional on the exam period, results in variation in recruits' academy peer group. Identification requires random assignment conditional on exam period because officers may sort into which exam period they participated but within exam period they have no control over their randomly determined order/lottery number.

To recover the effect of peer characteristic(s) on an outcome of interest, I estimate the following model:

$$Y_i = \pi_1 \overline{X}_{c(-i)} + \pi_2 X_i + \eta_{p(i)} + \epsilon_i \tag{1}$$

where Y_i is officer i's outcome (e.g., average low-level arrests per shift), X_i is officer-specific characteristics, $\overline{X}_{c(-i)} = \frac{1}{N_c - 1} \sum_{j \in c(i) \setminus i} X_j$ is the corresponding peer characteristic for officer i's cohort c, excluding officer i themselves (referred to as a leave-out mean or jackknife measure of peer composition) of peer characteristic(s) X, and $\eta_{p(i)}$ denotes the exam period (p) fixed effect. Equation (1) can be estimated with OLS.

A central identifying assumption required for an unbiased estimate of π_1, or the effect of peer characteristic(s) $\overline{X}_{c(-i)}$ on outcome Y_i, is that peer groups are randomly assigned, ensuring that peer composition is unrelated to unobservable officer-level characteristics which also influence outcomes, ϵ_i. The lottery number system should provide such exogenous group assignments, and I probe the validity of this assumption in the next section.[7]

[7] In this setting, I cannot distinguish between endogenous and exogenous effects of peers (Manski (1993)), which means I cannot directly disentangle the effect of officers being affected by minority peers due to their

Beyond the random assignment of officers to cohorts, this design has additional complications for interpreting estimates of π_1 as a peer effect, stemming from officers being randomly assigned to cohorts and not to peer characteristics. First, it is not possible to determine if an estimate of π_1 is due to differences in cohort composition in characteristic X or if some unobservable characteristics correlated with X are unbalanced across cohorts because there is no exogenous variation in any single peer characteristic. This issue applies to correlated observables and unobservables, meaning an officer who is 'treated' with higher shares of Black peers will also be 'treated' with peers who are observably more female and older in addition to any number of correlated unobserved dimensions such as personality or bias. While I attempt to address this issue for observables by introducing additional peer controls, I cannot overcome the fundamental issue of unbalanced unobservables.

Similarly, the identifying variation for π_1 comes from finite-sample variation in peer groups, and this variation disappears as group (cohort) sizes increase, potentially leading to spurious results which pass as peer effects (Angrist (2014)). This is akin to a weak instruments problem because exogenous assignment to one of many cohorts is used to generate variation in the peer characteristics of interest, and any violation of the exclusion restriction is amplified by weak instruments (Bound, Jaeger, and Baker (1995)). Furthermore, if we view observable characteristics as proxies for unobserved characteristics — such as racial bias or preferences for aggressive policing — then because proxy variables are mismeasured, estimated peer effects will either be attenuated if cohort assignment is random or inflated if random assignment is violated under classical measurement error (Acemoglu and Angrist (2000), Angrist (2014), Feld and Zölitz (2017)). While all of the criticisms of peer designs in Angrist (2014) cannot be addressed, in Appendix A.3.4, I provide additional results in an attempt to determine the extent of these issues, including adding measurement error to cohort compositions and applying 2SLS and split-sample IV procedures (Angrist and Krueger (1995)).

Interpreting the results in light of potential policy changes presents additional complications. I model peer effects as linear-in-means, so the results cannot speak to the optimal assignment of officers to cohorts or optimal cohort composition, as increasing a peer characteristic in one cohort means reducing it in another while holding total composition fixed (e.g., swapping a white officer in one cohort with a Black officer in another). Furthermore, the research design is not that of a systematic change in overall diversity, rather the identifying variation comes from finite-sample variation in compositions. As a result, while changing the overall composition of the academy, consistent with a change in hiring policy, would result in

behavior or characteristics. I assume there are no correlated effects (e.g., instructor effects)– given the large amount of courses recruits are taught during the academy, it is unlikely that a cohort with 40% minority composition would receive different institutional environments or instructors than a cohort with 50% minority composition starting a month later. However, this is further explored in Section 6.2.

net changes in cohort composition, the results from this research design are only externally valid to such policy changes if the mechanisms through which peer composition influences outcomes is also externally valid.

4.2 Random Assignment

I test for violations of the central identifying assumption that officers are randomly assigned to cohorts within exam periods. Table 3 displays the p-value of a joint F-test resulting from a multinomial logistic regression of assigned cohort on officer characteristics for each of the three exams separately and pooled with a restricted (officer race, gender, and start age) and expanded set (military experience, Spanish-speaking ability, and education) of predictors. Across all samples and both sets of predictors, the results indicate that officer characteristics do not predict which cohort an individual is assigned to, with only one test (Exam 2006 with expanded predictors) being even marginally significant.

Next, I test if officer characteristics are strongly associated with cohort composition. Table 4 displays the results of regressing an officer's cohort characteristics — average start age, share male, share minority, and start date — on their individual characteristics. Based on Columns (1)-(3), officer characteristics explain very little of the variation in cohort composition with economically insignificant coefficients. As expected, individual officer characteristics are generally negatively associated with cohort composition of that characteristic — since cohort shares exclude the officer in question, it reduces the pool of officers with that characteristic, as noted in Guryan, Kroft, and Notowidigdo (2009).

Given that cohorts begin successively and not all at the same time, there may be some amount of selection out of the academy by officers who give up, find other jobs, are no longer eligible (too old, moved out of Chicago, could not pass the physical exam, etc.). Column (4) of Table 4 regresses when the officer started at the academy (in years since 2009) on officer characteristics to determine if this delayed entrance and selection significantly alters the composition of recruits. Only officer start age has a statistically significant effect and is unsurprisingly positively associated with start date. Overall, this evidence suggests that attrition is not significantly impacting the composition of cohorts or associated with differences in officer unobservables. Furthermore, applicants wait over a year before the first cohort is called in, meaning the least committed applicants likely select out once they receive their numbers and understand where they are in the pool.

4.3 The Effect of Cohorts on Arrests

Given that officers are randomly assigned to cohorts, not peer characteristics, this environment is best suited for studying the role of a cohort on officers' future outcomes. If there is any economically significant relationship between peer composition and officer outcomes, we expect that assigned cohort would also have a significant reduced form relationship with the outcomes of interest. If not, relationships between peer composition and outcomes are more likely to be spurious. Alternatively, cohorts may have strong influence on officer outcomes even if peer characteristics such as race or gender have no effect.

To recover the relationship between cohorts and outcomes, I recover estimates of cohort-effects on outcomes after controlling for officer level characteristics and exam period fixed effects by estimating:

$$Y_i = \chi_{c(i)}\rho + X_i\beta + \alpha_{p(i)} + \mu_i \tag{2}$$

with OLS, where $\chi_{c(i)}$ are cohort indicators (excluding one per exam period), X_i are controls for officer race, gender, military experience, education, and start age, and $\alpha_{p(i)}$ are exam period fixed effects. I first test whether estimated cohort effects ($\hat{\rho}$) are jointly zero. Column (1) of Table 5 displays the corresponding Wald test p-values and rejects the null hypothesis that cohort effects in the pooled sample (and in the Exam 2010 and Exam 2013 subsamples) are jointly zero for officers' mean low-level and serious arrests using heteroskedasticity-robust standard errors.

Next, I assess the magnitude of the cohort effects and quantify the amount of variation in Y_i that can be explained by cohort-specific effects by recovering the variance of $\hat{\rho}$. For comparison, I partial-out the influence of officer characteristics and exam effects on the outcome by computing $\tilde{Y}_i = Y_i - X_i\hat{\beta} - \hat{\alpha}_{p(i)}$, such that $Var(\tilde{Y}_i) = Var(\hat{\rho}\chi_{c(i)}) + Var(\mu_i)$. As Column (3) of Table 5 shows that one standard deviation (SD) of cohort effects equates to 2.7 low level arrests and 0.5 serious arrests per officer per 100 shifts, and Column (5) shows that 1 SD in the cohort effects accounts for 25% and 17% of the a standard deviation of the residualized outcomes for low-level and serious arrests.

The variance, and thus standard deviation, of estimated cohort effects is biased upward due to sampling error meaning these estimates are overstated (Krueger and Summers (1988), Aaronson, Barrow, and Sander (2007), Kline, Saggio, and Sølvsten (2020)). To account for this, I implement Kline, Rose, and Walters (2022) (KRW) heteroskedasticity robust correction, with corrected standard deviations in Column (4), and Column (6) displays the share of 1 SD in the outcome accounted for by 1 SD in the corrected standard deviation. For low level arrests, sampling error accounts for between 4% to 17% of the estimated standard deviations, though the KRW correction cannot be computed for the Exam 2006 sample. The KRW corrected standard deviations for the pooled sample, Exam 2010, and Exam 2013 samples indicate that a 1 SD increase in cohort effects equates to between 1.9 and 2.6 additional low level arrests per officer per 100 shifts. For context, this 1 SD change in cohort effects on low level arrests is substantial: it is similar to the Black-White gap in total arrests per 100 shifts as estimated in Ba, Knox, et al. (2021) of around 2 arrests per 100 shifts, and it equates to around a 20% of the mean low-level arrests per shift in Table 2.

However, for serious arrests, sampling error appears to account for all or a substantial amount of the variation in cohort effects: in the pooled sample, around 95% of 1 SD is due to sampling error while in the Exam 2013 sample around 34% is due to sampling error— though sampling error is sufficiently large such that corrections cannot be computed in the Exam 2006 and Exam 2010 samples. For the pooled and Exam 2013 samples, a 1 SD increase in bias corrected cohort effects equates to an increase in serious arrests per 100 shifts per officer by 0.02 and 0.3— meaning a 1 SD increase bias corrected cohort effects accounts for less than 10% of a standard deviation of the partialled outcome in Exam 2013 and less than 1% of a standard deviation in the pooled sample. Compared with the means in Table 2, a 1 SD increase in cohort effects will increase an officer's serious arrests per shift by at most 6% of the mean.

Overall, while cohort effects account for a sizable portion of the variation in low-level arrests, any estimated effects of peer composition on serious arrests should be interpreted with caution due to the weak reduced form relationship. The implications are that cohort-specific factors drive some part of the differences in officers. Their magnitudes suggest these factors are important for policy, particularly if the aim is to reduce the number of low-level arrests officers make, as the 1 SD change is sizable compared with both the mean outcomes and prior estimates of differences in arrest behavior across officers. These cohort effects may be driven my a variety of factors, such as peer composition, instructors, field training officers, or timing and policing environment. In the following sections, I focus explore the role of peers in the academy in explaining these effects.

5 Main Results

5.1 Effect of Peer Composition on Arrests

Table 6 displays the results from estimating equation (1) with OLS, with the variables of interest being leave-out means of officer characteristics at the cohort level, with the dependent variables scaled to be in units of average arrests per 100 shifts, and standard errors clustered at the cohort level. Columns (1)-(6) display difference specifications while panels A and B correspond to average low-level arrests and serious arrests per shift respectively.

In Column (1) of Panel A, I find that Black peers have an economically significant but noisy effect on low-level arrests, with a coefficient of -11.89 ($p>0.1$) on (cohort) share Black. This indicates that a 5pp increase in Black peers (roughly 1 SD) — equivalent to replacing 4 white officers with Black officers in the average cohort — reduces an officer's future low-level arrests by 0.59 per 100 shifts, roughly a 6% decline relative to the mean. However, this is not statistically significant and we cannot reject that a 5pp increase in cohort share Black would actually increase low-level arrests by 2% compared to the mean or decrease them by 13% compared to the mean. For serious arrests, higher shares of Black peers cause a modest positive effect which is marginally significant, equivalent to a 5pp increase in Black peers increasing serious arrests by 3% compared with the mean. Non-Black minorities display effects are neither statistically or economically significant for either low-level or serious arrests.

Column (2) focuses on peer gender composition, specifically the influence of higher shares of female peers compared with male peers. The point estimates are similar to those of Black peers though more precise, with a coefficient of -13.74 ($p=0.07$). A 5pp increase in the share of a cohort which is female (roughly 1 SD) reduces an officer's low-level arrests by 6% compared with the mean and we fail to reject effects between a 13% decrease and a 0.4% increase. For serious arrests, the effect is statistically significant at the 5% level, and a 5pp increase in female share increases serious arrests by 3% relative to the mean.

Column (3) and (4) focus on the role of peer age. In Column (3), I split officers into 'older' and 'younger' groups by whether they started at the academy above or below the median age of 28.24 years. While the effects on serious arrests are statistically and economically insignificant, older peers have a sizable influence on low-level arrests: a 10pp increase in the share of a cohort which is above 28.24 years old (just below 1 SD of 13pp), reduces officers' low-level arrests by 0.88 ($p<0.05$) arrests per 100 shifts or an 8% decrease relative to the mean. This estimate is relatively precise compared with prior estimates, and with the 95% confidence intervals we can rule out declines larger than 15.7% and smaller than 0.7% compared with the mean.

Column (4) reduces the cutoff age between 'older' and 'younger' officers to being 27 (the 37th percentile). The motivation for this is the non-linear relationship between age and outcomes as shown in Figure 1 for Black and non-Black minority males, and it effectively divides between the 'youngest' officers and the mid-aged and older officers. The relevance of this division is reflected in Figure 3 which displays estimates of equation (1) using alternative age cutoffs between 25 and 34 years old (the 14th and 80th percentiles). I find that largest and most precise estimates of the influence of 'older' peers is when the age cutoff is between 26 and

28 years, and the estimates tend to decrease in magnitude as the age cutoff increases. This is consistent with peer age being mostly important as differentiating between the relatively young peers compared with the middle-aged and older peers.

The estimates in Column (4) of Panel A are larger than those in Column (3) and relatively more precise, though effects on serious arrests are still small and not statistically significant. A 10pp increase in a cohort's share of officers older than 27 (1SD being 13.4pp), decreases an officer's future low-level arrests by 1.39 per 100 shifts, equivalent to a 13% relative to the mean. This estimate is relatively precise, as we can reject effects larger than a 22.3% decline and smaller than 3.5% decline compared with the mean.

Cohorts, not demographics, are randomly assigned, and race, age, and gender are correlated. As a result, the prior estimates do not determine which characteristic(s) is driving the results. Column (5) combines the specifications in Columns (2) and (4), and shows that the effects of older (>27) peers and female peers are not due to female officers also tending to be older: point estimates are almost identical and both estimates are more precise for both serious and low-level arrests. The 95% confidence intervals for a 5pp increase in share female ranges between declines of 11.9% to 0.9% relative to the mean, and a 10pp increase in share older peers ranges from a 21.4% to a 4.1% decline relative to the mean. For low-level arrests, the effects of female and older peers are both statistically significant after controlling the family-wise error rate using the Holm (1979) adjustment.

Figure 3 also displays the same coefficients as those in Column (5), with both share female and share age > cutoffs between 25 and 34. While the coefficient on share female is stable across cutoff ages and almost always significant at the 10% level at least, it is most precise when using cutoff ages between 26 and 28. Again, this is consistent with grouping peer ages by those who are 'young' versus those who are not young as explaining more of the variation in the outcome than when the age cutoff is increased. For the remainder of the paper, I will focus on 27 as the older vs. younger cutoff, as it explains more variation than older cutoffs and results in more precise effects.

Finally, Column (6) combines the specifications in Column (1) and Column (5). Shares of Black and non-Black minority peers are both small and not statistically significant, while the effects of female and older peers remain large and statistically significant. With four parameters of interest, Holm (1979) adjusted p-values are larger, with share female having $q = 0.138$ and share age > 27 having $q = 0.066$ in Panel A. Overall, these results indicate that peer age and gender composition have an economically and statistically significant influence on future low-level arrests.

5.2 Distinguishing between Behavior and Assignments

The prior results pose two main questions for policy implications and external validity. First, through what channel does peer composition influence arrests? Second, why do peer characteristics matter in changing arrest outcomes? I explore the first question in this section by distinguishing between the two primary channels through which academy peers influence officers' future arrest outcomes: behavior and assignments. If peers influence each others' behavior then we would expect similar results in other environments and hiring policy changes could result in behavioral spillovers and net changes in arrest outcomes. Conversely, if peer composition affects an officer's assignment, the generalizability of the relationship between peer composition and future arrests may be limited, as total assignments would remain fixed and any net increase in any officer demographic would not produce a net change in assignments.

5.2.1 Measuring Behavior and Assignments

To separately estimate the influence of cohort composition on the portions of officers' arrests attributable to assignments versus those attributable to their behavior, I disentangle the contributions of officers' behavior and their assignments on arrests. Specifically, I decompose an officer i's arrests on a given shift t into an officer-specific effect (θ_i), the effect of their assignment ($\gamma_{a(i,t)}$), the influence of time-varying factors (V_{it}), and exogenous shocks ($\epsilon_{i,t}$). Similar to the teacher-value added literature (Rothstein (2010), Chetty, Friedman, and Rockoff (2014)), we can think of θ_i and $\gamma_{a(i,t)}$ as the officer's and the assignment's additive contribution to arrests made. I interpret officer effects as a measure of an officer's individual propensity to make arrests holding fixed their working environment, reflecting their personal preferences for enforcement, biases, and beliefs.

In the main specification, I define an assignment as the interaction between officer i's assigned district and truncated beat code (b), their role (r), their shift number (s), and the year, month, and day of the week (w_t): $a(i,t) = brsw_t(i)$.[8] I recover estimates for both θ_i and $\gamma_{a(i,t)}$ by estimating the following model:

$$Arrest_{it} = \theta_i + \gamma_{a(i,t)} + \beta V_{it} + \epsilon_{it} \tag{3}$$

[8]Formally, b is the numeric beat code with the last numerical digit removed, and r is the exact role designated by the full beat code. For example, beat code "2533" has a role of 'beat officer' and the beat is truncated to "253" which indicates the sector they work in (a group of contiguous geographic beats, and beats are on average less than 1 square mile). Beat code "2463A" has a role of tactical team C officer, as does beat code "2463C", and both have the same truncated beat as "246" (which does not map to a geographic sector), so their $brsw_t$'s are identical if they also work in the same watch, day of week, month, and year.

where $Arrest_{it}$ is the number of arrests officer i made during their on-duty time on date t, and V_{it} contains a polynomial of officer tenure. The data contain over 7.8 million officer-shift observations with 13,000 officer effects (θ_i) and approximately 580,000 assignment effects (γ_{brsw_t}).

I assume that conditional on a polynomial of officer tenure and officer and assignment fixed effects that: current and future shocks to arrest counts are orthogonal to past observables, and shocks to arrest counts are not serially correlated across shifts. Regarding officer-assignment sorting, identification requires that officer i's movement between assignments cannot be related to ϵ_{it}; however, there is no restriction on officers sorting into or out of assignments based on their fixed effects. Importantly, the CPD's operational schedule reinforces the inability of officers to select shifts on specific days or civilian pools (see Ba, Knox, et al. (2021) for more detail), which reduces concerns over officers selecting working conditions based on shocks.[9]

Figure 2 explores the relationship between officer-level observables (race, gender, and age bins) and officers' estimated arrest propensities ($\hat{\theta}_i$) and averaged assignment effects over an officer's career ($\bar{\hat{\gamma}}_{a(i,t)}$), decomposing their respective effects on arrests from Figure 1.[10] For low-level arrests (Panel A), while both assignment and officer effects match the general relationships between race, gender, and age as seen in the average arrest rates, officer effects contribute significantly more. For example, Black female officers make an average of almost 10 fewer low-level arrests per shift than young white male officers, but while differences in assignment effects account for 3 low-level arrests, differences in officer effects account for 7 low-level arrests. Across all demographics, assignment effects contribute significantly less to the differences than officer effects. There is relatively little sorting on assignments for serious arrests (Panel B). However, Black officers tend to work in higher serious arrest assignments, and differences in behavioral effects are small as well, though female officers tend to have lower officer effects overall. Consistent with the arrest patterns in Figure 1, these results suggest that older, female, Black, and, to a lesser extent, non-Black minorities have lower individual propensities to make low-level arrests on average, suggesting preferences for less

[9]For recovering officer and assignment effects, we require an additional condition. Following the literature on worker and firm fixed effects (Abowd, Kramarz, and Margolis (1999), Abowd, Creecy, and Kramarz (2002), Card, Heining, and Kline (2013), Bonhomme et al. (2022)), I can recover θ_i and $\gamma_{a(i,t)}$ only for officers and assignments that are within a connected component after the officer-assignment data has been transformed into a bipartite graph. Given movement across assignments and the long time series, the largest connected component contains over 95% of all officers and assignments.

[10]While not a perfect decomposition of average arrests, as tenure effects explain some portion of arrests, estimates of θ_i and $\bar{\gamma}_{a(i,t)}$ together explain over 88% and 93% of the variation in average serious and low-level arrests for sample officers, respectively. In both cases, officer effects explain more variation than assignment effects, with officer and assignment effects within exam periods being positively correlated for low-level arrests and negatively and more weakly correlated for serious arrests.

aggressive policing, with a similar pattern in terms of sorting into less aggressive (lower low-level assignment effects) assignments.

5.2.2 Effect of Peer Composition on Assignments and Behavior

With measures of the contributions of behavior and assignment to arresting outcomes, I attempt to determine through which channel peer composition influences arrests. This is primarily an exploratory exercise because estimating the relationship between peer composition and $\hat{\theta}_i$ requires holding fixed $\overline{\hat{\gamma}}_{a(i,t)}$ which peer composition may also influence, and vice versa.

I begin by estimating the relationship between peer composition and district-level assignments. Cohort composition may influence an officer's working environment by influencing their unit assignment. Because new officers have no seniority, they can be assigned to specific districts/units based on departmental demand, to avoid excessively long commutes, or to have officers reflect the communities they police, based on conversations with a retired officer.

For example, if Black officers are more likely to be initially placed in majority Black districts, with differential arrest opportunities, having more Black peers may influence their assignments and thus their arrests. Columns (1)-(4) of Table 7 display the relationship between officer's working district characteristics and their cohort peer composition and their characteristics, using the main specification from Column (6) of Table 6. Unlike the relationship between peer composition and arrests, these results suggest that peer race, age or gender do not have a strong relationship with district characteristics, with generally small and highly imprecise effects — for example, no joint F-test on the peer composition coefficients rejects at the 5% level. The lack of an effect on district-level characteristics can be explained by new officers being generally assigned to districts in which officers of higher tenure are more likely to exit with higher Black population shares and higher crime rates (Ba, Bayer, et al. (2021)). This suggests that peer composition influencing district-level assignments is unlikely to explain the relationship between peer composition and future arrests.

Next, I focus more directly on the influence of peer composition on officers' average assignments effects on arrests ($\overline{\hat{\gamma}}_{a(i,t)}$). Columns (5) and (6) display the effects of peer composition on officers' averaged assignment effects for low-level and serious arrests per 100 shifts. For low-level arrests, higher shares of Black, non-Black minority, older, and female peers increase officers' future assignment effects, meaning they have a positive effect on arrests, all else equal — though the only effect which is marginally significant is for non-Black minority peers. For serious arrests, the same peer groups decrease assignment effects with economically significant effects and statistically significant effects for non-Black minority and older (> 27) peers. The effects on assignments tend to be opposite of the effects on average arrests discussed in the main results. Overall, I find no evidence that effects on assignments are the primary channel for peer composition's influence on arrests.

Columns (7) and (8) display the results for the effects of peer composition on the measure of officer behavior, their individual arrest propensity ($\hat{\theta}_i$). The impact of peer composition on officer arrest propensities is consistent with the effects on average arrests, though larger and generally more precisely estimated, consistent with the attenuating effects of noisy and opposite-signed effects of peer composition on assignments. A 5pp increase in the share of female and older peers decreases an officer's low-level arrest propensity by 1.25 ($p<0.05$) and 1.15 ($p<0.01$) arrests per 100 shifts. Black and non-Black minority peers have economically significant negative effects but neither are statistically significant (-0.55 ($p>0.1$) and -0.38 ($p>0.1$) for a 5pp increase per 100 shifts). For serious arrests, 5pp increases in cohort shares of Black, non-Black minority, female, and older peers increases arrest propensities by 0.28 ($p=0.06$), 0.14 ($p>0.1$), 0.42 ($p<0.01$), and 0.34 ($p<0.01$) per 100 shifts.

Overall, these results suggest that police academy composition influences arrest outcomes primarily through changes in behavior. Effects on officers' individual arrest fixed effects are consistent with the main results for average arrests. In contrast, officers' district characteristics and average assignment fixed effects are more weakly influenced by peer composition, and the latter's effects are generally in the opposite direction compared with the main results.[11] This suggests that main results are potentially externally valid and overall changes in recruit composition, such as changes to hiring policy, would result in overall changes in outcomes through peer effects in addition to officer-level differences.

5.3 Additional Outcomes

5.3.1 Persistence of Effects

It is important to determine if these effects are short-term or if they persist into an officer's career. While the careers observed for sample officers range from almost 10 for Exam 2006 to less than two years for some Exam 2013 officers, I can explore the persistence of peer compositions' effects on officer outcomes over a moderate time frame. I divide officers' careers into roughly half-year (180 days) periods, then recover average arrests per shift and separate officer effects for officers during each period. I re-estimate equation (1) on average arrests and

[11] The assignment-level results for serious arrests do not indicate a clear crowding-out or crowding-in effect across characteristics — where crowding out would be indicated by, for example, Black officers working in lower low-level arrest assignments pushing their peers to work in higher low-level arrest assignments. While for low-level arrests, the effects are more consistent with a crowding-out effect, where officers with peers who prefer lower low-level arrest assignments crowd out others' from working in those assignments, though the results are noisy and the precise mechanism is unclear.

officer effects for that tenure period, including sample officers who appeared in at least three periods. I do this for tenure periods one through nine, covering at most the first 4.5 years of an officer's career as a full officer, about six years after they entered the academy — though the sample becomes increasingly small toward the final periods. Figure 4 displays the results.

The effects of Black and non-Black minority peers on low-level average arrests remains near zero and noisy over time, while their effects on officer effects grow larger (more negative). Female and older peers display consistently negative effects on average low-level arrests and officer effects until around 3 years (6 periods); while the effect of female peers moves toward zero for both outcomes after that, the effect of older peers remains consistently negative and close to the baseline estimates. Effects on serious arrests almost all revert to zero or flip sign over the periods. Overall, the results indicate that effects on low-level arrests are relatively persistent even at 3 years after the officer exited the academy, though the available sample becomes increasingly small over time.

5.4 Arrest Subtypes and Quality

As peer composition influences officer arrests, with evidence consistent with this being the result of differences in behavior, it is important to determine if these effects are not detrimental to overall officer performance. Specifically, the decrease in low-level arrests and increase in serious arrests could be due to officers refocusing their efforts on more serious crime types or making lower-quality arrests and worsening criminal justice outcomes.

The classification of arrests into two categories obscures significant heterogeneity in term of what forms of arrests are made, e.g., serious arrests contains both index violent crimes, such as robberies, and nonindex property crimes, such as vandalism. I disaggregate serious and low-level arrest categories into multiple subtypes and re-estimate equation (1) using the main specification with results in Table B.4, using both average arrests (Panel A) and officer arrest propensities (Panel B) as the outcomes of interest. For serious crimes, results are inconsistent across arrests and propensities, though property arrests are generally positively affected. For low-level crimes, in both average arrests and officer effects, peer effects operate through fewer drug, 'other' miscellaneous nonindex (e.g., disorderly conduct), municipal code violations, and warrant arrests. Notably, while peer effects are negative for average weapon arrests, this is likely due to assignment effects, as officer propensities to make weapon arrests have small positive effects (while traffic arrests display the opposite pattern). Overall, these results suggest that officers are shifting focus from lower to higher value crimes or, at worst, simply away from low-severity crimes without reducing effort applied to high-severity crimes.

Next, one feature of my data is that I can observe the outcomes of arrests in court, which enables me to measure one form of the quality of arrests. To study the effect of peer diversity on arrest quality, I consider arrests as those which result in a guilty outcome in court as high-quality and arrests that result in a non-guilty outcome in court as low-quality, though in reality a guilty finding does not always map to a correct conviction nor does a non-guilty outcome always mean the defendant was innocent. By comparing the effect of peer composition on the high (guilty) and low (non-guilty) quality arrests, I can infer its impact on arrest quality. In Table B.3, I present the results for officers' average arrests and arrest propensities in Panel A and B, respectively, with guilty and non-guilty low-level arrests in Columns (1) and (2) and the same for serious arrests in Columns (3) and (4).

For both average arrests and officer effects, peer effects generally operate similarly on both guilty and non-guilty arrests. For low-level arrests, older and female peers have a larger effect on non-guilty compared with guilty arrests, though the declines in guilty arrests are more precisely estimated. However, a larger proportion of low-level arrests are non-guilty, with the guilty over non-guilty ratio of means being 0.21. In contrast, the relative declines in guilty arrests are larger, suggesting a reduction in average arrest quality, but the estimates are too imprecise to make a definitive statement. Changes in serious arrests are similarly imprecise for non-guilty arrests, though the results suggest a decrease in arrest quality on average.

For officer effects, estimates are relatively more precise, and the coefficients for non-guilty arrests are larger. For low-level arrests, declines in guilty arrest propensity are much smaller than declines in non-guilty arrest propensity, such that older and female peers have a larger effect relative to their standard deviations, with the ratio of standard deviations of guilty to not-guilty low-level officer arrest propensities in the sample (0.26), implying an improvement in officers' ability to make higher quality arrests. Again, estimates are too imprecise to confidently say officers' propensity to make high-quality arrests improved. For serious arrests, the estimates suggest that Black, non-Black minority, female, and older peers increase an officer's propensity to make high-quality arrests (compared a standard deviation ratio of 0.43), though comparisons with the ratio of means (0.3) suggest that only non-Black minority peers improve quality. For serious arrest propensities, comparisons with the mean ratio indicate increases in quality while comparisons with the ratio of standard deviations suggest decreases. Overall, I find evidence for increases and decreases in arrest quality, depending on the outcome and the measure of a quality increase.[12]

[12]In prior versions of the paper, the results were consistent with an increase in arrest quality. This was driven by 1. a focus on officer effects and not average arrests, and 2. using shrunken fixed effects as the dependent variable which increased precision and produced much smaller magnitudes for the change in officer effects for guilty arrests. However, shrinking officer effects as the dependent variable introduces bias, so this analysis has been removed.

6 Mechanisms and Policy Implications

The prior results are consistent with peer age and gender being paramount factors in influencing future arrest outcomes through influencing officer behavior, though peer race has an economically large but not statistically significant effect on officer behavior. The primary questions are by what mechanism are peers influencing officer behavior and arrests? And what are the resulting policy implications? I begin by proposing and testing multiple mechanisms with different policy implications.

6.1 Primary Mechanisms

There are two primary mechanisms which can explain why peer composition influences outcomes and behavior. Peers may influence each other through their preferences — peers who prefer more aggressive policing may influence their cohort-mates to have more aggressive preferences as well — or socialization between minority and non-minority officers could reduce racial bias and thus reduce arrests of minority civilians. For policy, if officer race, not officer preferences, is the primary factor in altering peer behavior, then hiring policy would want to focus primarily on recruiting minorities. However, if preference spillovers rather than interracial contact are driving the results, then factors that are observable proxies for preferences should be considered when making changes to hiring policy. Given that Black peers do not have the largest effects compared with female and older peers, it is unlikely that interracial socialization is the primary driver of the effects, though it could play a secondary role.[13]

Officers may adopt the preferences of their cohort-mates due to shifts in culture, social identity, or personality (Akerlof and Kranton (2000), Anwar, Bayer, and Hjalmarsson (2019), Golsteyn, Non, and Zölitz (2021)). For example, having more peers who prefer aggressive policing will cause an officer to police more aggressively in the future. As previously discussed, the differences in low-level arrests (Figure 1) driven by differences in arrest propensities (Figure 2) across age, race, and gender bins are consistent with differences in preferences for aggressive policing of low-level crimes. Under an interpretation of low-level arrest propensities as reflective of preferences for aggressive policing, the prior results indicate that female, older, and to a lesser extent minority officers have lower individual preferences for aggressive policing of low-level offenses, which could then influence their peers' future behavior.

[13]The existing literature on policing emphasizes the role of race in policing (Ba, Knox, et al. (2021), Goncalves and Mello (2021), Hoekstra and Sloan (2022)), and the peer effects of diversity literature suggest a positive interracial socialization mechanism may be important (Boisjoly et al. (2006), Carrell, Hoekstra, and West (2019)). Since non-whites comprise over 90% of new officer arrests, reducing prejudices against such groups could be an important channel in affecting arrests.

As a first step, I explore the potential for minority peers' effects being mediated through their age or gender. In particular, the role of race may be partially confounded by the correlation between officers being older and officers being Black. Furthermore, it is unclear if all older or female peers influence outcomes equally, or if the effects are driven by one subset, for example white female peers. Larger effects for minority peers compared to white peers, conditional on age and gender would also be consistent with the patterns displayed at the officer-level in Figures 1 and 2, and thus consistent with a preference mechanism. To test this, I introduce interactions of peer race and peer age and gender bins into the main specification with average arrests and officer fixed effects as the outcomes of interest, with results displayed in Table 8.

Panel A of Table 8 estimates equation (1) focusing on the effects of the share of older (>27) minority peers, younger minority peers, older white peers, with younger white peers being the excluded group, as well as including the share of female peers. For low-level arrests, female peers and older minority peers have the largest effects (-15.1 ($p<0.05$) and -14 ($p<0.05$)), while older white peers have a negative effect that is economically large but not statistically significant (-9.6 ($p>0.1$)), and younger minorities have a positive and highly imprecise effect. For officer effects for low-level arrests, both minority peer groups have negative effects though older minorities have a much larger and precise effect (-32.3 ($p<0.01$)) and younger minorities have a small and statistically nonsignificant effect (-5.1 ($p>0.1$)). Older white peers also have a larger and more precisely estimated effect (-22.1 ($p<0.01$)), though smaller than the effect of older minorities, and female peers maintain a large negative effect (-26.8 ($p<0.05$)).

Panel B disaggregates minorities into Black and non-Black minorities and interacts them with the age groups. The main result from this is that older Black peers have a larger effect than older non-Black minority peers, and both are larger than that of older white peers. Finally, Panel C shows that the effects of female peers are largest for minority female peers, though white female peers also display negative, though smaller and not statistically significant effects on outcomes. However, I cannot reject that any of the effects of the older peer groups are statistically different or that the effects of minority female and white female peers are statistically different. Nevertheless, these patterns in peer effects on low-level arrests and propensities are consistent with peers' preferences influencing future preferences for policing low-level crimes, as the groups with the largest effects are also those who make the fewest arrests and have the lowest arrest propensities individually. For serious arrests, the effects are generally opposite of those for low-level arrests, and the economically and statistically significant effects are driven by older peers, with larger effects for older Black and to a lesser extent non-Black minority peers, and female peers.

While these patterns display minimal evidence for interracial socialization driving the results, we can provide an additional set of tests. Suppose interracial socialization and bias reduction drive the results. In that case, whites should be more strongly influenced than minorities by higher shares of minorities (regardless of the minority peers' preferences for aggressive policing). Alternatively, if preferences are the primary mechanism, less aggressive peers should influence all officers, with no stronger effect on whites. I test for such effects by re-estimating equation (1) with additional interaction terms and alternative specification, with results shown in Table B.5. As shown in the table, white officers do generally experience larger effects from minority peers than minority officers do but no additional effect from white (male) peers. However, the main effect on all officers of older minority peers is generally larger than any additional effect on white officers, and in general the is not sufficient power to precisely estimate any interaction effect. Overall, this is consistent with the main effects being due to preference spillovers on all officers, with only modest evidence for a secondary effect of interracial socialization on white officers.

Finally, I construct a proxy for officers' 'preferences' as their predicted arrest rates and officer effects based solely on their race, gender, and age bin. To avoid issues with officers and cohorts contributing to their own predicted propensities, I exclude each cohort from the data and regress officer outcomes on their interacted race, gender, and age bins and include exam period and cohort fixed effects (as in equation (2) using the age, gender, and race bins as in Figure 1). Then I construct predicted values for the officers in the excluded cohorts, and within each cohort I construct leave-one-out means of the predicted values. Table 9 presents the estimated effects of the officer's own predicted outcome and that of their peers. For simpler interpretation, I standardize the peer average predicted arrests and propensities to being 1 SD increases. For low-level arrests, consistent with preference spillovers, 1 SD increases in predicted peer arrests and arrest propensities have economically and statistically significant effects on the officer's arrests and propensities. Effects are between 0.7 and 1.5 arrests per 100 shifts, which is entirely consistent, in terms of magnitudes, as 1 SD changes in peer composition using cohort shares of more or less aggressive officer groups. For serious arrests, predicted peer preferences are actually opposite signed and inconsistent with a preference mechanism.

These results are highly consistent with the central role of peer preferences for low-level arrests: higher shares of officers from groups with preferences for less aggressive policing consistently result in fewer low-level arrests and lower officer propensities for low-level arrests. However, there is also evidence for a supplementary role of interracial socialization resulting in additional effects on white officers. The robustness of these results, using the specification in Panel A of Table 8 are discussed in Appendix A.3. The main implication for policy is that the underlying preferences of recruits, and the demographics which are proxies for these preferences, may matter more for reducing aggressive policing that solely focusing on minority status.

6.2 Alternate Behavioral Mechanisms

Separately from interracial socialization and the influence of peer preferences, cohort composition may result in different working and training environments, resulting in peer effects through other channels: 1. Academy peer composition may change officers' preferences for their future co-workers, resulting in contemporaneous peer effects. 2. If field training officers are matched to recruits based on composition, then one's police academy cohort composition may result in different trainers, influencing outcomes later on. 3. If peer composition changes how instructors train their officers, this could result in future behavioral changes. In each case, such mechanisms would suggest alternative targets for interventions (e.g., trainers) rather than altering hiring policy.

I begin by estimating the relationship between peer composition and co-workers and trainers, with results shown in Table B.6. Columns (1)-(3) provide the relationship between cohort composition, using the main specification, and the average composition of an officer's co-workers as a full officer (same unit, sector, watch, and day). I find some effects consistent with cohort composition influencing selection into co-worker groups: more female and, to a lesser extent, older peers lead to higher shares of female co-workers and lower shares of white co-workers. While consistent with prior work on how exposure to diversity influences future interpersonal relationships (Baker, Mayer, and Puller (2011), Carrell, Hoekstra, and West (2019), Dahl, Kotsadam, and Rooth (2021)), these effects are generally minor, with a 1 SD increase in female cohort share increasing female co-worker shares by less than 1 percentage point (0.6 ($p<0.01$)pp). Overall, the effect sizes are too small to explain changes in officer behavior through contemporaneous peer effects.

Similarly, cohort composition may change the composition or instruction of officers' trainers, who themselves influence outcomes later on (Adger, Ross, and Sloan (2022)). If this is the case, then policy should focus on altering training behavior rather than officer hiring policy. I test for this in two ways. First, I use data on officers' field training officers (FTO) who perform evaluations during their probationary periods to see if peer composition influences FTO composition. As shown in Columns (4)-(6), peer composition does not have a statistically or economically significant effect on FTOs being female or white and has a minor impact on FTO age. Again, these effect sizes are not sufficiently large to explain the established effects of peer composition on officer arrests.

Finally, I perform an indirect test to see if instructors might change their behavior in response to class compositions. Using data on which officers consistently attended academy trainings together, I construct within-cohort 'homerooms' — see Appendix A.4 for more details. If instructors respond to homeroom compositions and this influences officer behavior, then the composition of within-cohort homerooms should drive the effects on future behavior. If officer behavior is altered through cohort composition rather than through influences on trainers, then the composition of these homerooms should not influence outcomes beyond their relationship with cohort composition. As shown in Column (7)-(10), while homeroom composition has a similar effect on future behavior as the main results, including cohort fixed effects generally renders these sub-cohort effects insignificant, inconsistent with instructors responding to group composition leading to altered outcomes.

6.3 Policy Implications

The prior results are consistent with officers' preferences for aggressive policing of low-level crimes influencing their peers' preferences and future arresting behavior after cohorts dissipate. Importantly, female and older minority and, to a lesser extent, older white officers tend to make fewer low-level arrests and have lower individual arrest propensities compared to their younger white male counterparts. Though imprecisely estimated, white officers experience an additional effect from minority peers, regardless of preferences, consistent with a secondary bias reduction effect.

Officers with less aggressive peers tend to increase their arrests for serious offenses. I interpret this as consistent with a reallocation of effort from low-level crimes, following adoption of peer preferences, to more serious ones. However, the serious arrest results are not persistent or entirely robust across analyses, as the effects on them is not robust to using predicted preferences and cohort effects themselves are largely noise. Furthermore, there are mixed results on the resulting change in arrest quality as measured by court outcomes. Nevertheless, recent research suggests that a reduction in arrests of low-level offenses may be welfare enhancing through improved outcomes of civilians not prosecuted for low-level crimes (Agan, Doleac, and Harvey (2023)) and no corresponding rise in crime (Cho, Goncalves, and Weisburst (2021)).

Crucially, the evidence is consistent with peers influencing behavior, rather than assignments which are more likely to be inflexible in aggregate, suggesting that the results may have some external validity. In particular, the results are most relevant for recruitment and hiring policy which can change the pool of officers and thus the average peer composition. The central implication is that preferences of officers hired matters for changing policing outcomes than their race alone. While policies such as procedural justice training have not produced significant changes in officer behavior (Roth and Sant'Anna (2021)) and common-place diversity trainings can be ineffective or counterproductive (Chang et al. (2019), Dobbin and Kalev (2016)), these results indicate that peer effects through recruitment of older, female, and minority officers can be an effective way of reducing aggressive policing. However, changing hiring policy to increase recruitment of minority officers may be counterproductive if it targets minorities with preferences for more aggressive policing. On the other hand, if hiring policy takes additional demographic features and officer preferences into account, then hiring officers with preferences for less aggressive policing may result in reductions in low-level arrests through two channels: the direct effect of the officer and their peer effect onto others in the academy.

While quantifying the effects of academy composition on crime or including all potential channels in order to rigorously conduct a welfare analysis is beyond the scope of the paper, I provide some simple counterfactuals for changes in hiring policy to illustrate the importance of gender and age in addition to race. Notably, these counterfactuals assume that the average and marginal recruit will be similar in their preferences conditional on demographic group. While this is not directly testable, it is plausible because of the large surplus of applicants that pass the CPD tests though the extent of this surplus is unknown and an important factor to consider when changing policy. Even more strongly, they assume no net assignment changes and that the estimated effects on officer fixed effects are an accurate representation of officer behavior. As such, the exact numbers should be interpreted with caution.

I first use estimates from the specifications in Panels A of Table 8 and Table B.5. Suppose, holding the size of cohorts fixed at 60 officers and the total number of recruits fixed and assignments are inflexible (using officer effects rather than average arrests), 5pp (3) more minority officers were hired in place of white officers. If the minority officers were younger (< 27) and the white officers were older (> 27), the minority officers will make similar numbers of low-level arrests individually; however, collectively their peer effects will increase low-level arrests by 0.85 per 100 shifts per officer in the cohort. On net, this hiring policy change would result in a 8% increase in total low-level arrests compared to the mean and a 3% decrease in total serious arrests from hiring additional younger minority officers in place of older white officers. The increases in low-level arrests for whites would be smaller than on minorities (a 4.6% increase compared with 12%), due to the additional interracial effects. On the other hand, if age effects are considered and older minority officers are hired, then the net effect will be a reduction by 5.5% in low-level arrests, and a 3.1% increase in serious arrests, compared to the mean — with more similar effect on whites and minorities for low-level arrests (-5.8% and -3.4% compared with their respective means). Similarly, suppose the share of male officers hired decreases by 5pp (3 officers) but racial and age compositions are held fixed, then using the specification in Panel A of Table 7 low-level arrests would decrease by 12.8%.

7 Conclusion

In this paper, I document the effect of peers in the police academy on police officers' future arrests. I find that the composition of an officer's cohort significantly affects their future low-level arrests. Specifically, higher shares of female peers and older peers of all races lead to fewer low-level arrests. However, in terms of point estimates, the effects of older minority peers are larger than the impact of older white peers. This reduction in low-level arrests is accompanied by a smaller increase in arrests for serious crimes, but I find mixed results on arrest quality, as measured by guilty findings in court. By decomposing arrests into behavior (officer-specific effects) and assignment effects, I find that the results are driven by peers influencing officer behavior, suggesting increased external validity and policy implications.

After considering multiple mechanisms by which peer composition could influence future behavior, including co-worker and trainer composition, the results are most consistent with a preference spillover effect. I find that peer groups (e.g., older and female) that tend to have lower individual preferences for aggressive policing of low-level offenses produce similar peer effects on officers. As a result, the interaction between race, age, and gender is essential as older minority and female minority officers exhibit the least aggressive policing and, correspondingly, the largest peer effects. However, I also find some evidence consistent with a smaller and supplementary interracial socialization effect reducing aggressive policing among white officers. Furthermore, the effects are generally persistent, lasting for at least four years after the academy ends.

Overall, the policy implications of these findings are promising for improving policing. The increased inclusion of minority, female, and older officers can reduce over-policing of low-level offenses through spillovers while not reducing arrests for more serious crimes, thereby potentially improving police-community relationships and public safety. Furthermore, policy changes which result in more recruitment of minorities who prefer more aggressive policing may have self-defeating effects if reductions in low-level arrests are the goal. For example, lowering minimum start ages to have more minority applicants may increase the recruitment of younger male minorities. This may nullify the impact of increased racial diversity due to higher peer preferences for aggressive policing. These results suggest that departments should consider additional characteristics, such as gender and age, when changing hiring policies.

References

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25 (1): 95–135.

Abowd, John M., Robert H. Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data." *Longitudinal Employer-Household Dynamics Technical Papers*, March.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67 (2): 251–333.

Acemoglu, Daron, and Joshua D. Angrist. 2000. "How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws." *NBER Macroeconomics Annual* 15 (January): 9–59.

Adger, Chandon, Matthew Ross, and CarlyWill Sloan. 2022. "The Effect of Field Training Officers on Police Use of Force." Working Paper.

Aendos. 2015. "Chicago Police 2016. Police Forums & Law Enforcement Forums @ Officer.com." August 21, 2015.

Agan, Amanda, Jennifer L Doleac, and Anna Harvey. 2023. "Misdemeanor Prosecution." *The Quarterly Journal of Economics*, January.

Ager, Philipp, Leonardo Bursztyn, Lukas Leucht, and Hans-Joachim Voth. 2021. "Killer Incentives: Rivalry, Performance and Risk-Taking Among German Fighter Pilots, 1939–45." *The Review of Economic Studies*, December.

Aizer, Anna, and Joseph J. Doyle. 2015. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges." *The Quarterly Journal of Economics* 130 (2): 759–803.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *The Quarterly Journal of Economics* 115 (3): 715–53.

———. 2010. *Identity Economics : How Our Identities Shape Our Work, Wages, and Well-Being*. Princeton: Princeton University Press.

Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics*, Special section articles on "what determined the dynamics of labour economics research in the past 25 years? Edited by joop hartog and and european association of labour economists 25th annual conference, turin, italy, 19-21 september 2013 edited by michele pellizzari, 30 (October): 98–108.

Angrist, Joshua D., G. W. Imbens, and A. B. Krueger. 1999. "Jackknife Instrumental Variables Estimation." *Journal of Applied Econometrics* 14 (1): 57–67.

Angrist, Joshua D., and Alan B. Krueger. 1995. "Split-Sample Instrumental Variables

Estimates of the Return to Schooling." *Journal of Business & Economic Statistics* 13 (2): 225–35.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." *The Quarterly Journal of Economics* 127 (2): 1017–55.

———. 2019. "Politics in the Courtroom: Political Ideology and Jury Decision Making." *Journal of the European Economic Association* 17 (3): 834–75.

Ater, Itai, Yehonatan Givati, and Oren Rigbi. 2014. "Organizational Structure, Police Activity and Crime." *Journal of Public Economics* 115 (July): 62–71.

Athey, Susan, Dean Eckles, and Guido W. Imbens. 2018. "Exact p-Values for Network Interference." *Journal of the American Statistical Association* 113 (521): 230–40.

Athey, Susan, and Guido Imbens. 2016. "The Econometrics of Randomized Experiments." *arXiv:1607.00698 [Econ, Stat]*, July. https://arxiv.org/abs/1607.00698.

Austen-Smith, David, and Roland G. Fryer Jr. 2005. "An Economic Analysis of 'Acting White'." *The Quarterly Journal of Economics* 120 (2): 551–83.

Ba, Bocar, Patrick Bayer, Nayoung Rim, Roman Rivera, and Modibo Sidibé. 2021. "Police Officer Assignment and Neighborhood Crime." Working Paper 29243. National Bureau of Economic Research.

Ba, Bocar, Dean Knox, Jonathan Mummolo, and Roman Rivera. 2021. "The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago." *Science* 371 (6530): 696–702.

Baker, Sara, Adalbert Mayer, and Steven L. Puller. 2011. "Do More Diverse Environments Increase the Diversity of Subsequent Interaction? Evidence from Random Dorm Assignment." *Economics Letters* 110 (2): 110–12.

Benjamin, Daniel J., James J. Choi, and A. Joshua Strickland. 2010. "Social Identity and Preferences." *The American Economic Review* 100 (4): 1913–28.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1): 289–300.

Bergé, Laurent R. 2018. "Efficient Estimation of Maximum Likelihood Models with Multiple FIxed-Effects: The r Package FENmlm." *CREA Discussion Papers*, no. 13: 39.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2013. "Under Pressure? The Effect of Peers on Outcomes of Young Adults." *Journal of Labor Economics* 31 (1): 119–53.

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity." *The American Economic Review* 96 (5): 1890–1905.

Bonhomme, Stephane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler. 2022. "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?" *Journal of Labor Economics*, March.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable Is Weak." *Journal of the American Statistical Association* 90 (430): 443–50.

Brenoe, Anne Ardila, and Ulf Zölitz. 2020. "Exposure to More Female Peers Widens the Gender Gap in STEM Participation." *Journal of Labor Economics* 38 (4): 1009–54.

Caeyers, Bet, and Marcel Fafchamps. 2016. "Exclusion Bias in the Estimation of Peer Effects." Working Paper 22565. National Bureau of Economic Research.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *The Quarterly Journal of Economics* 128 (3): 967–1015.

Carrell, Scott E., Mark Hoekstra, and Elira Kuka. 2018. "The Long-Run Effects of Disruptive Peers." *American Economic Review* 108 (11): 3377–3415.

Carrell, Scott E., Mark Hoekstra, and James E. West. 2019. "The Impact of College Diversity on Behavior Toward Minorities." *American Economic Journal: Economic Policy* 11 (4): 159–82.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81 (3): 855–82.

Chang, Edward H., Katherine L. Milkman, Dena M. Gromet, Robert W. Rebele, Cade Massey, Angela L. Duckworth, and Adam M. Grant. 2019. "The Mixed Effects of Online Diversity Training." *Proceedings of the National Academy of Sciences* 116 (16): 7778–83.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *The Quarterly Journal of Economics* 126 (4): 1593–1660.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers i: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–2632.

Chicago_mwk. 2010. "Chicago Police Academy 2010. Police Forums & Law Enforcement Forums @ Officer.com." January 7, 2010.

Cho, Sungwoo, Felipe Goncalves, and Emily Weisburst. 2021. "Do Police Make Too Many Arrests?" Working Paper.

Cox, Robynn, Jamein P. Cunningham, and Alberto Ortega. 2021. "The Impact of Affirmative

Action Litigation on Police Killings of Civilians." Working Paper.

CPD. 2011. "Personnel Transfer and Assignment Procedures – (FOP)." December 15, 2011.

———. 2017. "CPD 2017 FAQ." 2017.

———. 2018. "Field Training and Evaluation Program." June 5, 2018.

———. 2020. "Education and Training Division (ETD) |Chicago Police Department." 2020.

Csardi, Gabor, and Tamas Nepusz. 2005. "The Igraph Software Package for Complex Network Research." *InterJournal* Complex Systems (November): 1695.

Dahl, Gordon B, Andreas Kotsadam, and Dan-Olof Rooth. 2021. "Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams." *The Quarterly Journal of Economics* 136 (2): 987–1030.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108 (2): 201–40.

Dobbin, Frank, and Alexandra Kalev. 2016. "Why Diversity Programs Fail." *Harvard Business Review* 94 (7).

DOJ. 2016. "Advancing Diversity in Law Enforcement Report (October 2016)." Report. The United States Department of Justice: DOJ Civil Rights Division; Equal Employment Opportunity Commission.

Donohue, John J., and Steven D. Levitt. 2001. "The Impact of Legalized Abortion on Crime." *The Quarterly Journal of Economics* 116 (2): 379–420.

Feld, Jan, and Ulf Zölitz. 2017. "Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects." *Journal of Labor Economics*, January.

feredeathpsn. 2017. "Chicago Police Lottery Number. R/Police." June 29, 2017.

Ferguson, Joseph M, and Deborah Witzburg. 2021. "EVALUATION OF THE DEMOGRAPHIC IMPACTS OF THE CHICAGO POLICE DEPARTMENT's HIRING PROCESS." Report. City of Chicago: Office of the Inspector General.

Fisher, R. A. 1925. "Theory of Statistical Estimation." *Mathematical Proceedings of the Cambridge Philosophical Society* 22 (5): 700–725.

Fryer, Roland G., and Paul Torelli. 2010. "An Empirical Analysis of 'Acting White'." *Journal of Public Economics* 94 (5): 380–96.

Garner, Maryah, Anna Harvey, and Hunter Johnson. 2019. "Estimating Effects of Affirmative Action in Policing: A Replication and Extension." *International Review of Law and Economics*, November, 105881.

Golsteyn, Bart H. H., Arjan Non, and Ulf Zölitz. 2021. "The Impact of Peer Personality on Academic Achievement." *Journal of Political Economy* 129 (4): 1052–99.

Goncalves, Felipe, and Steven Mello. 2021. "A Few Bad Apples? Racial Bias in Policing."

American Economic Review 111 (5): 1406–41.

Gould, Eric D., Victor Lavy, and M. Daniele Paserman. 2009. "Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence." *The Economic Journal* 119 (540): 1243–69.

Gupta, Arpit, Christopher Hansman, and Ethan Frenchman. 2016. "The Heavy Costs of High Bail: Evidence from Judge Randomization." *Journal of Legal Studies* 45 (2): 471–505.

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo. 2009. "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments." *American Economic Journal: Applied Economics* 1 (4): 34–68.

Harvey, Anna, and Taylor Mattia. 2024. "Reducing Racial Disparities in Crime Victimization: Evidence from Employment Discrimination Litigation." *Journal of Urban Economics*, RACE, SOCIAL JUSTICE, & CITIES, 141 (May): 103459.

Hoekstra, Mark, and CarlyWill Sloan. 2022. "Does Race Matter for Police Use of Force? Evidence from 911 Calls." *American Economic Review* 112 (3): 827–60.

Holden, Richard, Michael Keane, and Matthew Lilley. 2021. "Peer Effects on the United States Supreme Court." *Quantitative Economics* 12 (3): 981–1019.

Holm, Sture. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics* 6 (2): 65–70.

Holz, Justin E., Roman Rivera, and Bocar A. Ba. 2023. "Peer Effects in Police Use of Force." *American Economic Journal: Economic Policy* 15 (2): 256–91.

Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." Working Paper 7867. National Bureau of Economic Research.

Kass, John, and Robert Blau. 1991. "POLICE HIRING LOTTERY LATEST DALEY HEADACHE." *Chicago Tribune*, August, 3.

Kline, Patrick, Evan K Rose, and Christopher R Walters. 2022. "SYSTEMIC DISCRIMINATION AMONG LARGE u.s. EMPLOYERS." *THE QUARTERLY JOURNAL OF ECONOMICS*, 121.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-out Estimation of Variance Components." *Econometrica* 88 (5): 1859–98.

Krueger, Alan B., and Lawrence H. Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica* 56 (2): 259.

Lavy, Victor, and Analía Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics* 3 (2): 1–33.

Leifeld, Philip. 2013. "Texreg: Conversion of Statistical Model Output in r to LATEX and HTML Tables." *Journal of Statistical Software* 55 (November): 1–24.

Linos, Elizabeth. 2018. "More Than Public Service: A Field Experiment on Job Advertisements and Diversity in the Police." *Journal of Public Administration Research and Theory* 28 (1): 67–85.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60 (3): 531–42.

McCrary, Justin. 2007. "The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police." *The American Economic Review* 97 (1): 318–53.

Michelman, Valerie, Joseph Price, and Seth D Zimmerman. 2021. "Old Boys' Clubs and Upward Mobility Among the Educational Elite." *The Quarterly Journal of Economics*, December.

Miller, Amalia R, and Carmit Segal. 2018. "Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence." *The Review of Economic Studies* 86 (5): 2220–47.

neverlose357. 2010. "2011 Chicago Police Academy. Police Forums & Law Enforcement Forums @ Officer.com." December 4, 2010.

Newman, M. E. J., and M. Girvan. 2004. "Finding and Evaluating Community Structure in Networks." *Physical Review E* 69 (2).

Nicholson-Crotty, Sean, Jill Nicholson-Crotty, and Sergio Fernandez. 2017. "Will More Black Cops Matter? Officer Race and Police-Involved Homicides of Black Citizens." *Public Administration Review* 77 (2): 206–16.

Owens, Emily, David Weisburd, Karen L. Amendola, and Geoffrey P. Alpert. 2018. "Can You Build a Better Cop?" *Criminology & Public Policy* 17 (1): 41–87.

Pritchard, Paige. 2013. "Do You Have What It Takes to Join the Chicago Police Department?" *Chicago Magazine*, August.

Ridgeway, Greg. 2020. "The Role of Individual Officer Characteristics in Police Shootings." *The ANNALS of the American Academy of Political and Social Science* 687 (1): 58–66.

Rivera, Roman. 2024. "Replication Data for: Do Peers Matter in the Police Academy?" American Economic Association[publisher]; Inter-university Consortium for Political; Social Research[distributor].

Roth, Jonathan, and Pedro H. C. Sant'Anna. 2021. "Efficient Estimation for Staggered Rollout Designs." Working Paper. https://arxiv.org/abs/2102.01291.

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *The Quarterly Journal of Economics* 125 (1): 175–214.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *The Quarterly Journal of Economics* 116 (2): 681–704.

———. 2011. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In *Handbook of the Economics of Education*, 3:249–77. Elsevier.

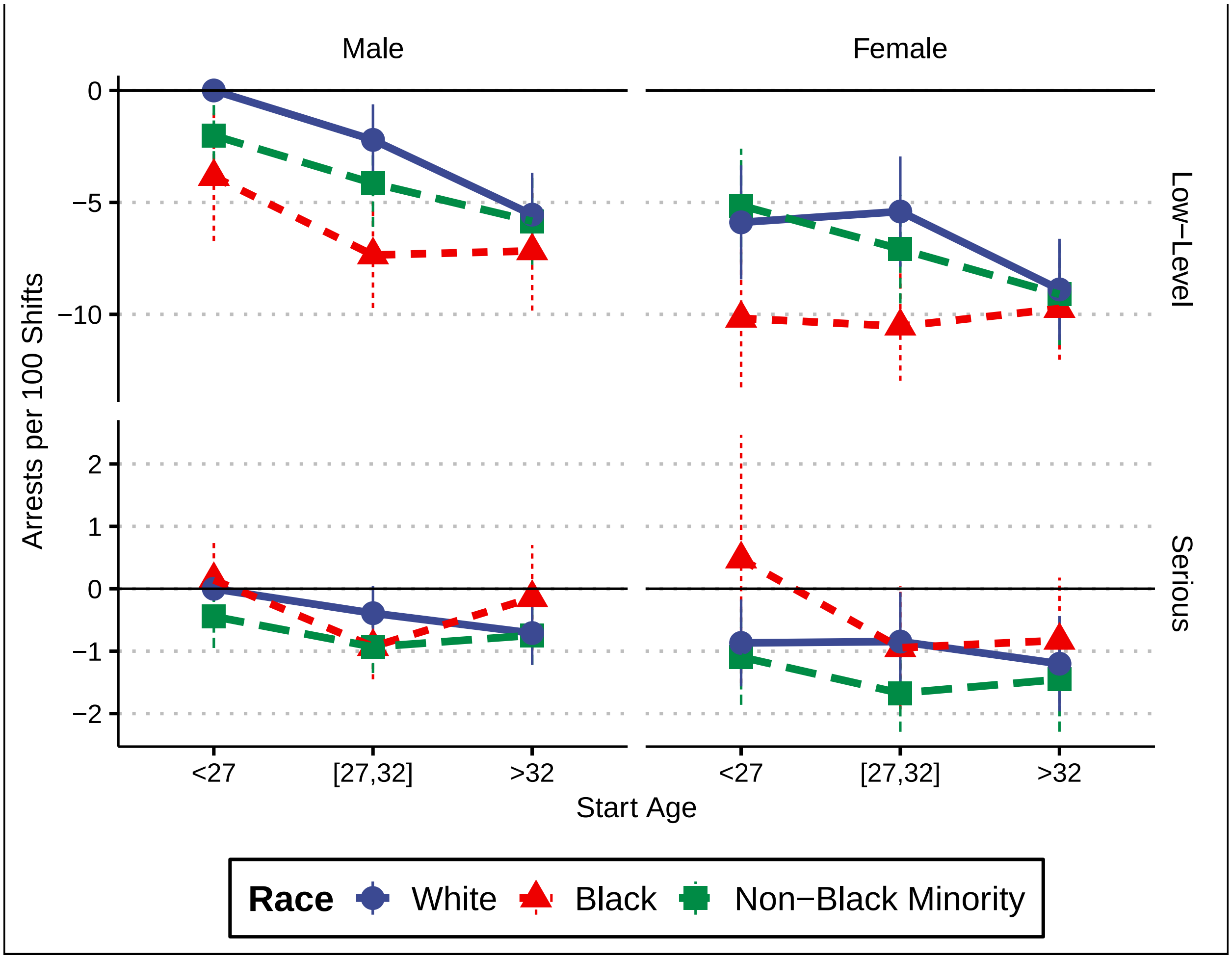
Schindler, David, and Mark Westcott. 2021. “Shocking Racial Attitudes: Black g.i.s in Europe.” *The Review of Economic Studies* 88 (1): 489–520.

Stevenson, Megan. 2018. “Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes.” *The Journal of Law, Economics, and Organization* 34 (4): 511–42.

Weisburst, Emily K. 2022. “‘Whose Help Is on the Way?’: The Importance of Individual Police Officers in Law Enforcement Outcomes.” *Journal of Human Resources*, March.

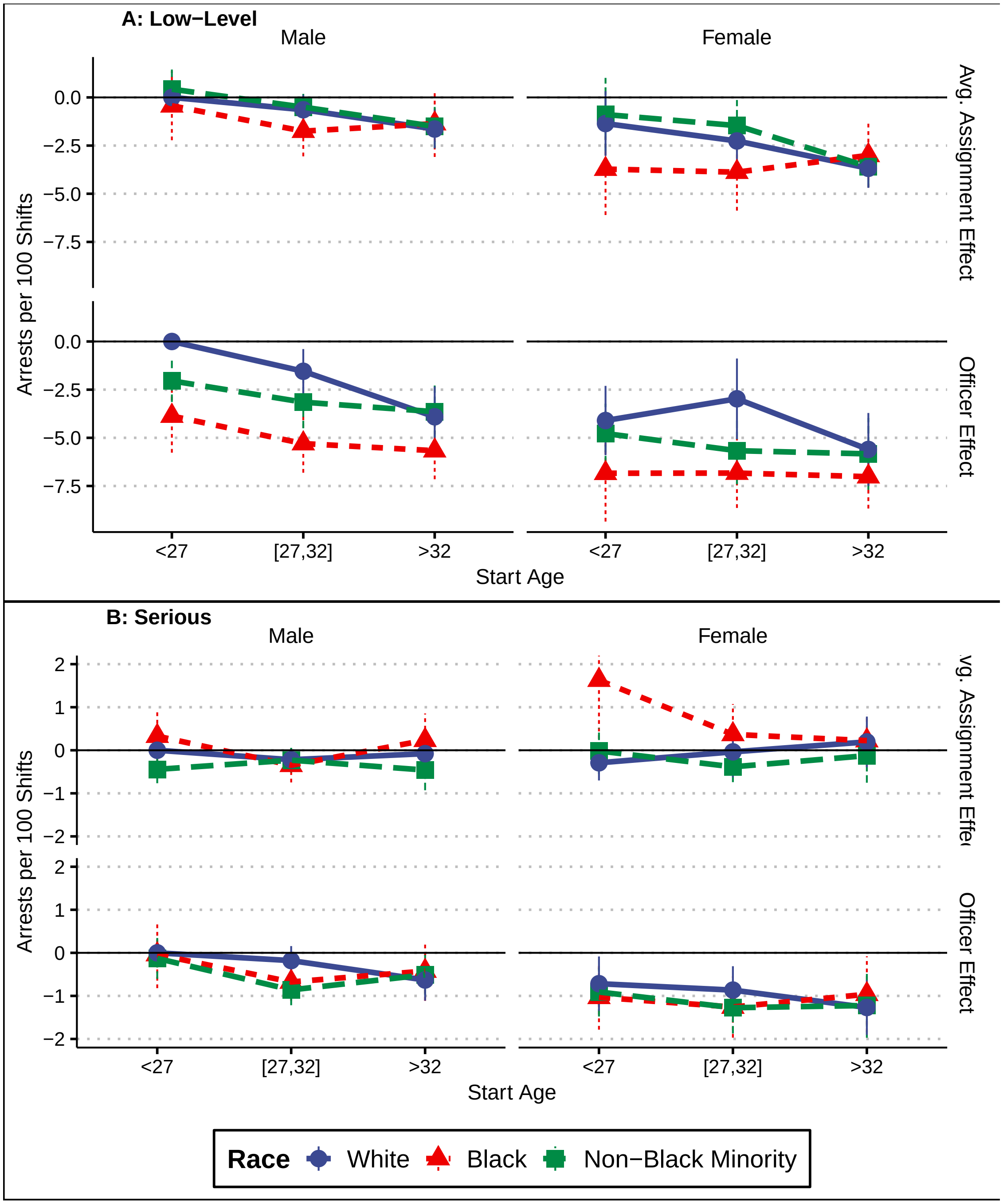
West, Jeremy. 2018. “Racial Bias in Police Investigations.” Working Paper.

Figure 1: Effect of Race, Gender, and Age on Arrests



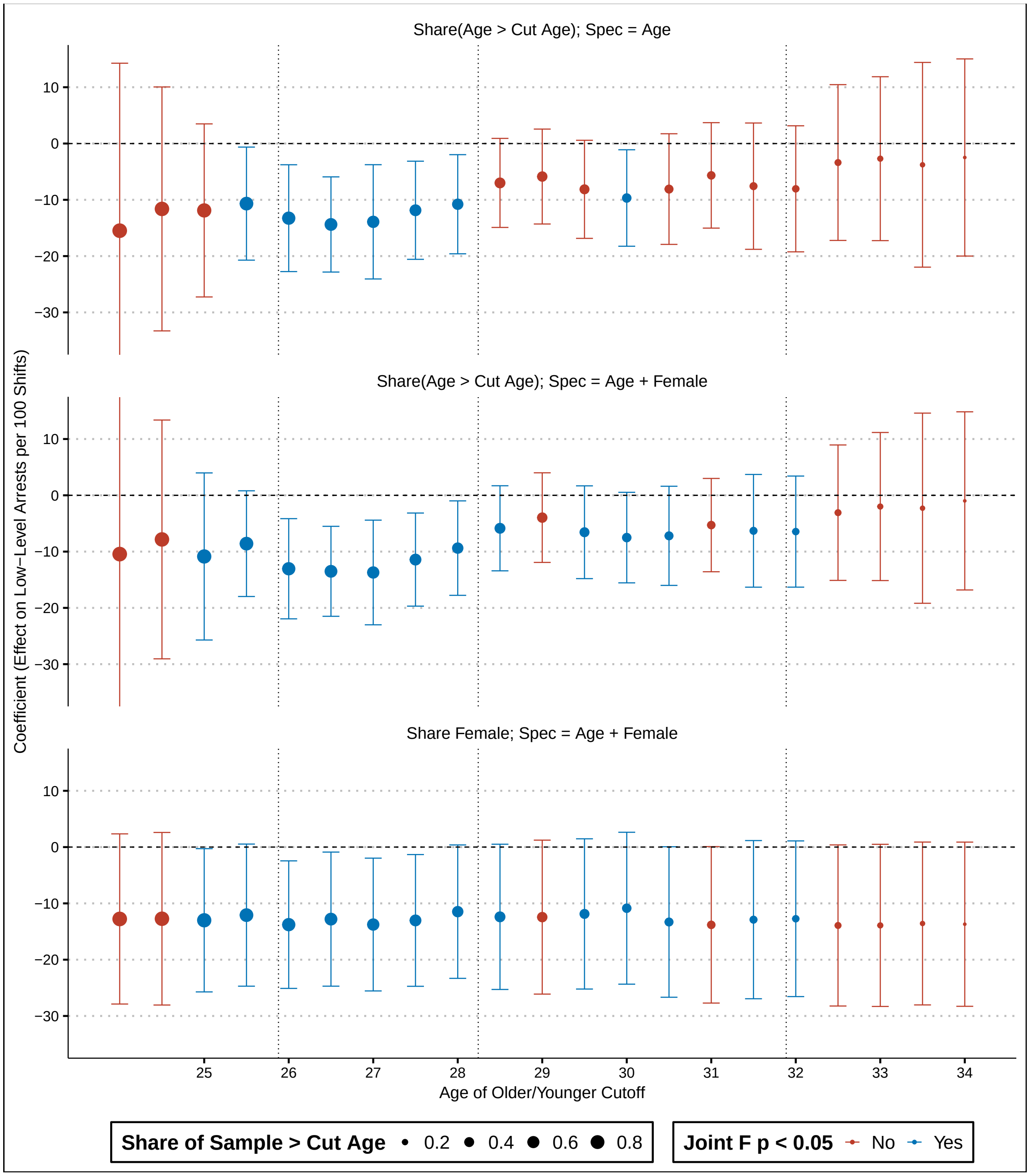
Note: Figure displays the coefficients of interacted race, gender, and age bins recovered by regression the analysis sample officers' average low-level and serious arrests on these bins, with white male officers starting at < 27 years old as the excluded group and including exam period fixed effects, with 95% confidence intervals computed from standard errors clustered at the cohort level.

Figure 2: Effect of Race, Gender, and Age on Average Assignment Effects and Officer Effects



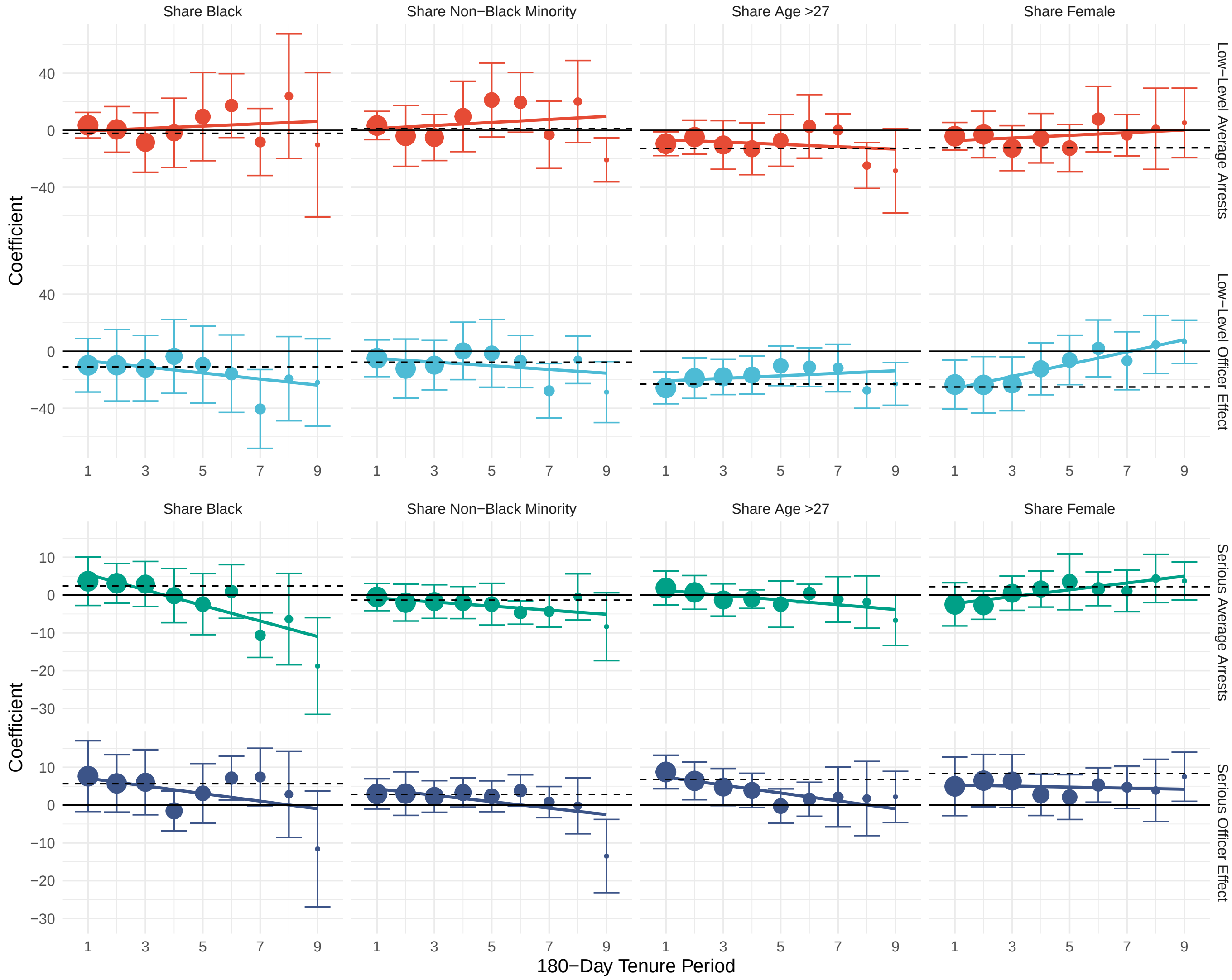
Note: Figure displays the coefficients of interacted race, gender, and age bins recovered by regressing those bins on analysis sample officers' average assignment effects and officer effects for both low-level and serious arrests recovered from estimating equation (3), with white male officers starting at < 27 years old as the excluded group and including exam period fixed effects, with 95% confidence intervals computed from standard errors clustered at the cohort level. The upper limit on the y-axis for Panel B is truncated at 2 to increase readability.

Figure 3: Effect of Peer Age and Peer Gender under Alternate Age Cutoffs



Note: Figure displays the coefficients of cohort share age > cut off and share female recovered estimating equation (1) as the cut age moves from 24 to 34, by regressing average low-level arrests on cohort share age > cut age (top panel) and cohort share age > cut age and share female (middle and bottom panel). All specifications contain exam period fixed effects and officer-level indicators (e.g., for if the officer is over the cut off age and if they are female in the bottom two panels). The Joint F p-value for each regression indicates whether we reject or fail to reject the null hypothesis that all cohort share coefficients are jointly equal to zero. 95% confidence intervals computed from standard errors clustered at the cohort level. The limits of the y-axis are truncated to -35 to 15 for readability. Dashed vertical lines correspond to the 25th, 50th, and 75th percentile in start ages.

Figure 4: Persistence of Effects on Outcomes



Note: Figure displays the effects of cohort shares Black, non-Black minority, older (age > 27), and female peers on average arrests and officer fixed effects for low-level and serious arrests over 180-day tenure periods. Officers in the analysis sample were assigned unique IDs for each 180-day tenure period in their careers, and average arrests and officer effects (estimated using equation (3) with modified officer indicators) were recovered for each period, with at most 2,296 observations. Coefficients are the result of estimating equation (1) with the main specification on average arrests and officer effects separately during the respective tenure period. All regressions include exam period fixed effects and controls for officer-level characteristics as shown in Table 7. Dotted black lines denote the respective coefficient from the main results. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Error bars indicate 95% confidence intervals based on standard errors clustered at cohort level. Point-size is proportional to the share of officers included in the estimation of the coefficients for that period. Colored lines are linear trends in coefficients weighted by sample size.

Table 1: Summary Statistics by Exam

	Pooled	Cohorts	Pooled	Cohorts	Pooled	Cohorts	Pooled
	Exam 2006		Exam 2010		Exam 2013		All Officers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	0.803	0.751	0.807	0.8	0.767	0.778	0.756
Female	0.197	0.249	0.193	0.2	0.233	0.222	0.244
White	0.475	0.458	0.513	0.506	0.489	0.488	0.471
Minority	0.525	0.542	0.487	0.494	0.511	0.512	0.529
Black	0.234	0.241	0.132	0.13	0.125	0.121	0.227
Hispanic	0.258	0.271	0.312	0.325	0.341	0.35	0.261
Asian/Native American	0.0328	0.0299	0.0436	0.0396	0.045	0.0407	0.041
Birth Year	1980.33	1980.5	1982.66	1982.7	1987.21	1987.15	1975.23
Start Age	29.74	29.68	30.07	30.16	28.18	28.04	29.04
Cohort Size	86.57	77.5	61.11	54.05	83.65	73.71	-
N	244	4	940	21	1112	17	11397

Note: Table compares the average characteristics of Exam 2006 (Columns (1) and (2)), Exam 2010 (Columns (3) and (4)), Exam 2013 (Columns (5) and (6)) to all of the officers in the panel data (Column (7)). Odd columns contain pooled average characteristics over recruits within the exam period, while even columns contain average characteristics of cohorts, including those recruits that do not appear in the main analysis due to attrition.

Table 2: Summary Statistics of Outcomes by Exam

	Exam 2006	Exam 2010	Exam 2013
	(1)	(2)	(3)
Total Arrests	0.18 (0.48)	0.175 (0.45)	0.154 (0.42)
White Arrestees	0.009 (0.1)	0.008 (0.09)	0.009 (0.1)
Black Arrestees	0.145 (0.43)	0.144 (0.41)	0.118 (0.37)
Hispanic Arrestees	0.024 (0.18)	0.023 (0.17)	0.025 (0.17)
Serious Arrests	0.044 (0.23)	0.052 (0.25)	0.049 (0.24)
Low-Level Arrests	0.135 (0.42)	0.123 (0.38)	0.105 (0.35)
Guilty Arrests	0.04 (0.21)	0.034 (0.19)	0.02 (0.15)
N	210528	531646	298992
Unique Officers	244	940	1112

Note: Table presents summary statistics for average arresting outcomes of sample officers once they enter the police force by exam period within the assignment panel data. Standard deviations are reported in parentheses. ***p $<$ 0.01; **p $<$ 0.05; *p $<$ 0.1

Table 3: Multinomial Logit for Cohort Assignment

Exam	Controls	Multinomial Logit P-Value	N Recruits	N Cohorts
Pooled	Black, Non-Black Minority, Gender, Start Age	0.750	2698	42
Pooled	+ Military, Spanish, High Edu.	0.361	2698	42
2006	Black, Non-Black Minority, Gender, Start Age	0.322	310	4
2006	+ Military, Spanish, High Edu.	0.099	310	4
2010	Black, Non-Black Minority, Gender, Start Age	0.854	1135	21
2010	+ Military, Spanish, High Edu.	0.523	1135	21
2013	Black, Non-Black Minority, Gender, Start Age	0.280	1253	17
2013	+ Military, Spanish, High Edu.	0.143	1253	17

Note: Table reports the p-value of the joint F-test on the coefficients of a multinomial logit regressing assigned cohort on officer characteristics for each exam period for two sets of controls. The limited controls include the officer being Black or not Black minority, start age, and gender; the second set of controls adds if they were in the military, if they speak Spanish, and if they have a Bachelors degree or higher. Pooled sample uses a base model of only exam fixed effects.

Table 4: Balance Regressions

	Cohort Mean Age	Cohort Share Male	Cohort Share Minority	Start Date (Years)
	(1)	(2)	(3)	(4)
Black	0.0338	-0.0029	-0.0025	0.054
	(0.0327)	(0.0025)	(0.0035)	(0.0401)
Non-Black Minority	0.0164	0.0018	-0.0055**	0.016
	(0.0168)	(0.0013)	(0.0026)	(0.0218)
Male	0.0079	-0.003	0.0015	-0.0416
	(0.0196)	(0.0032)	(0.0022)	(0.0262)
Start Age	0.0018	0	0.0005*	0.0093***
	(0.0025)	(0.0002)	(0.0003)	(0.0028)
Military	-0.0074	0.0093	-0.0037	-0.0151
	(0.0464)	(0.006)	(0.0058)	(0.0488)
Spanish-Speaking	0.0209	0.0025	0.0075	0.0145
	(0.0487)	(0.0035)	(0.0051)	(0.0622)
High Edu	-0.0371	0.0041*	-0.0029	-0.0191
	(0.0256)	(0.0022)	(0.0027)	(0.0321)
Exam 2010	0.4035	0.0314	-0.0483***	2.7013***
	(0.2557)	(0.0387)	(0.0141)	(0.2833)
Exam 2013	-1.5042***	-0.0001	-0.026	5.3238***
	(0.2657)	(0.0392)	(0.0173)	(0.3238)
(Intercept)	29.5954***	0.7586***	0.5333***	1.3305***
	(0.2307)	(0.0447)	(0.012)	(0.2687)
N	2698	2698	2698	2698
R2	0.731	0.107	0.07	0.888

Note: Table displays results for balance regression tests. Each column displays the coefficients of officer characteristics on their cohort composition (Columns (1)-(3)) and the cohort start date (Column (4)). Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. ***p < 0.01; **p < 0.05; *p < 0.1

Table 5: Reduced Form Effect of Cohorts on Arrests

	P-Value	SD(Partialled Y)	SD(Cohort-Effect) Raw	SD(Cohort-Effect) Bias Corrected (KRW)	Cohort-Effect Share of SD(Partialled Y) Raw	Cohort-Effect Share of SD(Partialled Y) Bias Corrected (KRW)
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Low-Level Arrests						
Pooled	<0.01	0.109	0.027	0.026	0.252	0.242
Exam 2006	0.91	0.122	0.006	-	0.047	-
Exam 2010	0.014	0.111	0.023	0.02	0.211	0.176
Exam 2013	<0.01	0.1	0.022	0.021	0.224	0.209
Mean Serious Arrests						
Pooled	<0.01	0.029	0.005	0.00022	0.166	0.008
Exam 2006	0.997	0.024	0.00037	-	0.015	-
Exam 2010	<0.01	0.028	0.006	-	0.203	-
Exam 2013	<0.01	0.03	0.005	0.003	0.149	0.099

Note: Table presents reduced form relationship between assigned cohort and future arresting behavior. Cohort-specific effects are recovered by estimating equation (2) with the dependent variable being partialled-out average officer arrests per shift after regressing them on officer race, gender, education, spanish-speaking, and age, cohort fixed effects, and exam period fixed effects (for pooled sample); then, the partialled outcome variable is constructed by subtracting fitted values on all officer observables and period effects, but not the cohort. Column (1) contains p-values of a joint F-test for cohort-specific effects being jointly zero, using robust standard errors. Column (2) contains the standard deviation of partialled out arrests, and Column (3) displays the standard deviation of cohort-specific effects — all are weighted by sample size. Column (4) displays bias-corrected standard deviations using the bias correction procedure from Kline, Rose, and Walters (2022) which corrects for sampling error which biases the standard deviations of the cohort effects. Column (5) and (6) display the ratio of standard deviation in cohort effects divided by the standard deviation in the partialled outcome, or the share of 1 SD in the outcome explained by cohort effects, with Column (5) using raw and Column (6) using bias corrected standard deviations of cohort effects.

Table 6: Effect of Peer Composition on Arrests

	(1)	(2)	(3)	(4)	(5)	(6)
A: Average Low-Level Arrests per 100 Shift						
Share Black	-0.119					-0.021
	(0.081)					(0.072)
	[p=0.15, q=0.299]					[p=0.77, q=1]
Share Non-Black Minority	-0.002					0.012
	(0.074)					(0.07)
	[p=0.976, q=0.976]					[p=0.864, q=1]
Share Female		-0.137*			-0.138**	-0.123**
		(0.074)			(0.06)	(0.06)
		[p=0.071, q=0.071]			[p=0.027, q=0.027]	[p=0.046, q=0.138]
Share Age >28.24			-0.088**			
			(0.041)			
			[p=0.039, q=0.039]			
Share Age >27				-0.139**	-0.137***	-0.128**
				(0.052)	(0.047)	(0.051)
				[p=0.01, q=0.01]	[p=0.006, q=0.012]	[p=0.017, q=0.066]
Adj. R2	0.023	0.029	0.025	0.029	0.053	0.064
Joint F p-value	0.282	0.064	0.033	0.007	<0.001	0.011
N	2296	2296	2296	2296	2296	2296
B: Average Serious Arrests per 100 Shift						
Share Black	0.03*					0.024
	(0.016)					(0.015)
	[p=0.068, q=0.135]					[p=0.13, q=0.52]
Share Non-Black Minority	-0.014					-0.013
	(0.01)					(0.011)
	[p=0.175, q=0.175]					[p=0.233, q=0.52]
Share Female		0.034**			0.033**	0.022
		(0.015)			(0.015)	(0.015)
		[p=0.033, q=0.033]			[p=0.036, q=0.072]	[p=0.154, q=0.52]
Share Age >28.24			-0.008			
			(0.011)			
			[p=0.438, q=0.438]			
Share Age >27				0.003	0.003	0.001
				(0.012)	(0.011)	(0.012)
				[p=0.783, q=0.783]	[p=0.815, q=0.815]	[p=0.946, q=0.946]
Adj. R2	0.008	0.011	0.003	0.007	0.016	0.021
Joint F p-value	0.015	0.027	0.434	0.782	0.092	0.03
N	2296	2296	2296	2296	2296	2296

Note: Table displays the OLS result for the effect of cohort composition on sample officers' average arrests per 100 shifts from estimating equation (1), with Panel A having low-level arrests and Panel B having serious arrests as the outcomes. All regressions include controls for officer-level indicators for group membership for all peer characteristics included (e.g., officer race being Black and officer race being non-Black minority in Column (1)), and all regressions include exam period fixed effects. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Joint F p-value refers to the p-value from a joint F-test with the null hypothesis being that all cohort share coefficients are jointly zero. Adj. R2 refers to the adjusted R-squared beyond the exam period effects (i.e, 'within' adjusted R-squared). Standard errors clustered at cohort level are in parentheses. P-values and adjusted p-values ('q') using the Holm (1979) correction for multiple hypotheses over cohort share variables are displayed in brackets. ***p < 0.01; **p < 0.05; *p < 0.1

Table 7: Effect of Peer Composition on Assignments and Behavior

	Average Assigned District (Std.)				Average Assignment Effect (Arrests per 100 Shifts)		Officer Effect (Arrests per 100 Shifts)	
	Violent	Property	Nonindex	Population Share Black	Low-Level	Serious	Low-Level	Serious
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	0.113*	0.256***	0.024	0.457***	-0.648	0.296*	-2.975***	-0.211
	(0.066)	(0.087)	(0.075)	(0.066)	(0.401)	(0.152)	(0.452)	(0.16)
Non-Black Minority	0.001	0.017	-0.051	-0.012	0.308	-0.237**	-1.406***	-0.279**
	(0.051)	(0.05)	(0.052)	(0.049)	(0.271)	(0.104)	(0.367)	(0.114)
Female	-0.026	0.018	-0.014	-0.002	-1.625***	0.144	-2.559***	-0.622***
	(0.047)	(0.054)	(0.05)	(0.044)	(0.325)	(0.103)	(0.366)	(0.134)
Age >27	-0.11**	-0.028	-0.099**	-0.137***	-1.131***	-0.119	-1.766***	-0.382***
	(0.048)	(0.039)	(0.047)	(0.044)	(0.257)	(0.087)	(0.319)	(0.103)
Share Black	-0.092	0.041	-0.277	-0.524	4.717	-1.952	-10.909	5.652*
	(0.988)	(1.039)	(0.972)	(0.796)	(4.557)	(2.585)	(12.592)	(2.975)
	[p=0.926, q=0.926]	[p=0.968, q=0.968]	[p=0.777, q=1]	[p=0.514, q=1]	[p=0.307, q=0.92]	[p=0.454, q=0.454]	[p=0.391, q=0.737]	[p=0.065, q=0.129]
Share Non-Black Minority	0.702	-0.895	0.183	-0.126	8.002*	-3.893**	-7.681	2.83
	(0.718)	(0.805)	(0.737)	(0.589)	(4.102)	(1.809)	(8.447)	(1.996)
	[p=0.334, q=0.812]	[p=0.273, q=0.818]	[p=0.805, q=1]	[p=0.832, q=1]	[p=0.058, q=0.232]	[p=0.037, q=0.145]	[p=0.369, q=0.737]	[p=0.164, q=0.164]
Share Female	0.91	1.574**	0.885	0.049	3.676	-3.235	-25.06**	8.364***
	(0.815)	(0.761)	(0.866)	(0.785)	(4.13)	(2.378)	(10.501)	(2.742)
	[p=0.271, q=0.812]	[p=0.045, q=0.18]	[p=0.313, q=1]	[p=0.951, q=1]	[p=0.379, q=0.92]	[p=0.181, q=0.362]	[p=0.022, q=0.065]	[p=0.004, q=0.012]
Share Age >27	-1.179	0.695	-0.919	-1.071	2.878	-4.332**	-23.008***	6.769***
	(0.799)	(0.652)	(0.824)	(0.683)	(4.422)	(1.999)	(6.915)	(1.927)
	[p=0.148, q=0.591]	[p=0.292, q=0.818]	[p=0.271, q=1]	[p=0.124, q=0.497]	[p=0.519, q=0.92]	[p=0.036, q=0.145]	[p=0.002, q=0.007]	[p=0.001, q=0.004]
Joint F p-value (Peer Composition)	0.404	0.076	0.705	0.32	0.052	<0.001	<0.001	<0.001
N	2291	2291	2291	2291	2296	2296	2296	2296

Note: Table displays the OLS result for the effect of cohort composition on sample officers' average assigned district characteristics (standardized) (not avaliable for 5 officers), average assignment and officer effects (arrests per 100 shifts) as dependent variables from estimating equation (1). All regressions include exam period fixed effects. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. Joint F p-value refers to the p-value from a joint F-test with the null hypothesis being that all cohort share coefficients are jointly zero. P-values and adjusted p-values ('q') using the Holm (1979) correction for multiple hypotheses over cohort share variables are displayed in brackets for cohort share variables. ***p < 0.01; **p < 0.05; *p < 0.1

Table 8: Interaction Effects

	Low-Level		Serious	
	Average Arrests	Officer Effects	Average Arrests	Officer Effects
	(1)	(2)	(3)	(4)
A: Pooled Minorities				
Share Minority and Age >27	-14**	-32.3***	0.3	10.5***
	(6.5)	(6.2)	(1.2)	(2)
	[p=0.037, q=0.11]	[p<0.001, q<0.001]	[p=0.837, q=1]	[p<0.001, q<0.001]
Share White and Age >27	-9.6	-22.1***	-0.1	7.1***
	(6.5)	(7.2)	(1.3)	(1.9)
	[p=0.147, q=0.295]	[p=0.004, q=0.011]	[p=0.911, q=1]	[p<0.001, q=0.001]
Share Minority and Age <27	10.1	-5.1	-1.9	3.9
	(8.1)	(10.3)	(2.3)	(3.1)
	[p=0.218, q=0.295]	[p=0.624, q=0.624]	[p=0.398, q=1]	[p=0.22, q=0.22]
Share Female	-15.1**	-26.8**	3.5**	9.1***
	(6.4)	(10.2)	(1.5)	(2.6)
	[p=0.022, q=0.088]	[p=0.012, q=0.025]	[p=0.028, q=0.11]	[p=0.001, q=0.002]
Joint F p-value	0.001	<0.001	0.209	<0.001
B: Black and Non-Black Minorities				
Share Black and Age >27	-21.4**	-34.7***	1.5	11.3***
	(10.1)	(10.3)	(2)	(3.5)
	[p=0.041, q=0.206]	[p=0.002, q=0.009]	[p=0.471, q=1]	[p=0.003, q=0.013]
Share Non-Black Minority and Age >27	-11.8*	-31***	-0.9	9.7***
	(6.7)	(8.3)	(1.2)	(2.3)
	[p=0.087, q=0.346]	[p<0.001, q=0.004]	[p=0.481, q=1]	[p<0.001, q<0.001]
Share White and Age >27	-6	-21.4**	0.2	7.6***
	(6.5)	(9.5)	(1.4)	(2.5)
	[p=0.363, q=0.727]	[p=0.031, q=0.093]	[p=0.912, q=1]	[p=0.004, q=0.015]
Share Black and Age <27	22.7	-5.8	3.2	9
	(18.9)	(32)	(3.1)	(7)
	[p=0.237, q=0.71]	[p=0.856, q=1]	[p=0.303, q=1]	[p=0.208, q=0.415]
Share Non-Black Minority and Age <27	4.9	-5.2	-3.5	2.4
	(9.6)	(12.6)	(2.4)	(3.9)
	[p=0.611, q=0.727]	[p=0.682, q=1]	[p=0.152, q=0.759]	[p=0.548, q=0.548]
Share Female	-14**	-25.6**	2.4	8.3***
	(6.2)	(10.8)	(1.5)	(2.8)
	[p=0.031, q=0.185]	[p=0.023, q=0.093]	[p=0.125, q=0.75]	[p=0.005, q=0.015]
Joint F p-value	0.011	<0.001	0.019	<0.001
C: Minority and White Female				
Share Minority and Age >27	-13.9**	-31.8***	0.2	10.3***
	(6.7)	(6.2)	(1.2)	(1.9)
	[p=0.042, q=0.186]	[p<0.001, q<0.001]	[p=0.867, q=1]	[p<0.001, q<0.001]
Share White and Age >27	-9.7	-22.6***	-0.1	7.3***
	(6.5)	(7.3)	(1.3)	(1.9)
	[p=0.142, q=0.427]	[p=0.003, q=0.014]	[p=0.936, q=1]	[p<0.001, q<0.001]
Share Minority and Age <27	9.6	-7.2	-1.7	4.8
	(7.9)	(10.9)	(2.2)	(3.1)
	[p=0.229, q=0.458]	[p=0.51, q=0.852]	[p=0.425, q=1]	[p=0.137, q=0.273]
Share Minority and Female	-16.4**	-32**	4**	11.2***
	(7.6)	(12)	(1.7)	(2.8)
	[p=0.037, q=0.186]	[p=0.011, q=0.033]	[p=0.025, q=0.124]	[p<0.001, q<0.001]
Share White and Female	-12.3	-14.5	2.4	3.9
	(11.7)	(18)	(2.6)	(5)
	[p=0.3, q=0.458]	[p=0.426, q=0.852]	[p=0.375, q=1]	[p=0.433, q=0.433]
Joint F p-value	0.003	<0.001	0.293	<0.001

Note: Table displays the OLS result for the effect of cohort composition on sample officers' average arrests and estimated officer effects, in units of arrests per 100 shifts using equation (1), with 2,296 observations. All regressions include exam period fixed effects and controls for officer-level characteristics corresponding to the included peer characteristics (e.g. an officer being a minority starting before age 27 in Panel A). Even columns contain interaction terms between an officer's race being white and cohort composition and additionally control for an officer being white. Officer effects are recovered from estimating equation (3). Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table 9: Effect of Predicted Peer Preferences

	Arrests per Shift			
	Average Arrests		Officer Effects	
	(1)	(2)	(3)	(4)
A: Low-Level Arrests				
Pred Low-Level Arrests	0.956***		0.655***	
	(0.094)		(0.062)	
Cohort Mean (Pred Low-Level Arrests) (1 SD)	0.009**		0.019***	
	(0.004)		(0.004)	
Pred Low-Level Officer Effect		1.491***		1.025***
		(0.148)		(0.099)
Cohort Mean (Pred Low-Level Officer Effect) (1 SD)		0.007*		0.015***
		(0.004)		(0.005)
Adj. R2	0.061	0.059	0.096	0.085
B: Serious Arrests				
Pred Serious Arrests	0.603***		0.494***	
	(0.133)		(0.123)	
Cohort Mean (Pred Serious Arrests) (1 SD)	-0.001		-0.004**	
	(0.001)		(0.002)	
Pred Serious Officer Effect		0.619***		0.578***
		(0.133)		(0.114)
Cohort Mean (Pred Serious Officer Effect) (1 SD)		-0.001*		-0.005***
		(0.001)		(0.002)
Adj. R2	0.009	0.011	0.026	0.036

Note: Table displays the results of regressing officer average arrests and officer effects for low-level (Panel A) and serious (Panel B) offenses on the officer's predicted arrests and officer effects and their cohort's mean predicted arrests and officer effects. Cohort means are standardized such that the coefficient corresponds to a 1 standard deviation increase. All specifications include exam period fixed effects. Predicted outcomes for each officer are computed by removing the officer's entire cohort from the sample, then estimating equation (2) with officer observables being the interaction between race, age, and gender bins as in Figure 1; predicted values for each officer in the excluded cohort are computed by using their race, age, and gender and the coefficients from the estimation; then each officer's cohort's mean predicted value is computed as a leave-one-out mean which excludes the officer themselves. Adj. R2 refers to the adjusted R-squared beyond the exam period effects (i.e, 'within' adjusted R-squared). Standard errors clustered at cohort level are in parentheses. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

A Appendix A

A.1 Entrance into the CPD Police Academy

To become an officer in the CPD, applicants must first meet multiple qualifications before applying to take the entrance exam. For example, by the time of starting at the academy, one must be a US citizen, a resident of Chicago, have sufficient credit hours at a college or university, and meet the age requirement (Pritchard (2013)). Potential applicants meeting these qualifications can apply to take the CPD entrance exam, and they will be notified of the test date and location after the application period ends (CPD (2017)).[14]

Applicants who pass the written exam are then assigned a random lottery number indicating the order in which they will be called into the academy. Random assignment to the academy was not always the case; it was introduced in the early 1990's in an attempt to increase diversity (Kass and Blau (1991)). After an applicant's number is drawn, they must pass a background check, drug screening, and medical, psychological and physical exams (Pritchard (2013)). Upon passing these requirements, potential officers are admitted into the academy.

There are usually tests once every 2 or 3 years (not including makeup exams)—but in 2006 there were four exams issued (one is labeled a '2005' exam in Figure B.1, but it took place in February 2006.) Generally, thousands of people take the CPD's written exam and a large portion of them meet the minimum passing score (see Figure B.1). Given the large number of passing applicants, many never have their numbers called before the applicant list is retired. Despite my best efforts, I have been unable to obtain any indication of when the applicant lists are retired (according to the CPD such documentation may not even exist). Also, applicants from a test are likely to be admitted possibly years after they took the test initially, and their entrance into the academy likely occurs while more applicants are taking a new test. This makes identifying which cohorts come from which tests (i.e., the pool from which officers are randomly assigned) difficult.

To the best of my knowledge, the Exam 2010 (July 2012 to May 2014) cohorts are an exception, and these cohorts all came from the same exam issued in December 2010 (see Figure B.1). The December 2010 exam was the last exam issued before the December 2013 exam. The only sizable cohort to enter in 2011 was on October 17, 2011, then about 8 months pass until the first sizable cohort of 2012 started on July 2, 2012. Following this,

[14]As late as the 2013 exam, veterans began to receive preference in their lottery numbers– though this is not well defined in the documentation. However, this preference is unlikely to be important considering almost all (over 95%) of recruits have military experience in the sample. This large number of veterans is consistent with more recent estimates from the Office of the Inspector General (Ferguson and Witzburg (2021)).

there were a total of 7 sizable cohorts starting between July and December 2012. Then, there is a continuous intake of cohorts until May 2014, when there is a three month gap until the next cohort. Given that it takes time for the CPD to draw in passing recruits and give them their multiple examinations, I believe these cohorts were all drawn from the December 2010 exam.

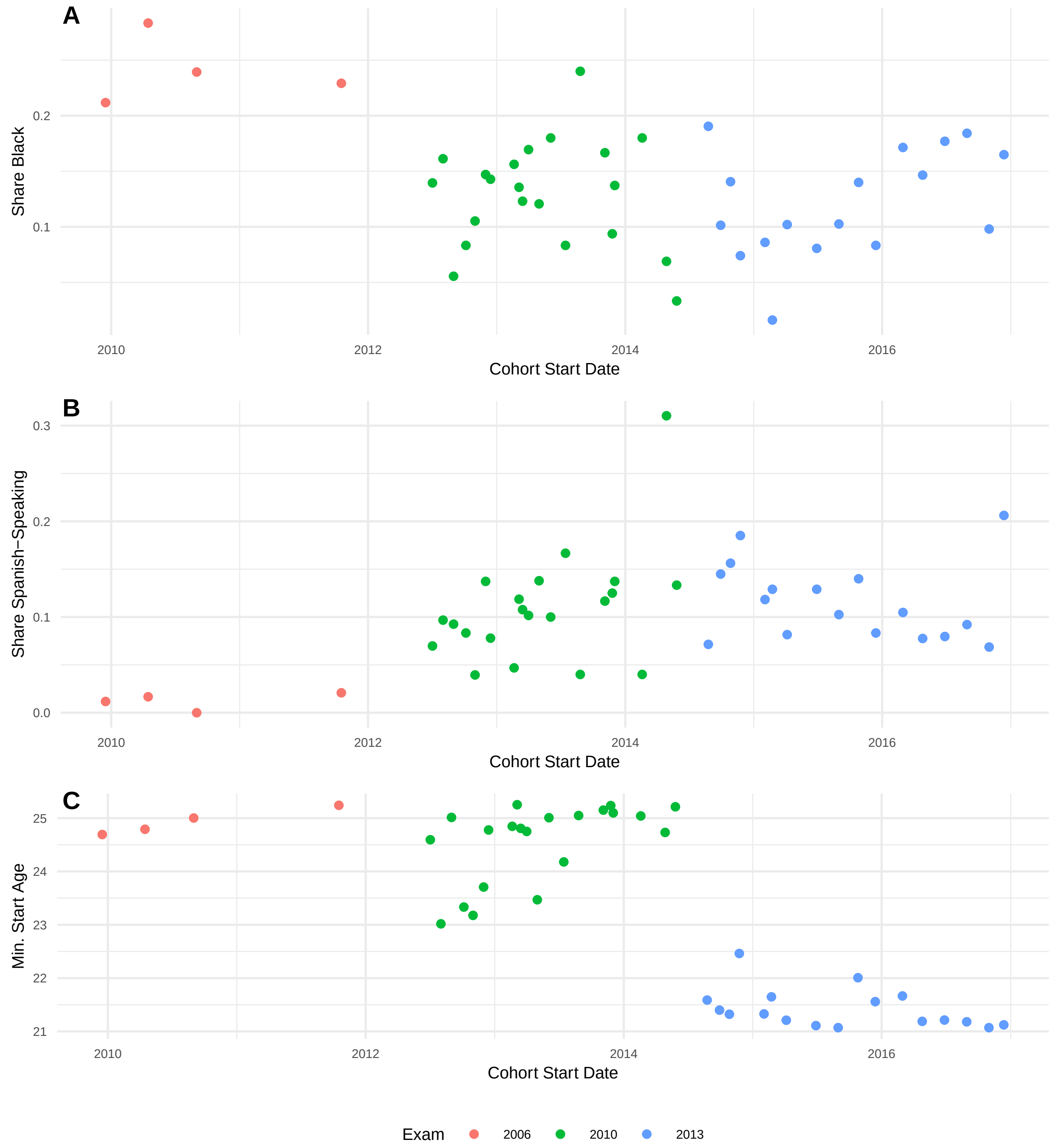
Further supporting this is the change in the composition of cohorts before and after 2012. As shown in Panel A of Figure A.1, the 2011 cohort has a higher share Black than almost every cohort in the 2012-2014 period, while it is within the range of the Exam 2006 cohorts (likely drawn from the 2006 tests). Similar patterns emerge when looking at share of the cohort which speaks Spanish (see Panel B of Figure A.1), where all of the 2006 cohorts have strictly smaller shares of Spanish speakers compared with any 2010 cohort. Finally, minimum start age (Panel C) increases successively for each of the pre-2010 cohorts (as expected since these recruits have been waiting at least 4 years to enter), while it decreases slightly in the first 2010 cohort and significantly in the second 2010 cohort. Anecdotally, an officer I spoke with who started the academy in 2012 confirmed that their cohort was comprised of 2010 test takers.

In separating the Exam 2006 cohorts (starting in 2009 and ending in 2011) from the Exam 2010 cohorts, and determining if all Exam 2006 cohorts actually came from the four 2006 exams (and not the 2004 exam), I use posts on a police forum (https://forum.officer.com/) in 2009, 2010, and 2011. One poster on November 17th, 2009, states: "Just got the call... the academy starts December 16th... My number is 1036, and I am a June 06 tester." (Chicago_mwk (2010), pg.29). December 16th, 2009, is the start date of the first cohort in my full sample. This is followed by a flurry of other posters stating their numbers also got called for the same start date. The only cohort before it was in March of 2009, which according to a poster in on March 6th 2009, "From what I know [the March 2006 cohort] it's a mix of Feb 06 and early June 06 testers." (Chicago_mwk (2010), pg.9). Overall, this indicates the 2009 and 2010 cohorts came from Exam 2006 test takers only.

Next, the main question is did the single 2011 (in October) cohort end the Exam 2006 cohorts or start the Exam 2010 cohorts? According to a different thread on the same site, a poster on December 4th, 2010, states: "With roughly 40 candidates ready for hire off the 2006 test, and a new test next week, its about time we started this thread. For those who are wondering, the last of the 2006 list (40 people) were scheduled to start on 01 November [2010] but according to my BI who I call twice every month, the class has been pushed back and only the fine folks at city hall know the date. In my humble opinion city hall is waiting on the new year [2011] to start our class because of the new budget and the new pension system for new hires." (neverlose357 (2010), pg.1) On September 30th, 2011, a poster states that their cohort ("2011-1") will "soon fill the halls of the Chicago Police Academy" (neverlose357 (2010), pg. 6), and another poster, on October 18th, 2011, (one day after the 2011 cohort starts in the data) states that the class has "About 50" recruits (49 in the data). The rest of this forum discusses the composition of this cohort. It is stated that this cohort will exhaust the rest of the 2006 applicants (at least 32) and fill the rest either with 2006 applicants who won appeals or 2010 exam-takers. So, based on these discussions, the single 2011 cohort finished off the Exam 2006 cohorts, and was potentially mixed with a small number of Exam 2010 takers– though this seems to be an unusual practice and only a result of the small number of potential recruits in the 2006 tests (neverlose357 (2010), pg. 3). While mixing a single cohort may produce issues, the exam period-specific effects discussed in Appendix A.3 indicate that the Exam 2006 period is not driving the results.

After May 2014, the cohorts until December 2016 (the last cohort in my sample) are from the 2013 test. The 2013 test recruits had the new feature that they were permitted to begin the academy at the age of 21, lower than the previous requirement of 23 (Pritchard (2013)). As can be seen in Panel C of Figure A.1, the lowest starting age per cohort drops to 21 after the May 2014 cohort. Thus, I can distinguish between the 2010 and 2013 test cohorts using this feature. The end of the 2013 test cohorts occurs after the final cohort in the full sample in December of 2016. Even though there was a test issued in April 2016, based on forum posts about 2016 recruitment the 2016 test-takers had not begun to be drawn in by the end of 2016. Following many 2016 test takers wondering when their cohorts would be drawn in, one poster stated on December 26, 2016, "People that took the exam in 2013 are still being processed. I believe about 9k people passed the written exam this year" (Aendos (2015), pg. 138). So, I am confident that the Exam 2013 sample does not contain 2016 test cohorts. Based on the panels in Figure A.1, there is fairly consistent cohort composition across the Exam 2013 cohorts. While extending my cohorts beyond December 2016 is possible, because my panel data extends to 2018 (overlapping with court data and outcomes), including the first cohorts in 2017 would not contribute much to my analysis as these officers would have less than 6 months of observations in the panel data after their probationary period.

Figure A.1: Composition of Cohorts by Start Date



Note: Figure displays the share of cohorts with more than 10 starting members that are Black (Panel A) and speak Spanish (Panel B), and the lowest starting age (Panel C) by the cohort start date, from July 2009 to 2016. Exam denotes the time period during which the cohorts started and assumes cohorts in the same period were in the same test pool.

A.2 Attrition

If the likelihood of attrition from the sample is impacted by the composition of one's cohort, then results in my estimation may be driven by selection bias rather than actual peer effects. In Table A.1, I present regression results where each outcome is a form of attrition for officers out of the analysis sample. The outcome in Column (1) contains any form of attrition, Column (2) indicates if the officer was removed for a training time violation — spent too much or too little time in the academy or probationary period — and Column (3) is whether or not the officer appeared in the final assignment data (AA). Across all outcomes, peer composition (e.g., cohort share female, share minority, average age) are not statistically or economically significant predictors of any form of attrition with respect to cohort composition. Thus it is unlikely that attrition driven selection is driving my results.

Another form of attrition is sample attrition after the recruits exit the academy, become full officers, and are present in the assignment data, e.g. cohort diversity being related to when officers choose to retire or exit the assignment data. While this may cause some officers to be more represented in the sample than others, the fixed effects recovered for the analysis sample are generally based on over 100 observations for almost all officers (94.12% of recruits). Nevertheless, I test for sample-exiting attrition in Table A.2. Column (1) contains the relationship between cohort composition and officer characteristics on the officer's number of observations in the assignment data, Column (2)'s outcome is an indicator for whether the officer exists the salary or unit history data (which contains officers not in the assignment data) at the end of 2018, Column (3)'s outcome is an indicator for promoted by the end of 2018, and Column (4)'s outcome is being assigned to a non-geographic unit ('specialized') at the end of 2018. While generally small or not statistically significant, peer composition does influence number of observations and exits from the sample, though not to an extent where it would result in mass attrition biasing the sample composition.

Table A.1: Attrition from Sample

	Any Attrition	Time Violation	Not in Final AA
	(1)	(2)	(3)
Share Black	-0.366	-0.345	0.297
	(0.553)	(0.513)	(0.384)
Share Non-Black Minority	0.193	0.204	-0.168
	(0.188)	(0.213)	(0.201)
Mean Age	0.008	0.015	-0.007
	(0.033)	(0.03)	(0.022)
Share Female	0.333	0.35	-0.385
	(0.413)	(0.386)	(0.275)
Share High Edu	0.09	0.067	-0.097
	(0.219)	(0.201)	(0.148)
Share Spanish-Speaking	0.188	0.17	0.054
	(0.226)	(0.24)	(0.188)
Share Military	0.11	-0.016	-0.217
	(0.282)	(0.281)	(0.212)
Black	0.023	0.013	-0.007
	(0.019)	(0.019)	(0.015)
Non-Black Minority	0.021*	0.026*	-0.008
	(0.011)	(0.015)	(0.013)
Male	-0.029*	-0.022	0.03*
	(0.017)	(0.019)	(0.018)
Start Age	-0.001	-0.001	0.002
	(0.001)	(0.002)	(0.001)
Size	-0.001*	-0.001*	0.001*
	(0.001)	(0.001)	(0)
Exam 2010	-0.137*	-0.141**	0.027
	(0.08)	(0.068)	(0.048)
Exam 2013	-0.169***	-0.166***	0.064
	(0.054)	(0.049)	(0.041)
(Intercept)	-0.124	-0.177	1.278*
	(1.034)	(0.963)	(0.677)
N	2528	2698	2698
R2	0.034	0.031	0.015

Note: Table displays the OLS regression estimates of cohort and officer observables on officer attrition from the sample. The dependent variables for Columns (1)-(3) are: (1) whether the officer was dropped for any reason from the sample; (2) (2) whether or not the officer is dropped due to spending too much or too little time in the academy or probationary period; (3) whether or not the officer is not in the final assignment data. Standard errors clustered at cohort level are in parentheses. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table A.2: Attrition out of Sample

	N. Obs in Data	Exit Data	Promoted at End	Specialized Unit at End
	(1)	(2)	(3)	(4)
Share Black	-106.788	0.381*	-0.049	0.007
	(437.846)	(0.193)	(0.233)	(0.159)
Share Non-Black Minority	-312.58	-0.121	-0.198	-0.196
	(235.241)	(0.176)	(0.151)	(0.136)
Mean Age	-43.288**	0.016	-0.029**	-0.021**
	(20.394)	(0.01)	(0.011)	(0.009)
Share Female	-890.831**	-0.346***	-0.052	-0.052
	(347.306)	(0.114)	(0.188)	(0.157)
Share High Edu	114.16	0.11	-0.18	-0.181
	(231.012)	(0.111)	(0.134)	(0.109)
Share Spanish-Speaking	-151.734	-0.089	0.305*	0.127
	(340.834)	(0.131)	(0.155)	(0.192)
Share Military	-313.392	-0.168	0.354	0.24
	(300.708)	(0.171)	(0.236)	(0.213)
Black	-23.205	-0.007	-0.038**	-0.066***
	(16.198)	(0.018)	(0.018)	(0.017)
Non-Black Minority	12.686	0.01	-0.032**	-0.048***
	(11.999)	(0.012)	(0.015)	(0.013)
Male	98.614***	0.002	0.009	0.004
	(19.321)	(0.012)	(0.015)	(0.012)
Start Age	2.023	0.004***	0	-0.003**
	(1.354)	(0.001)	(0.001)	(0.001)
Size	-0.529	0	0.001**	0.001**
	(0.666)	(0)	(0)	(0)
Exam 2010	-267.771***	0.07***	-0.032	-0.018
	(55.814)	(0.02)	(0.028)	(0.024)
Exam 2013	-566.09***	0.187***	-0.262***	-0.19***
	(59.34)	(0.024)	(0.037)	(0.029)
(Intercept)	2559.216***	0.395	0.858**	0.785**
	(577.408)	(0.279)	(0.344)	(0.343)
N	2457	2567	2369	2369
R2	0.417	0.032	0.087	0.086
SD(Outcome)	302.73	0.27	0.31	0.25

Note: Table displays the linear regression estimates of cohort and officer observables on officer observations and other measures of attrition for the analysis sample. The dependent variables are the officer's number of observations (shifts) used to estimate fixed effects in the daily panel data (Column (1)), whether or not the officer is in the salary and unit history data which contains non-D1 officers and units outside of the assignment data (Column (2)), whether the officer has been promoted by the end of 2018 (Column (3)), whether the officer is in a specialized unit at the end of 2018 (Column (4)). Standard errors clustered at cohort level are in parentheses. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

A.3 Robustness

In this section, I discuss a variety of additional analyses to test the robustness of the results using the specification from Panel A in Table 8. Table B.7 presents results for average arrest outcomes, and Table B.8 presents results from analogous tests for officer arrest propensity outcomes when feasible.

A.3.1 Alternate Outcomes

As the number of arrests is count data, I estimate the relationship between peer composition and total arrests (controlling for total shifts) with a Poisson regression, and I recover alternative estimates of officer effects (equation (3)) using a fixed-effects Poisson regression and then re-estimate equation (1) using them. This model is potentially more reflective of the true data-generating process and allows for peer composition and officer effects to contribute to arrests in a non-linear fashion. Specifically,

$$\mathbb{E}[Arrest_{it}|\theta_i, \gamma_{brsw_t}, V_{it}] = exp(\theta_i + \gamma_{brsw_t} + \beta V_{it}) \tag{4}$$

for recovering fixed effects, and

$$\mathbb{E}[Arrest_i|X_i, \overline{X}_{c(-i)}, \eta_{p(i)}, N_i] = exp(\pi_2 \overline{X}_{c(-i)} + \pi_1 X_i + \eta_{p(i)} + N_i) \tag{5}$$

for average arrests, where $Arrest_i$ is total arrests over the sample period and N_i is the number of observations (shifts). A negative binomial regression is unnecessary as the data are not overly dispersed.

Column (1) in both Tables B.7 and B.8 display the results. The results for low-level arrests and officer propensities and serious officer effects are qualitatively similar to the main results, though the estimates are not directly interpretable due to non-linearity. However, the average serious arrest coefficients are negative, inconsistent with the main results.

Another concern is that skewed arrest data drives the results: most shifts have no arrests, while very few have many. I test for the sensitivity of my results to this by binarizing the shift-level outcomes into whether an arrest was made on that shift. I recover alternative officer effects by estimating equation (3) as a linear probability model (LPM) with the dependent variable being if any arrest was made by officer i during their shift, and the average arrest outcome becomes the share of shifts in which an arrest was made. The results (Column (2) of both tables) are similar to the main results. Together, these tests indicate that the results were not driven by the reliance on a linear model in the first stage or the skewed distribution of arrests per shift.

As multiple officers can be listed on a single arrest, some arrests are double-counted in my analysis. This may be an issue if cohort composition influences assignments in which only single-officer arrests generally occur. I reproduce my results by counting arrests for only the primary arresting officer. Column (3) in both tables displays the results, similar to the main results but smaller, as expected. Next, I re-categorize the arrests based on the FBI index crimes such that arrests are serious if they are for index crimes and low-level if non-index to determine if my categorization of serious and low-level is spuriously producing results.[15] The coefficients using these alternate definitions (Column (4) of both tables) are generally consistent with the main results.

A.3.2 Alternative Samples and Controls

As three exam periods come from different periods, it is important to ensure that no single exam period drives the results. I redo the analysis for Exam 2010 and Exam 2013 separately, as displayed in Columns (5) and (6). While the results are generally noisier due to the smaller sample sizes, the point estimates, particularly for older minorities and female peers, are consistent with the main results indicating their effects are not specific to one exam period. Next, to alleviate concerns over selectively dropping officers for training time violations or lack of observations, I include all officers in the sample cohorts for whom average arrests or officer fixed effects could be recovered, producing similar results in Column (7).

With additional information about officers, I can add more controls about both officer and cohort compositions to test the robustness of the main results and ensure that other correlated peer features, such as education, are not driving the results. These additional controls include officer-level and cohort shares of Spanish-speaking, high education (bachelor's degree or above), and military experience. The results in Column (8) are similar to the main results.

Focusing on the behavior versus assignment analysis, I test the robustness of the results on alternate assignment definitions in Columns (9)-(10) of Table B.8 using officer effects recovered from re-estimating equation (3) using alternative assignment definitions. First, I repeat my analysis using the more granular assignment fixed effects ('MDSBs') from Ba, Knox, et al. (2021).[16] Column (9) presents results similar to the main results though slightly larger. Next, I relax the assignment effects by separating them into two components, time effects (interacting year, month, day of week, and shift) and role effects (interacting beat description and unit and year), which are included additively ($\gamma_{a(i,t)} = \gamma_{sw_t} + \gamma_{bry}$). Then, I recover officer and assignment fixed effects using equation (3). This relaxes the stringency of assignment effects by looking at unit and role (rather than distinguishing between the same role in different sectors within the same unit) and ensures that the results are not due to the interaction of role and granular time effects. The results for officer effects are displayed in Columns (10) and are consistent with the main results.

[15]Index crimes are murder, rape, robbery, aggravated assault, burglary, theft, motor vehicle theft, and arson; and non-index crimes are all others but exclude warrant arrests as the exact crime type is not known.

[16]These control for the interaction between year-month, day of the week, shift, and exact beat code ('MDSB'), whereas I interact assignment role with a truncated beat code, year-month, day of week, and shift.

A.3.3 Age Cutoffs

The main results use 27 as the age cutoff between older and younger officers. I test the robustness of my results against alternative cutoffs to ensure the effects of older peers are not due to the exact age cutoff. Figure B.5 displays the main specification coefficients using age cutoffs from 25 (the 14th percentile) to 33 (80th percentile), with the 25th, 50th, and 75th percentiles denoted by vertical lines. While alternative cutoffs produce qualitatively similar effects, they generally become noisier and smaller at the upper edges of the age distribution, particularly for white officers. The results are consistent with the impact of older peers being driven by the exclusion of relatively young officers who are also police low-level crimes most aggressively (see Figure 2). The coefficients on female peers are consistent across outcomes and age cuts, consistent with less heterogeneity in female preferences by age group in Figure 2. These results indicate that the 27-year cutoff is not spuriously producing economically significant results though the precision of the estimates changes with cutoffs.

A.3.4 Measurement Error and Instruments

Angrist (2014) outlines many issues common in similar peer effects studies which incorrectly produce or overstate peer effects. First, as noted in Acemoglu and Angrist (2000), peer effects can be over-estimated due to classical measurement error in the peer characteristics — which is likely the case in this study if peer race, age, and gender are seen as proxies for preferences. However, Feld and Zölitz (2017) shows that classical measurement error only amplifies effects if assignment is non-random; otherwise, it attenuates effects. To test for this, I follow Carrell, Hoekstra, and Kuka (2018) by adding measurement error to cohort compositions. Figure B.6 displays the results of adding increasing amounts of measurement error to cohort composition in terms of race, gender, and age. Adding measurement error to race or age does not amplify any coefficients (other than for young minority peers occasionally), with error in race modestly attenuating the effects of older minority peers, while error in age significantly attenuates the effects of all older peer groups — though older white peers are most influenced consistent with the additional effects of minority peers beyond age. Adding error to gender attenuates the coefficient on female peers. These results are not consistent with the main results over-estimated due to measurement error and non-random peer assignment.

Second, Angrist (2014) discusses the relationship between individual outcomes and group averages and understanding social spillovers through an instrumental variables framework, wherein estimated peer effects may be spurious or exaggerated. In Table B.9, I first display the results of regressing the main outcomes on officer-level characteristics used in the main specification (Panel A). I contrast this with the 2SLS estimates of regressing main outcomes on cohort compositions by instrumenting for officer-level characteristics with cohort indicators as instruments. The 2SLS estimates in Panel B far exceed the OLS estimates in Panel A and the first-stage R^2 is small, consistent with a social-multiplier effect. As expected, cohort indicators are weak instruments for officer characteristics (small first stage F-statistic). To test the robustness of the 2SLS results, I recover the effect of peer composition using split-sample instrumental variables (SSIV) by randomly splitting each cohort in half and using one half's composition as an instrument for the composition of the other half (Angrist and Krueger (1995)). SSIV reduces concerns over the inclusion of an individual's own observation contributing to both cohort composition and average outcomes (Chetty et al. (2011)), the many weak instruments bias, and the implicit jackknife instrumental variables design (Angrist, Imbens, and Krueger (1999)) when using leave-out means of peer composition. Panel C displays the SSIV results. Though the estimates are significantly less precise than the main results, consistent with the sample size being cut in half, the coefficients are generally directionally similar to those of the main and 2SLS results, particularly the effects of older minorities.

A.3.5 Inference

The main specification has multiple variables of interest, requiring adjustments for multiple hypothesis testing. Table B.10 displays the p-values for the main specification along with adjusted p-values using the Benjamini and Hochberg (1995) (BH) adjustment controlling the false discovery rate and Holm (1979) adjustment controlling for the family-wise error rate. Overall, while the officer effects results are unaffected by either method, and the BH adjustment does not change the main conclusions for low-level average arrests at the 10% level (only female peers and older minority peers are significant), only female peers result survives the Holm correction at the 10% level. The average serious arrest results have large adjusted (and unadjusted) p-values.

Additionally, traditional inference techniques do not necessarily apply to many (quasi-)experimental designs, particularly peer effects studies where inter-group variation results from finite-sample bias. Recent peer effect studies use randomization inference to construct p-values for estimates (Carrell, Sacerdote, and West (2013), Caeyers and Fafchamps (2016), Carrell, Hoekstra, and West (2019)), consistent with the guidance in Athey and Imbens (2016). I construct p-values using randomization inference, which provides a distribution of estimates under the null hypothesis that peer composition has no effect on outcomes. Column (2) and (6) in Table B.10 shows that the randomization inference p-values are generally smaller than or similar to those in the main results, and I discuss the process in more detail below.

Randomization inference (or randomization-based inference) allows us to construct an empirical distribution of coefficients under the null hypothesis, that peers have no effect on the outcomes of interest. This is preferable to traditional asymptotic inference in which the error in estimates is a result of sampling error because in such environments, there is no sampling error: the sample of CPD recruits between 2009 and 2016 is the population. Such methods have their origin in Fisher (1925), wherein one wants to test to see if they can reject the 'sharp' null hypothesis that the treatment has no effect on the outcome of interest, and much of this section will follow Athey, Eckles, and Imbens (2018). Let us generalize equation (1) (removing superscript k for simplicity) as a potential outcomes function:

$$Y_i(P_i = \overline{X}_{c(i)}) = \alpha_{p(i)} + \pi_1 \overline{X}_{c(-i)} + v_i$$

Then the potential outcomes function for an individual, Y_i, takes in a value for i's peer composition P and tells us what the outcome (e.g., mean arrests or officer fixed effect) would be had they had peer composition P in the academy. As discussed in Athey, Eckles, and Imbens (2018), under a sharp null hypothesis of no effect, given some treatment assignment P' and the realized outcomes for that specific assignment $Y_i(P')$, one can infer the value of the outcome at any other treatment assignment. Essentially if under the null that $\pi_1 = 0$, then $Y_i(P) = Y_i(P') \forall P, P' \in \mathbb{P}$ where P' is any possible peer composition and $\mathbb{P}$ is the space of all possible treatment (peer) assignments. The intuition is that if the true peer effect is zero ($\pi_1 = 0$), then it should not matter what treatment (peer composition) is assigned.

Now, we can test this null hypothesis. We can generate test statistics based on the distribution of estimated treatment effects (π_1^r), the 'randomization distribution', when the treatment status is randomly assigned. With this distribution of estimated peer effects under the randomized treatments, we compare the estimate from our actual data ($\hat{\pi}_1$) to the randomization distribution and recover the p-value– the likelihood of finding an effect more extreme than the one estimated under the null hypothesis that treatment has no effect. Again, borrowing from Athey, Eckles, and Imbens (2018):

$$p\text{-}value = Pr(|\hat{\pi}_1(Y_i(P = \overline{X}_{c(-i)}))| \geq |\pi_1^r(Y_i(P'))|)$$

With this p-value, we can assess the likelihood that the estimate recovered from the actual data ($\hat{\pi}_1$) is consistent with the null hypothesis that the peer effect is null.

In practice, constructing the randomization distribution can be done in two ways. (1) (Re-assigning Treatment) Randomly re-assigning individuals to cohorts within exams and ensuring cohort sizes remain the same and thus constructing randomized treatments ($\bar{X}^r_{c^r(i)}$), then estimate:

$$\hat{\theta}_i = \alpha_{p(i)} + \pi_1^r \overline{X}^r_{c^r(i)} + \pi_2 X_i + v_i$$

Or (2) (Re-assigning Outcomes), randomly re-assigning outcomes to individuals (θ_i^r):

$$\hat{\theta}_i^r = \alpha_{p(i)} + \pi_1^r \overline{X}_{c(-i)} + \pi_2^r X_i + v_i$$

Both methods produce similar results, and I proceed by using method (1). In either case, this procedure can be repeated N number of times (I perform 1,000 iterations for method (1)) with each iteration producing an estimate of π_1^r. Then, the coefficient using the actual data, $\hat{\pi}_1$ can be compared with the distribution of $\hat{\pi}_1^r$ to obtain a p-values as discussed above. Method (1) is used in Caeyers and Fafchamps (2016) and Michelman, Price, and Zimmerman (2021), while Method (2) is used in Carrell, Sacerdote, and West (2013) and Carrell, Hoekstra, and West (2019).

In practice because of sample attrition, method (1) involves re-drawing cohorts (within exams) using recruits in the final sample and those who are dropped from it. Furthermore, the method takes the error in the outcome (e.g., Y_i is an estimate with measurement error if we use officer fixed effects) as given. In both cases, two-sided p-values are computed by ranking the coefficient in the main results within the distribution of placebo coefficients.

A.4 Small Class Effects (Homerooms)

While many classes were composed of almost all the officers in one's cohort, smaller sub-cohort groups ("homerooms") are identifiable when restricting to classes with fewer than 30 recruits. I use data on individual classes the officers took while in the academy to see if recruits in small group (homeroom) composition is driving the effects of cohort composition on the outcomes. If this is the case, then it is more likely that instructor effects are a contributing factor.

The training data provided lists the set of classes each probationary officer took during their time at the academy. This includes classes on the data base access, report writing, terrorism, chemical and radioactive events, and use of force. Many classes are large containing almost all (or a large portion) of a cohort's members. A subset of courses contain fewer officers per class, meaning there is larger within-cohort variation on which cohort members attended these courses together.

I use the set of trainings during the academy that full sample officers took which had fewer than 30 officers attend and a sufficiently high share of the classes being from the same cohort. With this set of courses, I created a weighted undirected network of recruits within cohorts and use the "edge betweenness" clustering algorithm (Newman and Girvan (2004)) (implemented in the igraph package in R (Csardi and Nepusz (2005))) in order to partition these networks into sub-communities of officers that had the strongest ties based on classes taken together. I refer to these sub-cohorts as homerooms.

After some filters, the final sample of officers in the homerooms (also in the full sample) is 2,038 in 102 homerooms. Not all recruits are present in the final homeroom data due to matching issues and filters (88.76% of full sample officers are in the final homeroom data) and I restrict to homerooms with between 14 and 30 recruits. Due to the smaller size of these homerooms, there is much more variation in compositions. For example, there is 2.5 times more variation in cohort share minority for homerooms relative to cohorts. Nevertheless, homeroom and cohort compositions are highly correlated.

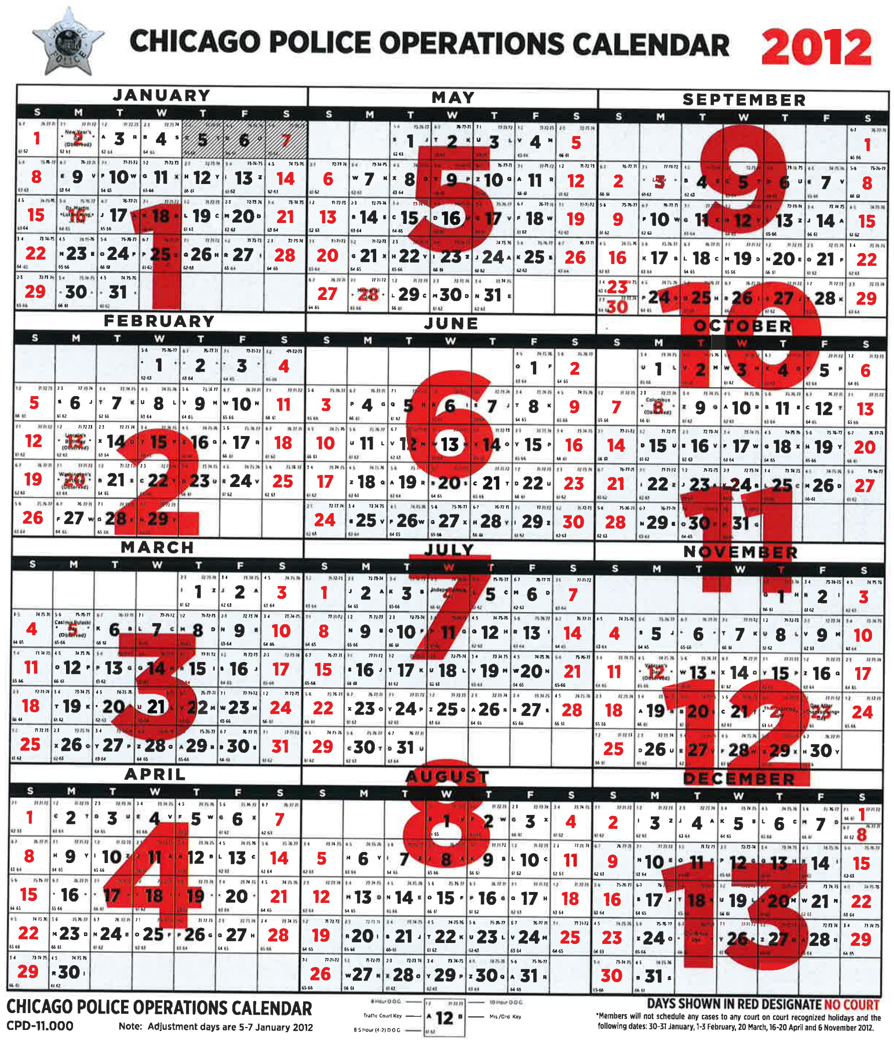
B Appendix B - Additional Figures and Tables

Figure B.1: CPD Exam Information

Exam	Date of administration	Attended	Passed	Failed
Police Entrance 1999	3/15/1999; 3/16/1999	3,967	No info available	No info available
Police Entrance 1999	1/5/2000	2,517	No info available	No info available
Police Entrance 2000	7/1/2000	2,053	No info available	No info available
Police Entrance 2000	1/4/2001	1,829	No info available	No info available
Police Entrance 2001	5/19/2001	1,923	No info available	No info available
Police Entrance 2002	1/12/2002	3,150	No info available	No info available
Police Entrance 2003	11/22/2003	3,875	No info available	No info available
Police Entrance 2004	11/20/2004	4,163	No info available	No info available
Police Entrance 2005	2/18/2006; 2/19/2006	4,061	3,338	723
Police Entrance 2006-1	6/4/2006	1,508	1,255	253
Police Entrance 2006-2	8/6/2006	1,025	863	162
Police Entrance 2006-3	11/5/2006	1,795	1,487	308
Police Entrance 2010	12/11/2010	8,621	7,689	932
Police Entrance 2010 make up	makeups: 3/12/2011; 6/11/2011; 9/25/2011; 12/3/2011; 6/2/2013; 12/1/2012; 3/9/2013	No info available	No info available	No info available
Police Entrance 2013	12/14/2013 & military makeups (6/28/2014; 12/7/2014; 6/13/2015; 12/6/2015)	14,788	12,877	1,911
Police Entrance 2016	4/16/2016 & make ups :12/3/2016; 12/4/2016	10,199	9,023	1,176
Police Entrance Spring 2017	4/1/2017-4/2/2017	8,620	7,437	1,183
Police Entrance Winter 2017	12/16/2017,12/17/2017 & makeup: 2/24/2018	7,294	6,418	876
Police Entrance Spring 2018	5/5/2018 & 5/6/2018 & makeup: 6/23/2018	4,273	3,789	484
Police Entrance Winter 2018	12/8/2018	4,433	3,964	469
Police Entrance Winter 2018 make up	3/9/2019	Hasn't occurred	N/A	N/A

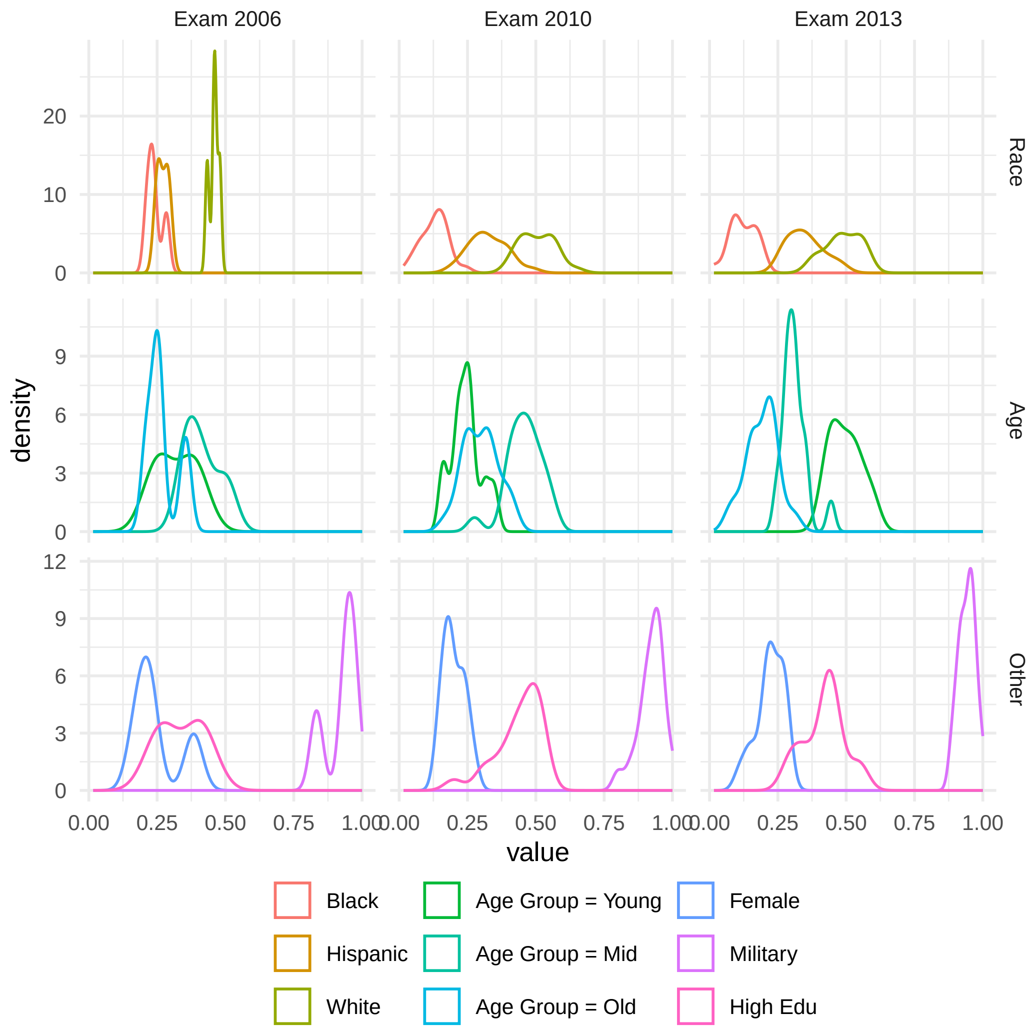
Note: Figure displays information on CPD entrance exam information, the date of the exam and the numbers of applicants that attended, passed, and failed the exam.

Figure B.2: CPD Operations Calendar (2012)



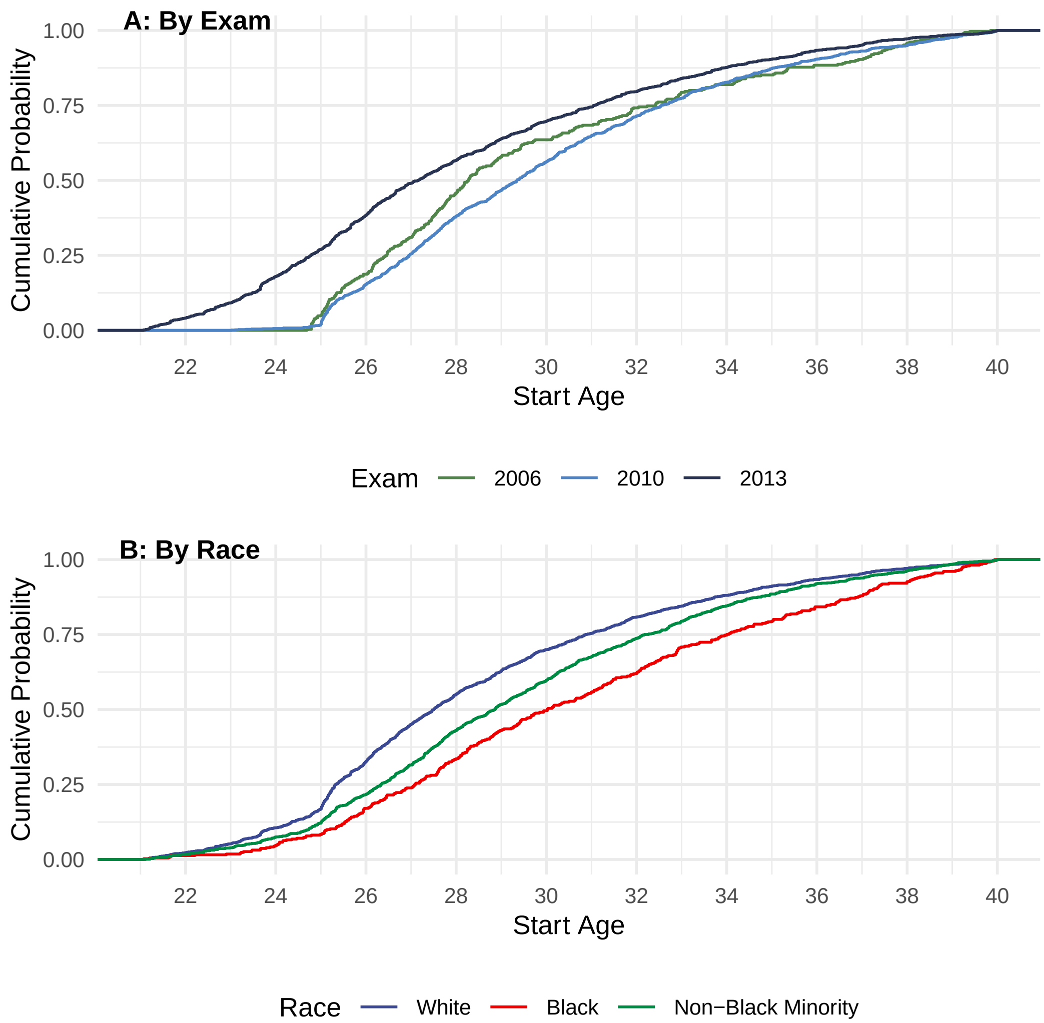
Note: Figure displays an example of the CPD operations calendar for the year 2012.

Figure B.3: Cohort Composition



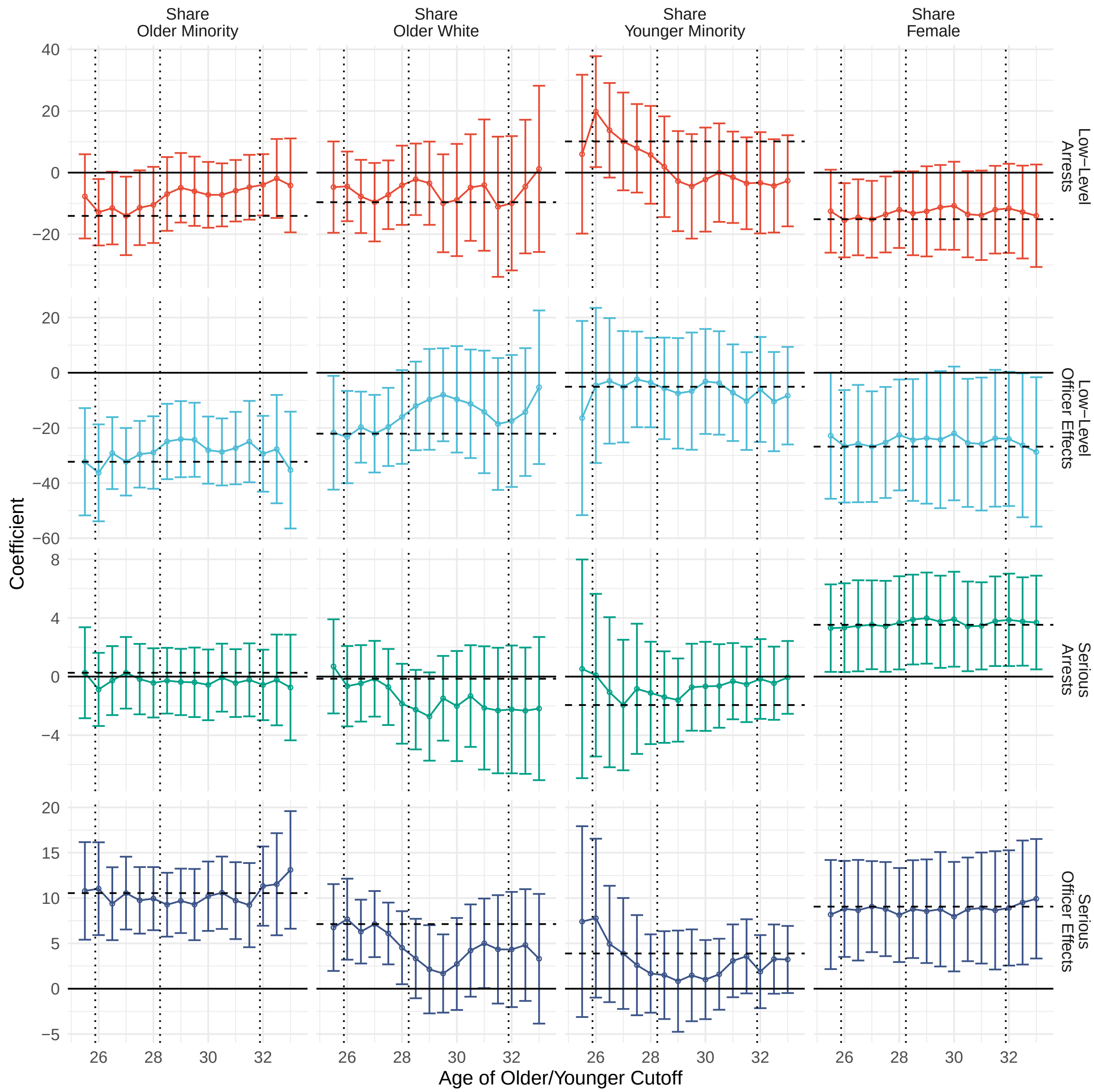
Note: Figure displays the distributions of cohort compositions for Exam periods 2006, 2010, and 2013 for characteristics including race (share Black, Hispanic, white), age (young = <27, mid=[27,32], and old= >32), gender (share female), and shares of those with military experience and high education (Bachelors or above).

Figure B.4: CDF of New Officer Start Ages



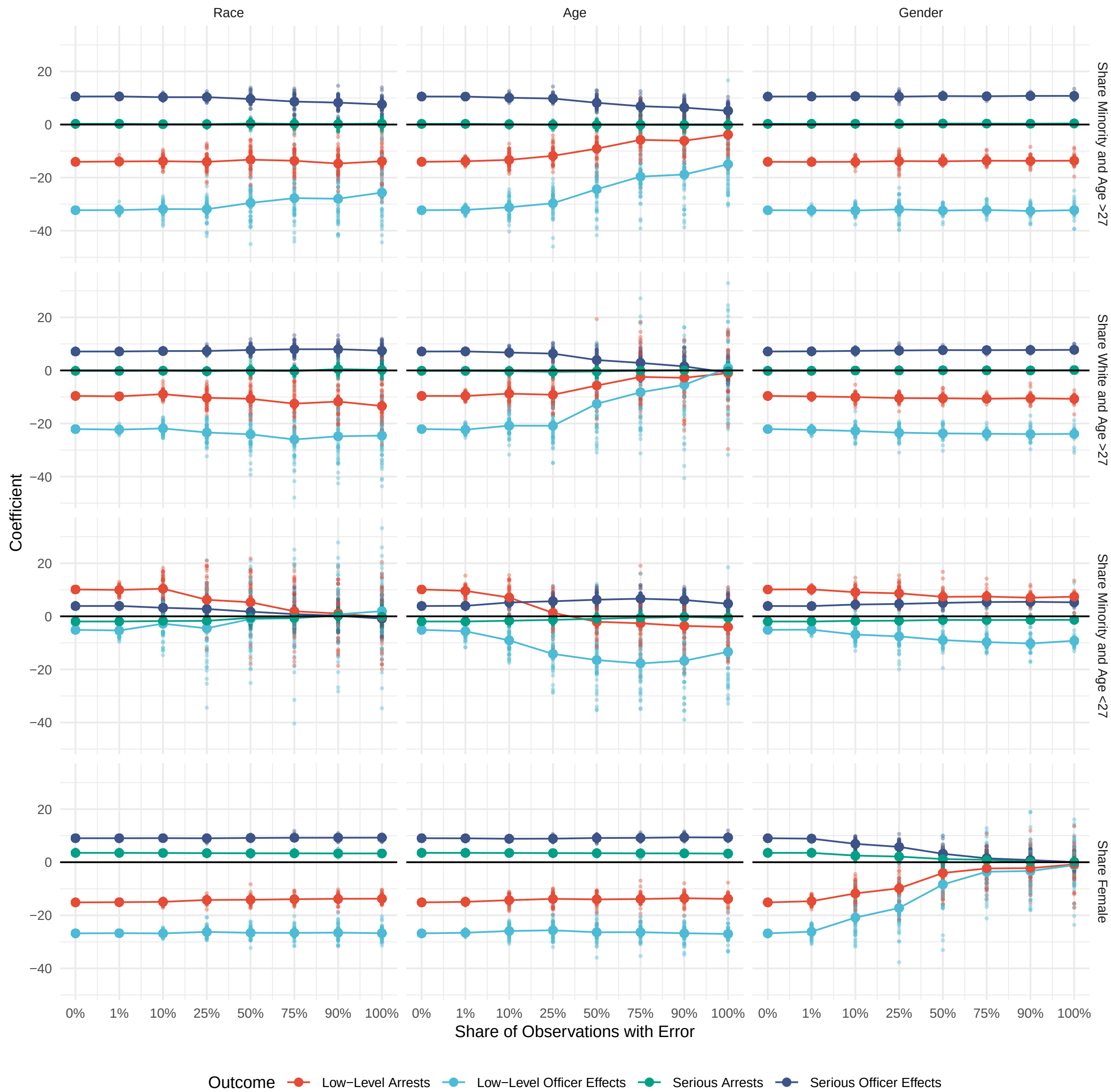
Note: Figure displays the cumulative distributions of officer start ages in cohorts for each Exam 2006, 2010, and 2013 (Panel A) and for each race group (Panel B). The top figure illustrates that officers cannot begin at the academy after the age of 40 or before the age of 23 prior to Exam 2013 and 21 in Exam 2013 due to a policy change.

Figure B.5: Change in Coefficients with Alternative Age Cutoffs



Note: Figure visualizes how coefficients change as the age cutoff between 'older' and 'younger' peers changes from 25 years to 33 years at half-year increments using the main specification (except using different age cuts whereas 27 was used in the main specification) and main outcomes in units of arrests per 100 shifts, with 2,296 observations. Vertical dotted lines denote the 25th, 50th, and 75th percentiles in the age distribution, and horizontal dashed lines denote the coefficient in the main results corresponding to using 27 as the older/younger age cutoff. Error bars correspond to 95% confidence intervals using standard errors clustered at the cohort level.

Figure B.6: Change in Coefficients with Measurement Error



Note: Figure visualizes how coefficients change as measurement error is added to peer race, age, and gender. Coefficients are the effects of cohort shares of older (starting age > 27) minority and white, share younger minority, and share female peers on the main outcomes. Measurement error is induced by taking the initial sample and assigning racial group (minority and white), age group (older or younger than 27), and gender (male or female) based on a uniform random variable for some share ('Share Error') of the sample, then new peer compositions are computed. For each share error of observations with measurement error, this exercise is repeated 50 times, and each faint dot corresponds to a particular run. The larger dots (on per value of share error), are the mean coefficients across runs. Officer effects are individual officer fixed effects estimated using equation (3). Coefficients estimated using equation (1) using controls for officer group membership and exam period fixed effects.

Table B.1: Additional Summary Statistics of Cohort Composition

	Min	Median	Mean	Max	IQR	SD
Cohort Share White	0.38	0.48	0.49	0.65	0.10	0.06
Cohort Share Hispanic	0.19	0.32	0.33	0.48	0.08	0.07
Cohort Share Black	0.02	0.14	0.14	0.28	0.08	0.06
Cohort Share Asian/Native American	0.00	0.04	0.04	0.08	0.03	0.02
Cohort Share Minority	0.35	0.52	0.51	0.62	0.10	0.06
Cohort Share Non-Black Minority	0.19	0.36	0.37	0.52	0.09	0.07
Cohort Share Female	0.11	0.21	0.21	0.38	0.06	0.05
Cohort Share Military	0.80	0.94	0.93	1.00	0.05	0.04
Cohort Share High Edu	0.20	0.43	0.42	0.55	0.11	0.08
Cohort Share Start Age	26.75	29.43	29.25	31.03	2.08	1.18

Note: Table presents summary statistics of cohort compositions across all periods including minimum, median, mean, maximum, interquartile range, and standard deviation. Statistics computed across 42 cohort observations not weighted by cohort size.

Table B.2: Correlations Across Demographics

	White	Non-Black Minority	Black	Female	Male	Start Age
Exam 2006						
White	1.00	-0.61	-0.51	-0.15	0.15	-0.26
Non-Black Minority	-0.61	1.00	-0.37	-0.03	0.03	0.12
Black	-0.51	-0.37	1.00	0.21	-0.21	0.17
Female	-0.15	-0.03	0.21	1.00	-1.00	0.02
Male	0.15	0.03	-0.21	-1.00	1.00	-0.02
Start Age	-0.26	0.12	0.17	0.02	-0.02	1.00
Exam 2010						
White	1.00	-0.77	-0.39	-0.10	0.10	-0.19
Non-Black Minority	-0.77	1.00	-0.29	0.03	-0.03	0.09
Black	-0.39	-0.29	1.00	0.10	-0.10	0.15
Female	-0.10	0.03	0.10	1.00	-1.00	0.14
Male	0.10	-0.03	-0.10	-1.00	1.00	-0.14
Start Age	-0.19	0.09	0.15	0.14	-0.14	1.00
Exam 2013						
White	1.00	-0.77	-0.37	-0.07	0.07	-0.12
Non-Black Minority	-0.77	1.00	-0.30	0.05	-0.05	0.04
Black	-0.37	-0.30	1.00	0.04	-0.04	0.13
Female	-0.07	0.05	0.04	1.00	-1.00	0.11
Male	0.07	-0.05	-0.04	-1.00	1.00	-0.11
Start Age	-0.12	0.04	0.13	0.11	-0.11	1.00

Note: Table presents correlations across officer demographics within exam periods.

Table B.3: Effect of Peer Composition on Arrest Quality

	Arrests per 100 Shifts			
	Low-Level		Serious	
	Guilty	Non-Guilty	Guilty	Non-Guilty
	(1)	(2)	(3)	(4)
A: Average Arrests				
Share Black	-1.2	-1.3	0.5	0.9
	(1.5)	(5.4)	(0.6)	(1.2)
Share Non-Black Minority	-0.3	0.3	-0.5	-1.2
	(1.3)	(5.1)	(0.4)	(0.8)
Share Female	-3.2***	-8.8**	-0.5	1.6
	(1.2)	(4.3)	(0.5)	(1.1)
Share Age >27	-3***	-9**	-0.7*	-0.1
	(0.9)	(3.9)	(0.4)	(0.9)
B: Officer Effects				
Share Black	-2	-8.8	1.6	3.3**
	(2.1)	(10.4)	(0.9)	(1.5)
Share Non-Black Minority	-1.2	-7.1	0.6	1.8
	(1.4)	(6.7)	(0.5)	(1.1)
Share Female	-5***	-20.1**	1.8**	4.9***
	(1.7)	(8.6)	(0.8)	(1.5)
Share Age >27	-4.4***	-18.6***	1.4**	3.8***
	(1.2)	(5.7)	(0.5)	(1.1)

Note: Table displays the OLS result for the effect of cohort composition on sample officers' average arrests (Panel A) and estimated officer effects (Panel B) for arrests eventually resulting in a guilty finding and those not resulting in a guilty finding (non-guilty), in units of arrests per 100 shifts using equation (1), with 2,296 observations. All regressions include exam period fixed effects and controls for officer-level characteristics. Officer effects are recovered from estimating equation (3). Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table B.4: Effect of Peer Composition on Arrest Subtypes

	Arrests per 100 Shifts									
	Serious				Low-Level					
	Index Violent	Nonindex Violent	Index Property	Nonindex Property	Drug	Traffic	Weapon	Municipal	Warrant	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A: Average Arrests										
Share Black	0.5	0.1	1.8*	-0.1	-1.5	0.5	-0.3	0.5	-0.4	-0.6
	(0.5)	(0.7)	(1)	(0.2)	(1.8)	(1.7)	(1)	(0.6)	(2.1)	(1.4)
Share Non-Black Minority	0	-0.6	-0.8	0	0.6	0.8	0.2	-0.1	1	-0.9
	(0.3)	(0.6)	(0.8)	(0.1)	(1.7)	(1.8)	(0.7)	(0.4)	(1.3)	(1.6)
Share Female	-0.2	1.7**	0.5	0.2	-5.1***	-2.2	-1.4*	-1**	-1.5	-0.6
	(0.4)	(0.7)	(0.7)	(0.2)	(1.4)	(1.4)	(0.8)	(0.4)	(1.8)	(1.2)
Share Age >27	-0.7**	0.3	0.2	0.2	-1.8	-0.9	-1.5***	-0.9*	-2.9**	-3.6***
	(0.3)	(0.5)	(0.7)	(0.2)	(1.7)	(1.2)	(0.5)	(0.5)	(1.3)	(1)
B: Officer Effects										
Share Black	0.8**	-0.3	5.3	-0.1	-6.3	-2.1	1.1	0.5	-0.7	-2.1
	(0.4)	(1.1)	(3.5)	(0.2)	(4.8)	(2.4)	(0.7)	(0.5)	(2.3)	(2.3)
Share Non-Black Minority	0.2	-0.2	2.8	0	-2.9	-1.1	0.7	-0.2	-0.4	-2.8
	(0.2)	(0.6)	(2)	(0.2)	(3.1)	(1.8)	(0.6)	(0.3)	(1.3)	(1.7)
Share Female	-0.1	0.5	7.7**	0.4**	-11.3***	-4.1**	0.7	-1**	-3.2	-3.9**
	(0.2)	(0.7)	(3.1)	(0.2)	(4.2)	(1.9)	(0.6)	(0.4)	(2)	(1.9)
Share Age >27	-0.4**	-0.6	7.6***	0.3**	-8.3***	-2.7**	0.7	-0.8**	-4.3***	-5***
	(0.2)	(0.5)	(2.2)	(0.1)	(2.8)	(1.3)	(0.5)	(0.3)	(1.4)	(1.2)

Note: Table displays the OLS result for the effect of cohort composition on sample officers' average arrests (Panel A) and estimated officer effects (Panel B), in units of arrests per 100 shifts using equation (1), with 2,296 observations. All regressions exam period fixed effects and controls for officer-level characteristics. Officer effects are recovered from estimating equation (3). See Appendix C for more details on crime type classification. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table B.5: Additional Effects on White Officers

	Low-Level				Serious			
	Average Arrests		Officer Effects		Average Arrests		Officer Effects	
	Base	x PO White	Base	x PO White	Base	x PO White	Base	x PO White
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: Pooled Minorities								
Share Minority and Age >27	-10.64	-7.63	-30.24***	-4.56	-0.33	1.13	10.72***	-0.32
	(7.34)	(8.1)	(6.53)	(5.23)	(1.55)	(1.59)	(2.15)	(1.34)
Share White and Age >27	-10.94	1.91	-23.96***	3.22	-0.57	0.84	6.2***	1.81
	(6.58)	(5.9)	(6.99)	(4.03)	(1.43)	(1.67)	(2.06)	(1.51)
Share Minority and Age <27	16.35*	-13.86*	-1.45	-8.15	-1.86	-0.17	3.72	0.34
	(8.83)	(7.61)	(10.27)	(5.64)	(2.46)	(1.7)	(3.24)	(1.66)
Share Female	-15.34**		-26.99**		3.51**		9***	
	(6.39)		(10.25)		(1.54)		(2.56)	
Joint F p-value	0.002		<0.001		0.377		<0.001	
B: Black and Non-Black Minorities								
Share Black and Age >27	-21.74	-0.57	-34.89***	-0.33	1.77	-0.63	12.13***	-1.87
	(13.22)	(18.34)	(11.23)	(10.54)	(2.37)	(2.54)	(3.71)	(1.94)
Share Non-Black Minority and Age >27	-5.52	-13.56*	-27.04***	-8.51	-2.15	2.52	10.33***	-1.27
	(6.62)	(7.58)	(7.87)	(5.29)	(1.54)	(1.77)	(2.4)	(1.82)
Share White and Age >27	-6.44		-21.69**		0.18		7.51***	
	(6.37)		(9.48)		(1.42)		(2.46)	
Share Black and Age <27	27.18	-12.06	-0.86	-12.02	5.17	-3.74	8.49	0.75
	(18.11)	(17.94)	(32.11)	(10.19)	(3.7)	(4.56)	(7.07)	(2.82)
Share Non-Black Minority and Age <27	12.4	-16.27	-0.58	-10.14	-3.64	0.31	3.51	-2.44
	(9.76)	(10.43)	(12.53)	(6.85)	(2.5)	(1.89)	(4.19)	(2.02)
Share Female	-14.02**		-25.62**		2.37		8.25***	
	(6.24)		(10.87)		(1.53)		(2.79)	
Joint F p-value	0.013		<0.001		0.021		<0.001	
C: Minority and White Female								
Share Minority and Age >27	-14.06**		-31.89***		0.24		10.36***	
	(6.61)		(6.17)		(1.25)		(1.92)	
Share White and Age >27	-9.55		-22.5***		-0.04		7.34***	
	(6.48)		(7.24)		(1.3)		(1.87)	
Share Minority and Age <27	9.45		-7.39		-1.71		4.82	
	(7.79)		(10.8)		(2.13)		(3.14)	
Share Minority and Female	-8.46	-15.95	-29.65**	-4.65	7.66***	-7.46***	12.55***	-2.75
	(8.41)	(9.66)	(11.68)	(6.86)	(2)	(2.4)	(2.83)	(1.72)
Share White and Female	-3.16	-17.57	-6.97	-14.78	1.93	1.1	2.5	2.94
	(15.05)	(17.42)	(18.26)	(11.27)	(2.73)	(3.23)	(5.05)	(2.96)
Joint F p-value	0.002		<0.001		0.011		<0.001	

Note: Table displays the OLS result for the effect of cohort composition on sample officers' average arrests and estimated officer effects, in units of arrests per 100 shifts using equation (1), with 2,296 observations. All regressions include exam period fixed effects and controls for officer-level characteristics corresponding to the included peer characteristics (e.g. an officer being a minority starting before age 27 in Panel A). Even columns contain interaction terms between an officer's race being white and cohort composition and additionally control for an officer being white. Officer effects are recovered from estimating equation (3). Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table B.6: Effect of Peer Composition on Co-Workers, Trainers, and Homerooms

	Co-Workers			FTOs			Officer Effects			
							Serious		Low-Level	
	Share Female	Share White	Mean Age	Share Female	Share White	Mean Age	Homeroom	Homeroom with Cohort FE	Homeroom	Homeroom with Cohort FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Share Minority and Age >27	0.08**	-0.1	0.09	0.06	0.06	-3.64	4.55***	1.27	-12.98***	-1.62
	(0.04)	(0.07)	(0.45)	(0.11)	(0.19)	(6)	(1.34)	(0.89)	(4.01)	(2.4)
Share White and Age >27	0.05*	-0.07	0.48	0.02	0.04	4.88	2.18**	0.37	-6.59	-0.1
	(0.03)	(0.07)	(0.4)	(0.09)	(0.19)	(5.21)	(1)	(1.2)	(4.32)	(2.19)
Share Minority and Age <27	0.04	-0.08	0.01	0.24	0.15	-3.42	3.24***	1.18	-6.53	-3.71
	(0.04)	(0.09)	(0.41)	(0.17)	(0.2)	(8.18)	(1.04)	(0.78)	(4.98)	(2.42)
Share Female	0.13***	-0.14*	-0.95*	-0.01	-0.02	-14.74***	4.75***	-0.9	-13.13**	-0.16
	(0.03)	(0.07)	(0.51)	(0.08)	(0.14)	(5.27)	(1.57)	(1.2)	(5.71)	(2.25)
N	2291	2291	2291	1732	1732	1732	2038	2038	2038	2038

Note: Table displays the OLS result for the effect of cohort composition on sample officers' co-worker composition (same sector, watch, and day), field training officer (FTO) composition, and average officer effects (in units of arrests per 100 shifts) using equation (1). All regressions include exam period fixed effects and controls for officer-level characteristics. Officer effects are recovered from estimating equation (3). Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition for Columns (1)-(6). Homerooms are sub-cohorts constructed using individual class training data as described in Appendix A.3. Homeroom shares (Columns (7)-(10)) are computed as the leave-out mean of the officer's homeroom's initial composition. Columns (8) and (10) include cohort fixed effects. Standard errors clustered at cohort level are in parentheses. ***p < 0.01; **p < 0.05; *p < 0.1

Table B.7: Robustness Tests for Average Arrests

	Poisson	LPM	First PO	FBI Index /Nonindex	Exam 2010	Exam 2013	Include Dropped	All Controls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: Low-Level Arrests								
Share Minority and Age >27	-193.8***	-10.7**	-7.3**	-11.2*	-14.4	-11.9	-14.6**	-17.6**
	(75)	(5.2)	(3.6)	(5.8)	(10.1)	(9.8)	(5.9)	(6.9)
Share White and Age >27	-120.8*	-7.4	-4.3	-6.7	1.2	-19.7	-9.8	-8.8
	(71.2)	(5.2)	(3.6)	(5.6)	(10.3)	(12.8)	(6.1)	(6.1)
Share Minority and Age <27	59.1	8.3	4.5	7	6.9	17.3	8.9	6.9
	(98)	(6.5)	(4.6)	(7.1)	(11.7)	(12.8)	(7.6)	(8.3)
Share Female	-155.3**	-12**	-7.1**	-10.7**	-19.8*	-17	-14.2**	-21.2***
	(61.4)	(5.3)	(3.3)	(5.2)	(10.9)	(14.5)	(6.6)	(6.4)
N	2296	2296	2296	2296	940	1112	2457	2296
B: Serious Arrests								
Share Minority and Age >27	-104.3**	0.7	0.7	0	-1.1	3*	-0.6	0.9
	(46.9)	(1.1)	(0.7)	(1)	(2.3)	(1.5)	(1.4)	(1.5)
Share White and Age >27	-70.7*	0	0.1	-0.7	0.8	-1.4	-0.6	-0.5
	(39.6)	(1.2)	(0.7)	(1)	(2.4)	(2)	(1.6)	(1.4)
Share Minority and Age <27	-61.7	-1.8	-0.8	-1.2	-0.6	-3.4	-1.8	-1.1
	(67.6)	(2.1)	(1.3)	(1.8)	(4.2)	(2.4)	(2.4)	(2.3)
Share Female	-42.2	3.2**	2.5***	1.3	5.5*	3.5	3.1*	3.1
	(38.1)	(1.5)	(0.7)	(1)	(3.1)	(2.6)	(1.6)	(1.9)
N	2296	2296	2296	2296	940	1112	2457	2296

Note: Table displays the OLS result from robustness tests with average arrests per 100 shifts as the outcomes from estimating equation (1), unless otherwise specified. All regressions include controls for officer-level indicators for group membership for all peer characteristics are included unless otherwise specified, and all regressions include exam period fixed effects. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. Column (1) presents results from estimating equation (5) as a Poison regression; Column (2) uses average number of shifts in which an officer made at least one arrest as the outcome; Column (3) uses average number of arrests per shift including only arrests where the officer was the first arresting officer; Column (4) reclassifies serious and low-level arrests as index and non-index based on the FBI UCR classification and excludes warrant arrests due to unknown crime types; Columns (5) and (6) subset to Exam 2010 and Exam 2013 officers only; Column (7) includes all officers in the initial cohorts for which average arrests could be recovered regardless of attrition; Column (8) includes controls for cohort shares and officer characteristics for spanish-speaking ability, military experience, and having a bachelor's degree or above. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table B.8: Robustness Tests for Officer Effects

	Poisson	LPM	First PO	FBI Index /Nonindex	Exam 2010	Exam 2013	Include Dropped	All Controls	MDSB	Unit-Role
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A: Low-Level Arrests										
Share Minority and Age >27	-406.2***	-24.6***	-18.3***	-28.1***	-33.1***	-31.2**	-33.4***	-33.6***	-33.8***	-23***
	(69.7)	(4.9)	(3.4)	(5.6)	(10.4)	(11.8)	(6.2)	(7.8)	(6.1)	(5.5)
Share White and Age >27	-313.2***	-16.6***	-11.5***	-18.6***	-10.2	-28.4*	-24.6***	-23***	-21.6***	-15.8***
	(78.2)	(5.6)	(3.9)	(6.4)	(12.2)	(15.1)	(7.5)	(7.3)	(7)	(5.7)
Share Minority and Age <27	-85.2	-3.5	-3.8	-5.7	-30*	18.9	-5.9	-5.1	-6	-1
	(115.1)	(8)	(5.7)	(9.2)	(14.6)	(16.1)	(10)	(12.1)	(10.4)	(9.2)
Share Female	-254.3**	-20.7**	-14**	-22.6**	-25.3	-41.9*	-25.1**	-33.8***	-26.9***	-22.3***
	(120.3)	(8.1)	(5.6)	(8.9)	(15.8)	(19.9)	(10.7)	(10.2)	(9.8)	(7.8)
N	2201	2296	2296	2296	940	1112	2456	2296	2296	2296
B: Serious Arrests										
Share Minority and Age >27	236.1***	8.4***	6.9***	11.3***	9***	11.5***	10.1***	10.4***	9.1***	13.3***
	(49.1)	(1.7)	(1.4)	(1.9)	(3.1)	(3.7)	(1.9)	(2.1)	(1.8)	(2.4)
Share White and Age >27	167.4***	5.4***	4.5***	7.5***	4	7.9*	7.2***	7.5***	5.8***	9.5***
	(44.1)	(1.5)	(1.3)	(2.1)	(3.3)	(3.8)	(1.9)	(1.9)	(1.6)	(2.5)
Share Minority and Age <27	77.6	2.9	2.5	4	9.8	-2.3	4.2	3.5	4.1	3.6
	(69.8)	(2.5)	(2.2)	(3)	(6.3)	(3.8)	(3.2)	(3.4)	(2.7)	(3.8)
Share Female	200.4***	7.1***	6.2***	8.3***	7.1	13.2***	8.8***	10***	7.1***	11.5***
	(57.2)	(2)	(1.7)	(2.8)	(5.4)	(4.1)	(2.7)	(2.8)	(2.3)	(3.3)
N	2222	2296	2296	2296	940	1112	2456	2296	2296	2296

Note: Table results from robustness tests with officer effects recovered from estimating equation (3), in units of arrests per 100 shifts, as the outcomes from estimating equation (1), unless otherwise specified. All regressions include controls for officer-level indicators for group membership for all peer characteristics are included unless otherwise specified, and all regressions include exam period fixed effects. Cohort shares are computed as the leave-out mean of the officer's cohort's initial composition. Standard errors clustered at cohort level are in parentheses. Column (1) presents results from estimating equation (1) using officer effects recovered from estimating equation (4) instead of equation (3); Column (2) uses whether an officer made an arrest during their shift as the outcome variable in equation (3) to recovered officer effects used as the outcome in equation (1); Column (3) uses officer effects from estimating equation (3) with first arresting officer arrests only as the outcome for equation (1); Column (4) reclassifies serious and low-level arrests as index and non-index based on the FBI UCR classification and excludes warrant arrests due to unknown crime types; Columns (5) and (6) subset to Exam 2010 and Exam 2013 officers only; Column (7) includes all officers in the initial cohorts for which officer effects could be recovered regardless of attrition; Column (8) includes controls for cohort shares and officer characteristics for spanish-speaking ability, military experience, and having a bachelor's degree or above. Column (9) uses officer effects recovered from re-estimating equation (3) using assignment fixed effects as described in Ba et al. (2021). Columns (10) uses officer recovered from re-estimating equation (3) with assignment effects broken into shift-year-month-day of week and unit-role-year effects. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table B.9: OLS, 2SLS, and SSIV Results

	Average Arrests per 100 Shifts			
	Low-Level		Serious	
	Arrests	Officer Effects	Arrests	Officer Effects
	(1)	(2)	(3)	(4)
A: Effect of Officer Characteristics (OLS)				
Minority and Age >27	-5.06***	-3.66***	-0.73***	-0.64***
	(0.72)	(0.45)	(0.15)	(0.14)
White and Age >27	-3.02***	-1.99***	-0.45**	-0.34**
	(0.66)	(0.49)	(0.17)	(0.15)
Minority and Age <27	-2.12***	-2.29***	-0.25	-0.14
	(0.71)	(0.4)	(0.21)	(0.21)
Female	-3.84***	-2.5***	-0.58***	-0.65***
	(0.5)	(0.36)	(0.15)	(0.13)
N	2296	2296	2296	2296
B: Instrumenting with Cohort Indicators (2SLS)				
Minority and Age >27	-14.14**	-24.89***	-0.33	7.79***
	(6.17)	(8.73)	(1.47)	(2.68)
White and Age >27	-13.1*	-21.03**	-1.44	5.35**
	(7.19)	(9.59)	(1.45)	(2.55)
Minority and Age <27	7.11	-8.34	-0.87	4.15
	(7.37)	(9.28)	(1.92)	(2.73)
Female	-14.13**	-14.44	2.26	4.56*
	(5.81)	(9.83)	(1.48)	(2.62)
Max First Stage F-Statistic	0.95	0.95	0.95	0.95
Max First Stage R2	0.016	0.016	0.016	0.016
N	2296	2296	2296	2296
C: Split-Sample IV (SSIV)				
Share Minority and Age >27	-16.88*	-43.69***	1.16	15.56***
	(9.29)	(15.41)	(2.94)	(5.67)
Share White and Age >27	-0.26	-15.25	2.02	9.69**
	(10.95)	(12.35)	(2.59)	(4.09)
Share Minority and Age <27	3.18	-42.37	-3.32	20.99
	(22.55)	(43.5)	(5.44)	(13.5)
Share Female	-3.56	3.71	1.36	-2.72
	(16.83)	(25.62)	(4.25)	(7.16)
N	1130	1130	1130	1130

Note: Table displays additional robustness tests relating to issues associated with estimating peer effects in Angrist (2014). All specifications include exam period fixed effects. Panel A provides effects of officer characteristics (corresponding to main specification peer characteristics) on main outcomes estimated with OLS. Panel B uses cohort indicators as instruments for cohort composition, with first stages that regress officer characteristics on cohort indicators, with the largest R2 and F-statistics across officer characteristics in the first stage reported. Panel C reports results from split-sample instrumental variables procedure in which each cohort is randomly split in half, and the composition of the first half is used as an instrument for the composition of the other half with officer-level controls included. Standard errors clustered at cohort level are in parentheses. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table B.10: Randomization Inference and Adjusted P-Values

	Low-Level Arrests				Serious Arrests			
	p-value	RI	Holm	BH	p-value	RI	Holm	BH
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: Outcome = Average Arrests								
Share Female	0.022	0.006	0.088	0.073	0.028	0.012	0.11	0.11
Share Minority and Age <27	0.218	0.143	0.295	0.218	0.398	0.283	1	0.796
Share Minority and Age >27	0.037	0.013	0.11	0.073	0.837	0.86	1	0.911
Share White and Age >27	0.147	0.09	0.295	0.196	0.911	0.917	1	0.911
B: Outcome = Officer Effects								
Share Female	0.012	0.001	0.025	0.016	0.001	0.001	0.002	0.001
Share Minority and Age <27	0.624	0.269	0.624	0.624	0.22	0.018	0.22	0.22
Share Minority and Age >27	<0.001	0.001	<0.001	<0.001	<0.001	0.001	<0.001	<0.001
Share White and Age >27	0.004	0.001	0.011	0.007	<0.001	0.001	0.001	<0.001

Note: Table displays p-values and adjusted p-values for the peer composition coefficients from the main specification. Columns (1)-(4) for low-level arrest outcomes, and repeated in Columns (5)-(8) for serious arrest outcomes, display p-values are computed from clustered standard errors at the cohort level (main results), 'RI' indicating randomization inference as discussed in Appendix A.3.5, 'Holm' indicating adjusted p-values using the Holm–Bonferroni method from Holm (1979) which controls for the family-wise error rate, and 'BH' indicating adjusted p-values using the Benjamini and Hochberg (1995) method which controls for the false discovery rate.

C Appendix C - Data

The data used in this study were obtained via FOIA request, in collaboration with the Invisible Institute and Chicago Data Collaborative, and generously shared by Rachel Ryley.

Demographics Data on officer demographics were obtained via multiple FOIA request to the Chicago Police Department. These data include information on officers extending as far back as the 1940's to the present (2021). The core demographic data includes name, race (ethnicity), start date, resignation date, and gender. Additional data sets relating to officer's language abilities were obtained for more recent officers (i.e., those in the data for this study), which were used to determine if the officer reported being able to speak Spanish. Similarly, whether or not an officer was in the military was also obtained for the present set of officers. Educational attainment records were also obtained indicating where, when, and what degree (if any) was obtained by each officer– this data is much less complete than other data sets but is most complete for officers starting around the Exam 2010 cohorts. For simplicity, educational data was summarized for this study as an indicator ("high edu") if the officer had reported obtaining a Bachelors degree or higher (e.g., masters, law degree, doctorate) before they started at the academy. The CPD's demographic data often combines race and ethnicity into a single variable. For expositional purposes and due to the data used, I classify 'Hispanic' as a distinct racial group.

Salary Salary data, obtained via FOIA to the Department of Human Resources, contains salary, pay grade (rank), and promotion information for officers between 2002 and 2020. This data is important as it allows us to focus on 'regular' police officers, i.e., D1 employees, and filter out promoted employees (sergeants, detectives, etc.). Importantly, this data contains officers' age at hire, allowing for very close approximation of their actual birth date and thus their exact age upon starting at the academy.

Unit History Officers' official unit assignments were obtained via FOIA to the CPD. This data indicates the dates on which an officer began and ended their official assignment to a specific unit.

Daily Assignments On a day to day basis, officers work specific beat assignments (alphanumeric codes that relate to function and location), are on specific watches, are or are not present for duty, are absent for some reason, are assigned to specific cars, and work between specific times. This information is contained within the daily assignment data, referred to in the text often as "AA" data. This data was obtained for the 22 (25 pre-2013) geographic units focused on in this study via FOIA request (for years 2010-2011 and 2016-2018) and shared by Rachel Ryley (for 2012-2015). Additional information on officer 'roles' were obtained via FOIA request to the CPD which gave descriptions of almost all beat assignment code to clarify their meaning.

Trainings A training data set, supplementary data set to the AA data, was obtained via FOIA request covering the period of the study. Specifically, this contains the name and start time of classes/trainings officers attended. This is particularly useful for identifying which officers were consistently trained together during the academy within their cohorts.

Arrests Data on adult arrests in Chicago were obtained via FOIA request to the CPD. This data includes arrestee information (race, age, gender), identifying officer information, arrest date and time, crime type and description, and the officer's arrest role (primary, secondary, or assisting). The arrest severity (Serious or Low-Level) is by crime type. Serious crimes include all violent and property index crimes, non-index property, and non-index violent crime (such as domestic violence and all forms of sexual assault). Index crimes are offenses on which the FBI collects data and tracks and publishes annually in the Uniform Crime Report (UCR). The eight index crimes are four violent and four property offenses: (violent) aggravated assault, robbery, murder, rape, (property) burglary, larceny, motor vehicle theft, and arson. For non-index crimes, I classify as 'serious', domestic violence is determined by whether the description indicates domestic battery or assault, and a few additional sexual assaults were classified based on whether the description indicates criminal sexual assault. Simple assaults and battery include crimes such as attempts at assault, child abuse, and threats of violence. I classify multiple types of deceptive practices as fraud. See *Crime* for crime code information. All other crimes (e.g., traffic, gambling, prostitution, drug) are considered low-level.

Court Court data from the Circuit Court of Cook County was obtained through collaboration with the Invisible Institute and Chicago Data Collaborative. This data is used to link specific arrests to cases and thus court outcomes (i.e., guilty finding, dropped case, etc.). It contains cases through 2019.

I define an arrest to be 'guilty' if the central booking number (CBN) is associated with any guilty finding; I consider an arrest not guilty if the CBN is associated with no guilty findings and at least one not guilty finding. If a CBN is associated with no guilty findings and no not guilty findings, and it has any dismissed cases, then I consider it dismissed. If a CBN does not appear in the court data, I classify the case as dropped. I group not guilty, dismissed, and dropped cases together and label them as 'non-guilty'. If a CBN is not classified as guilty, not guilty, or dismissed, but it is in the court data, then it only has incomplete/open cases, so it is classified as neither guilty nor non-guilty. A single CBN may have multiple charges or cases associated with it, and I use the method discussed above to provide a single outcome of an arrest which is conservative as only one guilty verdict on any charge is sufficient for an arrest to be 'guilty'.

Population Information on district populations for each year is obtained from American Community Survey 5-Year data, with census tracts spatially overlayed onto CPD districts using public district maps.

Crime Raw crime data is obtained from the Chicago Data Portal, downloaded in August of 2020. Crime is classified based on FBI codes into violent, property, and other crime. Violence-related crime FBI codes are 1A/B (homicide/manslaughter), 2 (criminal sexual assault / rape), 3 (robbery), 4A/B (aggravated assault/battery), 8A/B (simple assault/battery). Property-related crime FBI codes are 5 (burglary), 6 (theft), 7 (motor vehicle theft), 9 (arson), 10-13 (deceptive practices/fraud/stolen property), 14 (criminal damage). Index crime codes are 1A, 2, 3, 4A, 4B, 5, 6, 7, 9. All other crimes are classified as other and non-index, e.g., prostitution, gambling, trespassing, narcotics. Arrest data have the same classifications using FBI codes.