ONLINE APPENDIX:

Do Peers Matter in the Police Academy?

Roman Rivera

May 29, 2024

Contents

1 Additional Analyses

1.1 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).[1]

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 2.1, 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 2.1). 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)

[1] 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)

cohorts are an exception, and these cohorts all came from the same exam issued in December 2010 (see Figure 2.1). 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, 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 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 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 1.3 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 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 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.

1.2 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 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 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.

Figure 1: Composition of Cohorts by Start Date

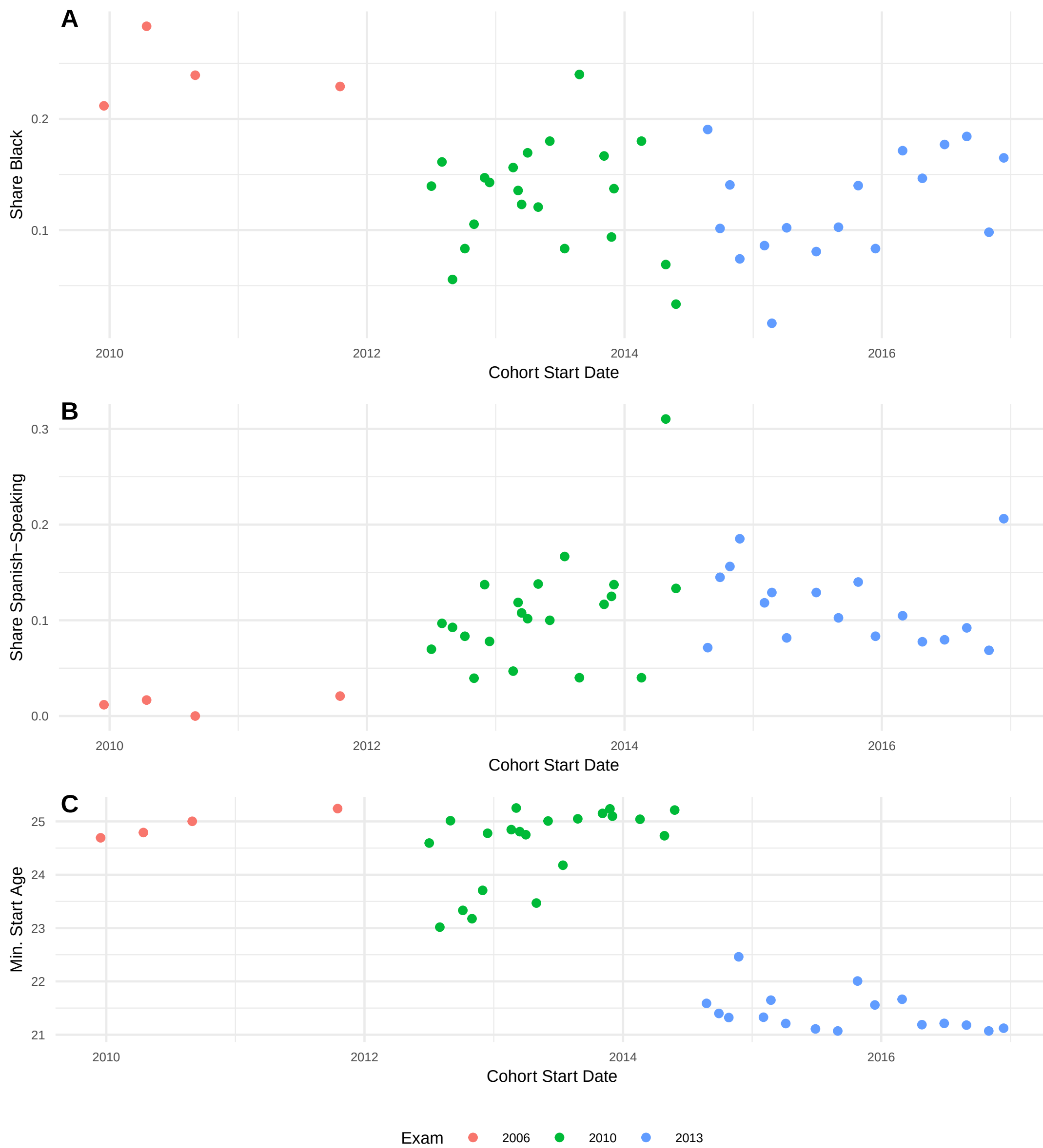

Figure displays the share of cohorts with more than 10 starting members that are Black (Panel A) and speak Spanish (Panel B), and the lowest starting age (Panel C) by the cohort start date, from July 2009 to 2016. Exam denotes the time period during which the cohorts started and assumes cohorts in the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

Table 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

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 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

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. $***p < 0.01$; $**p < 0.05$; $*p < 0.1$

1.3 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 in the main text. Table 2.7 presents results for average arrest outcomes, and Table 2.8 presents results from analogous tests for officer arrest propensity outcomes when feasible.

1.3.1 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) in the main text) using a fixed-effects Poisson regression and then re-estimate equation (1) in the main text 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,

$$E[Arrest_{it}|\theta_i, \gamma_{brsw_t}, V_{it}] = exp(\theta_i + \gamma_{brsw_t} + \beta V_{it}) \tag{1}$$

for recovering fixed effects, and

$$E[Arrest_i|X_i, \overline{X}_{c(-i)}, \eta_{p(i)}, N_i] = exp(\pi_2 \overline{X}_{c(-i)} + \pi_1 X_i + \eta_{p(i)} + N_i) \tag{2}$$

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 2.7 and 2.8 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) in the main text 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.[2] The coefficients using these alternate definitions (Column (4) of both tables) are generally consistent with the main results.

1.3.2 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 2.8 using officer effects recovered from re-estimating main text equation (3) using alternative assignment definitions. First, I repeat my analysis using the more granular assignment fixed effects ('MDSBs') from Ba et al., 2021.[3]

[2]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.

[3]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.

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) in the main text. 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.

1.3.3 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 2.5 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 in the main text). 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 in the main text. These results indicate that the 27-year cutoff is not spuriously producing economically significant results though the precision of the estimates changes with cutoffs.

1.3.4 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 2.6 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 2.9, 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 Alan B. 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 A. B. 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.

1.3.5 Inference

The main specification has multiple variables of interest, requiring adjustments for multiple hypothesis testing. Table 2.10 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 2.10 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) in the main text (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 $Y_i(P) = Y_i(P') \forall P, P' \in P$ where P' is any possible peer composition and 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:

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

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.

1.4 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.

2 Additional Figures and Tables

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

Figure 2.1: CPD Exam Information

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.

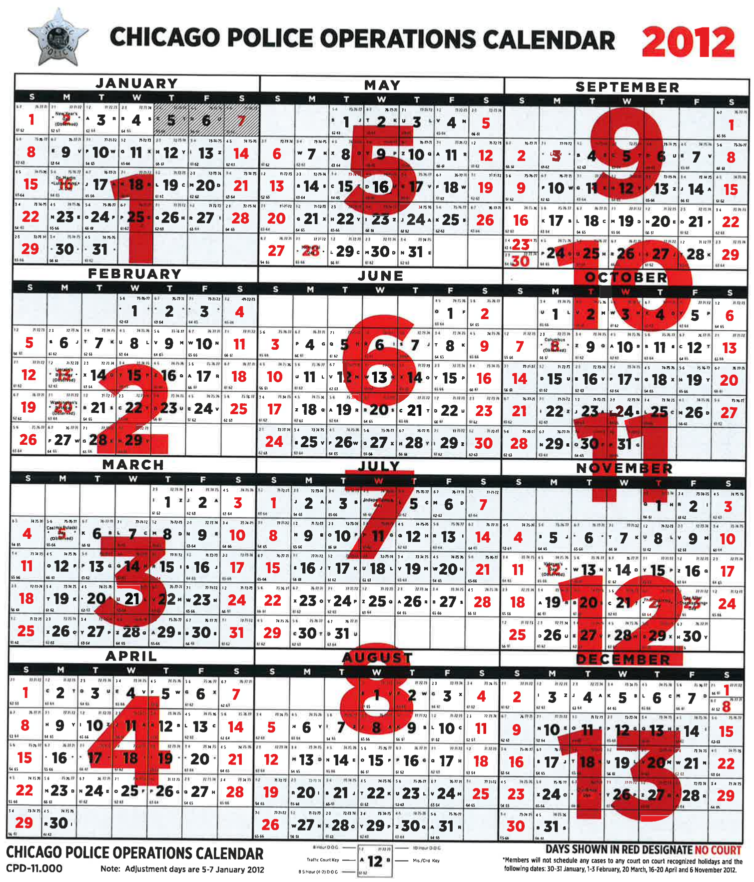

Figure 2.2: CPD Operations Calendar (2012)
Figure displays an example of the CPD operations calendar for the year 2012.

Table 2.1: 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

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.

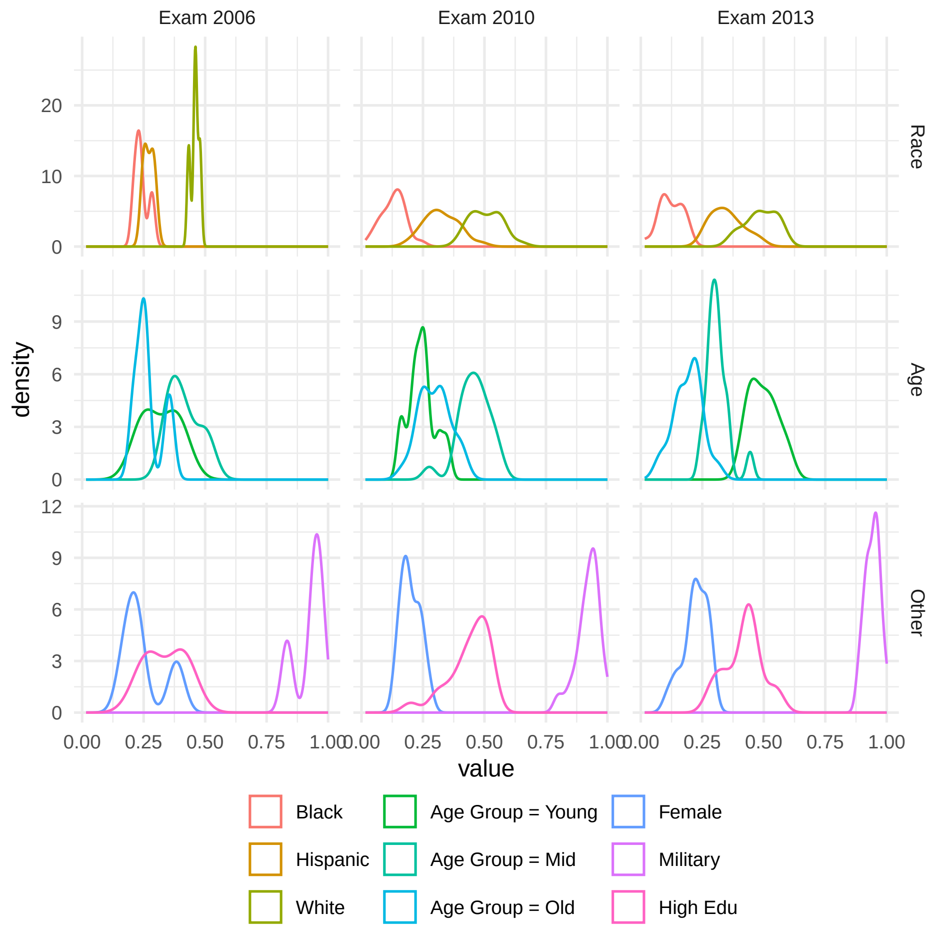

Figure 2.3: Cohort Composition

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).

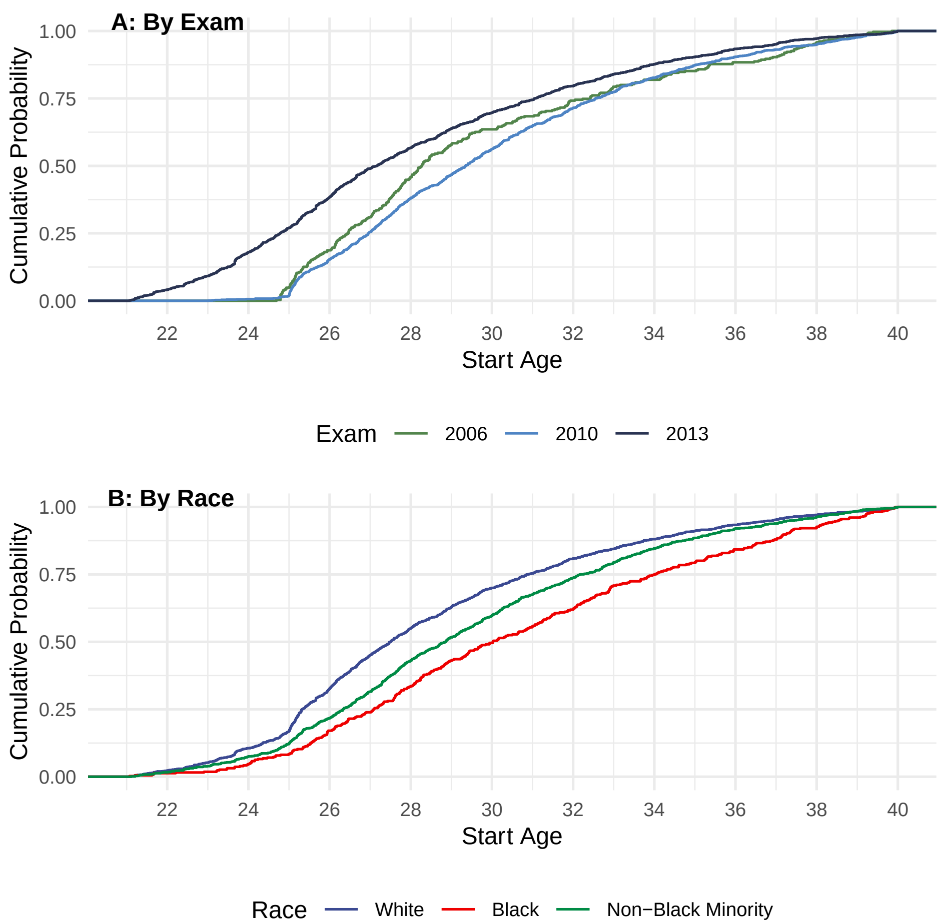

Figure 2.4: CDF of New Officer Start Ages

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.

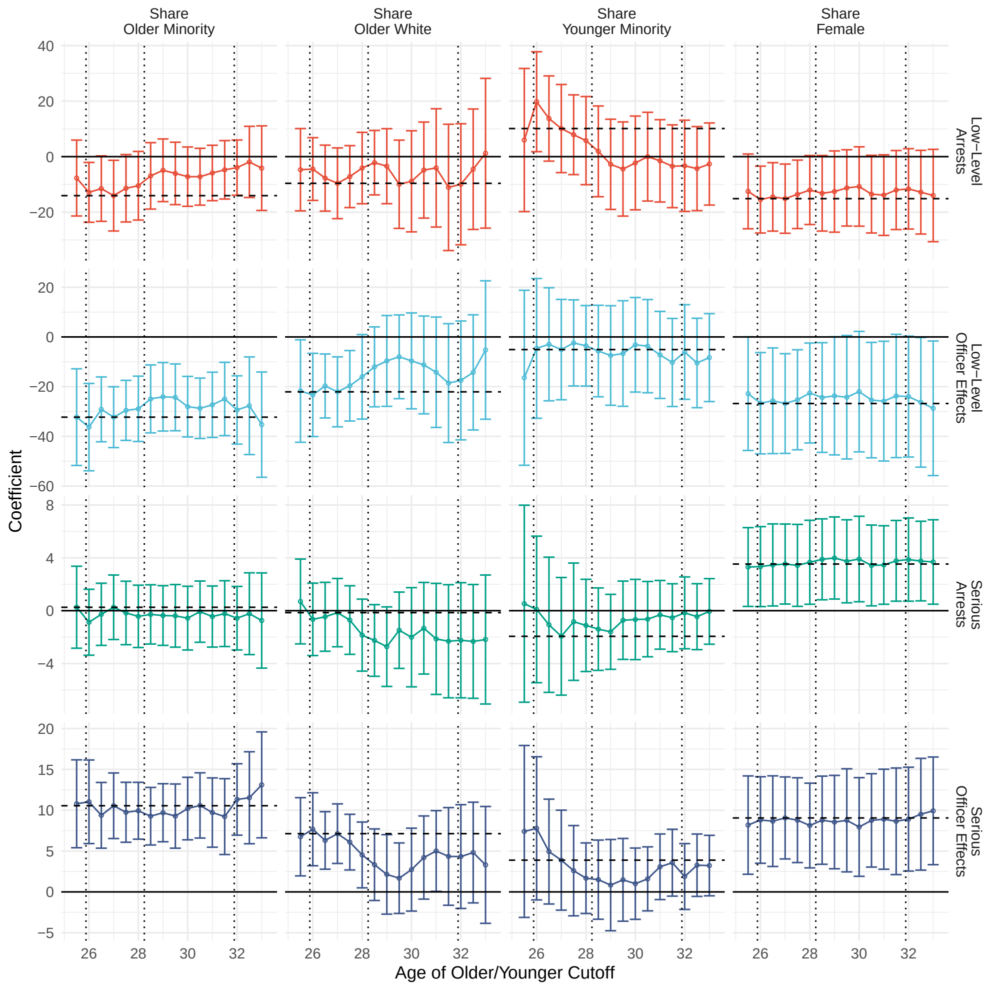

Figure 2.5: Change in Coefficients with Alternative Age Cutoffs

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.

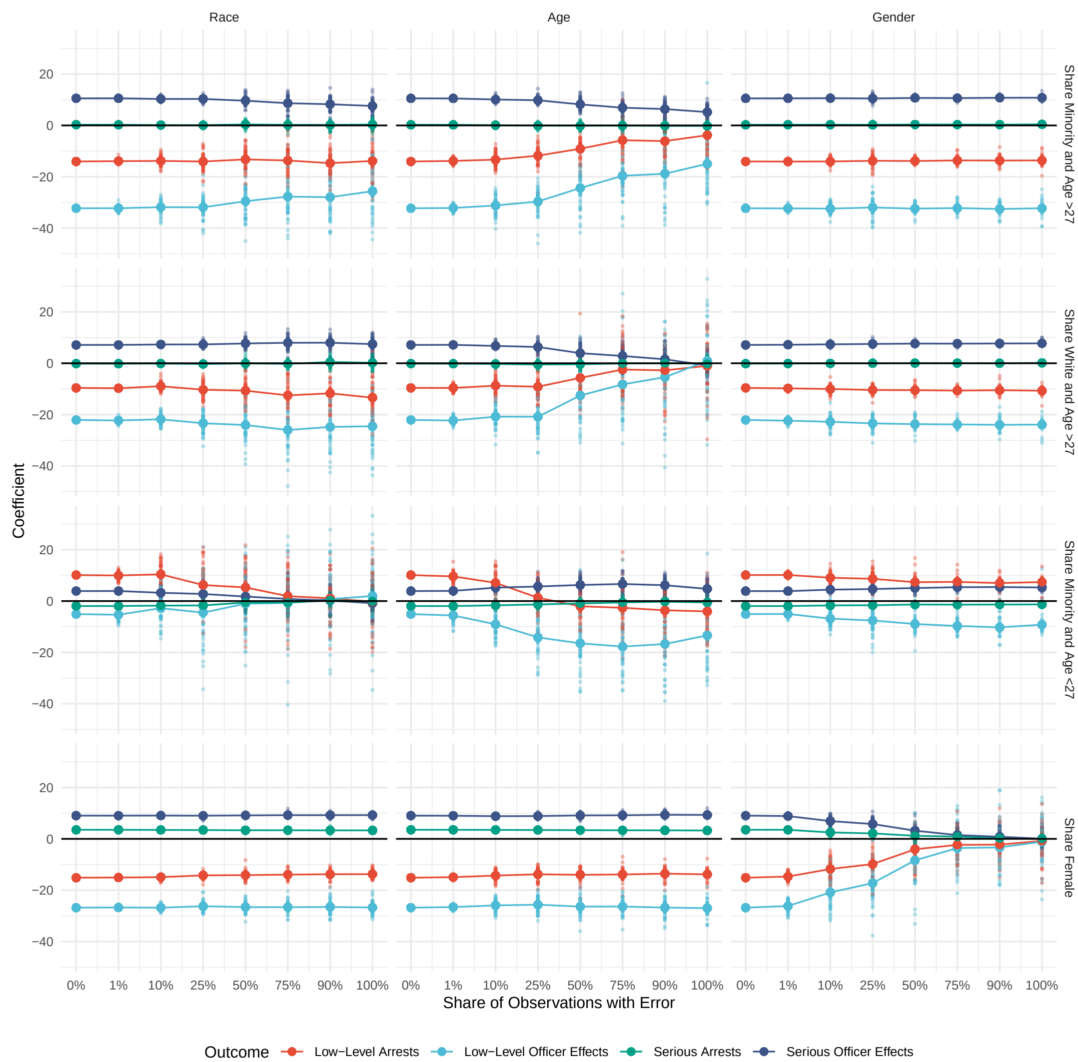

Figure 2.6: Change in Coefficients with Measurement Error

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 main text equation (3). Coefficients estimated using main text equation (1) using controls for officer group membership and exam period fixed effects.

Table 2.2: 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

Table presents correlations across officer demographics within exam periods.

Table 2.3: 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)

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 main text 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 main text 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 2.4: 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)

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 main text equation (1), with 2,296 observations. All regressions exam period fixed effects and controls for officer-level characteristics. Officer effects are recovered from estimating main text equation (3). See Appendix 3 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 2.5: 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	

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 main text 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 main text 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 2.6: 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

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 main text equation (1). All regressions include exam period fixed effects and controls for officer-level characteristics. Officer effects are recovered from estimating main text 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 1.3. 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 2.7: 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

Table displays the OLS result from robustness tests with average arrests per 100 shifts as the outcomes from estimating main text 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 (2) 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 2.8: 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

Table results from robustness tests with officer effects recovered from estimating main text equation (3), in units of arrests per 100 shifts, as the outcomes from estimating main text 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 main text equation (1) using officer effects recovered from estimating equation (1) instead of main text equation (3); Column (2) uses whether an officer made an arrest during their shift as the outcome variable in main text equation (3) to recovered officer effects used as the outcome in main text equation (1); Column (3) uses officer effects from estimating main text equation (3) with first arresting officer arrests only as the outcome for main text 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 main text equation (3) using assignment fixed effects as described in Ba et al. (2021). Columns (10) uses officer recovered from re-estimating main text 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 2.9: 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

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 2.10: 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

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 1.3.5, '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.

3 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.