The Relationship between Officer Misconduct and Conviction-less Arrests*

Bocar A. Ba† Nayoung Rim‡ Roman Rivera§

December 11, 2024

Given the use of an individual's arrest history for many economic and social outcomes, reducing conviction-less arrests (arrests that result in no charges or where the defendant is found not guilty) is an important policy goal. This paper examines which officers are making conviction-less arrests, and whether these arrests can be reduced with increased oversight. Using the Chicago Police Department's rotational duty calendar to obtain plausibly exogenous variation in the set of officers assigned to work on a particular day, we find that arrests made by officers with high misconduct are 10.5% less likely than the arrests made by no-misconduct officers to result in charges and are 14% more likely to have a "Not Guilty" outcome. We also analyze two events that increased the transparency of police misconduct through public disclosure of complaint records and find that increased oversight reduces conviction-less arrests, but with important nuances across misconduct profiles. While no- and low-misconduct officers appear more responsive to oversight mechanisms, high-misconduct officers may require more targeted interventions.

JEL Codes: K42, D73, J18
Keywords: police, misconduct, complaints, arrest quality, oversight

*We thank the Russell Sage Foundation and the Laura and John Arnold Foundation for financial support. The views expressed herein are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S Government. All errors are our own.

†Duke University, Department of Economics and NBER. Email: bocar.ba@duke.edu

‡U.S. Naval Academy, Department of Economics. Email: rim@usna.edu

§Princeton University. Email: romgariv@gmail.com

1 Introduction

There is extensive research on the negative consequences of entering the criminal justice system. For example, being arrested increases the probability of dropping out of high school (Hirschfield, 2009) and decreases the likelihood of receiving a college acceptance (Stewart and Uggen, 2020). In fact, many non-criminal justice actors—such as employers, public housing authorities, child protective service providers, and immigration enforcement officials—rely on arrest records for their own decision-making processes (Jain, 2015). For example, people with criminal records are ineligible for public housing and housing choice vouchers (Carey, 2004). In the employment sector, the vast majority of employers require background checks as a condition of employment[1], with some even receiving automatic notifications when an employee is arrested and fingerprinted.[2] These consequences are particularly alarming for arrests that never resulted in a conviction or where charges were dropped, as these records still remain on file and appear in background checks.[3]

Policymakers have implemented criminal remediation laws as a potential solution, but evaluations of these policies point to the need for an earlier intervention in the criminal justice process due to the scarring effect from having a criminal record (e.g., Agan et al., 2024b; Mueller-Smith and T. Schnepel, 2021). Building upon this insight, our paper examines a critical upstream factor: the role of the law enforcement officer. We examine which officers are making conviction-less arrests (defined as arrests that result in no charges or where the defendant is found not guilty), and whether these arrests can be reduced with increased oversight.

In the first part of the paper, we establish that officers with greater misconduct make different types of arrests than those with no misconduct. Not only are officers with misconduct 14%-79.6% more likely than officers with no misconduct to make an arrest, in general, but they are also 24%-132.5% more likely to make an arrest for less serious (non-index) crimes. More importantly, the arrests made by officers with misconduct are 4.9%-10.5% less likely to result in charges and are 2.8%-14% more likely to have a "Not Guilty" outcome compared to the arrests made by officers

[1] According to a 2012 survey by the Society for Human Resource Management, 86% of employers conducted criminal history checks for at least some jobs and 69% conducted checks on all job applicants.

[2] A specific example of this practice can be seen in New York, where arrest data are automatically transmitted to the NYC and NYS Departments of Education and the NYC Taxi and Limousine Commission (Jain, 2015, p. 839).

[3] See https://www.wsj.com/articles/as-arrest-records-rise-americans-find-consequences-can-last-a-lifetime-1408415402. Further, despite legal provisions allowing individuals to request record expungement, the actual take-up remains remarkably low. Prescott and Starr (2019) find that in Michigan, only 6.5% of those legally eligible obtained expungement within five years of becoming eligible.

with no misconduct.

To obtain plausibly causal estimates, we compare Chicago Police Department (CPD) officers working on the same shift but with differing levels of misconduct. A shift is defined as a unique combination of the day of the week + month-year (e.g., January 2014) + district + sector (which identifies the geographic area where the officer is working) + watch (which identifies the time of day that the officer is working). We measure misconduct by the total number of complaints made against the officer in the previous calendar year. We argue that we estimate causal estimates because the set of officers working on the same shift is as-good-as-randomly determined due to the rotational duty calendar.

After establishing these facts on the relationship between officer misconduct and arrest patterns, we turn to analyzing a policy change in the second part of the paper. Given the observed differences in arrest behavior between officers with varying levels of misconduct, a natural question is: can increased oversight influence officer behavior and potentially reduce these conviction-less arrests? To answer this question, we analyze two events that increased the transparency of police misconduct and, consequently, oversight.

On March 10, 2014, the Illinois Supreme Court ruled in *Kalven v. City of Chicago* that police-misconduct records are public information under the Illinois Freedom of Information Act (FOIA). Prior to this ruling, officer misconduct records were not public information. The court decision altered how much information about complaints against individual police officers could be seen by the public. This change was expected to impact policing practices. The Chicago police union's reaction showed this, with their president writing a front-page letter in the August 2014 newsletter discussing the ruling and its effects on CPD officers.

We conduct an event study to examine how officers responded to these two policy changes. We find that complaints fell by 37%-60% for all officers, with the largest drop among high-misconduct officers after the FOP newsletter. Regarding conviction-less arrests, there is a distinction between the court ruling and the union president's letter.

Immediately after the court ruling, officers reduce conviction-less arrests by changing their arrest behavior compared to the arrests these same officers made before the court ruling. The effects are strongest for no-misconduct officers and weakest for high-misconduct officers. After the FOP newsletter, however, we see heterogeneous effects by misconduct level. No-, low-, and

medium-misconduct officers change the types of arrests they make, potentially focusing on incidents that result in more charges—with the strongest effects for no-misconduct officers and weakest for medium-misconduct officers. In contrast, there is suggestive evidence that high-misconduct officers are withdrawing from policing after the union president's letter.

Our findings reveal that increased transparency of officer misconduct can reduce conviction-less arrests, but with important nuances across officer misconduct profiles. The effect is most pronounced among officers with no or low misconduct, suggesting that a blanket transparency intervention may have limited effectiveness for officers with high-misconduct. These results are in line with prior research indicating that police oversight can effectively influence officer behavior (Cheng and Long, 2018; Mummolo, 2018; Rivera and Ba, 2023; Rozema and Schanzenbach, 2023; Campbell, 2024). Notably, while existing studies typically examine aggregate effects across all officers, our paper provides a more granular analysis by distinguishing between high-, medium-, low-, and no-misconduct officers. This approach enables us to offer more targeted insights for developing nuanced oversight policies that can more precisely address variations in officer behavior and reduce conviction-less arrests.

This paper contributes to research on methods to mitigate the negative consequences of having a criminal record. Past studies have examined the effectiveness of various criminal remediation policies, ranging from criminal-record expungement (Jackson et al., 2017; Selbin et al., 2018; Prescott and Starr, 2019; Agan et al., 2024b) to the removal of criminal history questions from job applications (Agan and Starr, 2018; Craigie, 2020; Doleac and Hansen, 2020; Rose, 2021) to downgrading felonies to misdemeanors (Agan et al., 2024a) or not prosecuting misdemeanors (Agan et al., 2023). These studies find improved criminal justice outcomes and mixed employment outcomes. Importantly, the results collectively underscore a scarring effect from the criminal record (see also McWilliams and Hunter, 2021; Cullen et al., 2023).[4] We build on this and examine the role of the officer, an important upstream factor in the criminal justice process.

This paper also expands the current literature on alternatives to traditional arrest methods. Several counties and states have implemented reforms aimed at reducing unnecessary arrests and their

[4] McWilliams and Hunter (2021) finds that criminal record-related stigma negatively impacts the individual's quality of life, pointing to the importance of reducing criminal record stigma and discrimination. Relatedly, Cullen et al. (2023) implements a field experiment to test whether at what "price" businesses would be willing to hire workers with a criminal record, thereby mitigating the scarring effect.

negative impact on individuals. For example, an examination of Florida's Juvenile Civil Citation Program, which recommends police officers issue a civil citation instead of making an arrest for first-time misdemeanor offenses, finds that juvenile recidivism declined (Nadel et al., 2018; Krishna, 2024). Mueller-Smith and T. Schnepel (2021) examines two policy changes in Harris County, Texas that changed diversion rates and finds that diversion from a felony conviction for first-time offenders reduces future convictions and increases employment rates and earnings.[5] These findings also point to the promise of alternative response models, which are slowly gaining traction (Bell, 2021; Dee and Pyne, 2022; Ba et al., 2024). We examine two policy changes that increased the transparency of misconduct allegations and find that increased oversight reduces conviction-less arrests for most officers but less so for high-misconduct officers.

More broadly, this paper also relates to work that examines different dimensions of law enforcement practices and their impact on society (Heaton et al., 2017; Ouss and Rappaport, 2020; Ang et al., 2024; Cunningham et al., 2024). For example, Cunningham et al. (2024) finds that collective bargaining rights appear to create protective mechanisms that may shield problematic officers from meaningful accountability. As Ouss and Rappaport (2020) documents, there is an increasing societal intolerance of policing harms, highlighting the need for reform. Developing practices that strengthen community-police relations is important, especially as an erosion of trust in the police can have important ramifications for public safety (Ang et al., 2024; Sánchez De La Sierra et al., 2024). This paper examines the possibility of using officer misconduct as a policy lever to reduce conviction-less arrests.

The rest of the paper is organized as follows: Section 2 describes the data and provides background on CPD patrol assignments. Section 3 describes our identification strategy. Section 4 discusses our results, and Section 5 discusses additional analyses to explore heteroengeous effects. Section 6 presents the policy analysis on the court ruling and the Chicago police union president's letter. Section 7 concludes.

[5] Specifically, the diversion program in Texas is a deferred adjudication agreement wherein the defendant admits guilt without receiving a criminal conviction record and completes a probationary period of community supervision.

2 Data and Background

2.1 Data and Summary Statistics

Officer personnel data was acquired from the Chicago Police Department through Freedom of Information Act requests. We construct a panel dataset that contains a continuous record of officer shift assignments, complaints, and arrests between 2010 and 2016. We link officer arrests to court records, which was acquired through the Chicago Data Collaborative. The analysis dataset has 3.2 million observations on 8,843 patrol officers between 2010 and 2016.

Table 1 presents descriptive statistics on the analysis sample. The first row lists the number of unique officers in each misconduct category. These are not mutually exclusive categories as an officer may have 0 complaints one year and 3 complaints the next year. In terms of percentages, on average, 66 percent of officers between 2010-2016 have 0 complaints in a year, 22 percent have 1 complaint, 7 percent have 2 complaints, and 4 percent have 3 or more complaints. To reiterate what the data are showing, complaints are not common; an officer at the 90th percentile of complaints has 2 complaints a year.

Table 1: Descriptive Statistics

		Number of complaints last year:			
	Overall	0	1	2	3+
Unique number of officers	8,843	8,217	4,711	2,045	1,072
Average share of officers	1.000	0.664	0.222	0.073	0.041
Share with any arrest	0.053	0.046	0.055	0.071	0.095
Mean number of arrests	1.104	1.093	1.106	1.125	1.151
	(0.382)	(0.364)	(0.377)	(0.411)	(0.472)
Mean number of verdicts	1.051	1.046	1.044	1.059	1.103
	(0.799)	(0.786)	(0.784)	(0.844)	(0.870)

Notes: This table depicts summary statistics on the analysis sample of all Chicago Police Department patrol officers between 2010 and 2016. "Unique number of officers" is the unique number of officers between 2010-2016 in the relevant misconduct category. "Average share of officers" is the average over years. "Share with any arrest" is the share of all officer-shifts between 2010-2016 with any primary arrests. The mean number of arrests and verdicts are conditional on making an arrest. Standard deviations are in parentheses.

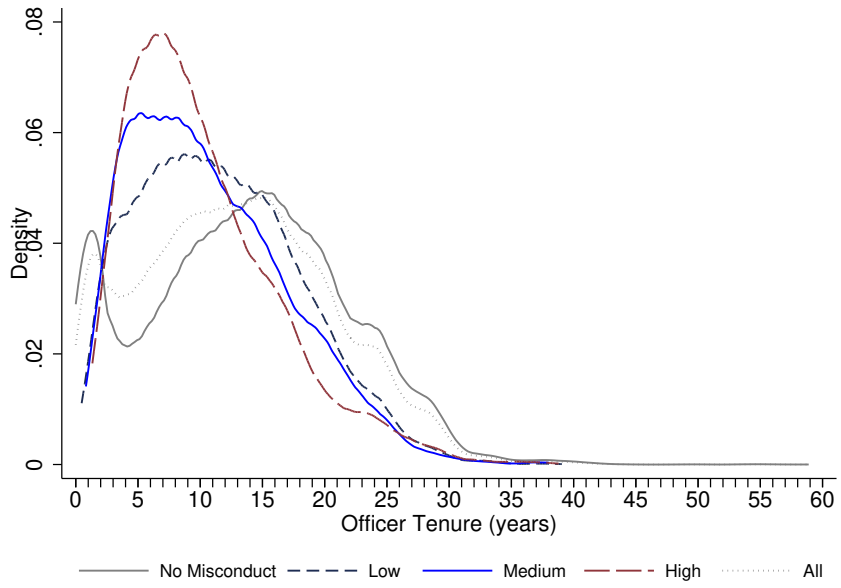
About 5 percent of officer-shifts between 2010-2016 have at least one arrest a shift. This translates to an average of 68 arrests a shift across all of Chicago. On any given shift, 5 percent of patrol officers with no complaints last year make at least one arrest as the primary arresting officer, while

9.5 percent of officers with 3 or more complaints last year on shift make at least one arrest as the primary arresting officer. These numbers suggest that officers with more complaints may be more active when they are on duty, which makes sense as complaints and arrests are positively correlated.

Conditional on making an arrest, the average officer makes 1.1 arrests per shift. This is not too different across the different misconduct levels. An officer with 3 or more complaints last year makes 1.15 arrests compared to the 1.09 arrests made by an officer with no complaints. This suggests that the heterogeneity is coming from the extensive margin, the likelihood of making an arrest, rather than the intensive margin.

The last two rows in Table 1 list the mean number and standard deviation of verdicts. Because an arrest can have multiple charges, the number is greater than one. Since we do not have charge data, we use verdicts as a proxy for charges. Most arrests have about about one verdict, which is slightly higher for officers with more misconduct.

Figure 1: Distribution of Officer Tenure by Misconduct Level



Notes: This figure depicts the distribution of officer tenure by misconduct level.

Figure 1 depicts the distribution of officer tenure by misconduct level. The dotted line, which is for all officers, indicates that a plurality of officers are between 8-20 years of experience in our data. If we look separately by misconduct level, most of the officers who had high misconduct are

relatively junior–between 5-10 years of experience (maroon, dashed line). Last, the no-misconduct officers tend to have more experience, with a hump around 16 years of experience (navy, dashed line).

2.2 Background on CPD patrol assignments

There are three components to an officer's assigned work environment: the geographic district or where they work; the watch or the time of day they work; and the calendar day that they work.

Chicago is geographically divided into 22 districts (see Appendix Figure A1). Officers are initially assigned to one of these districts based on available vacancies and departmental needs.[6] Districts are subdivided into beats. Our analysis works with sectors, which are aggregated beats. Sectors are determined by the second digit in a beat's ID number. There is an average of 3-4 beats per sector, and 3 sectors per district.

Officers are assigned to one of three watch periods, which are 8-hour long periods. Watch 1 is from midnight to 8 am, Watch 2 is from 8 am to 4 pm, and Watch 3 is from 4 pm to midnight. Officers bid every fall for their preferred watch period for the next calendar year. The bidding process for watches is based on seniority. Most officers are on Watch 1 and 3, while more senior officers are most likely to be on Watch 2 (Appendix Figure A2).

Officers are also assigned to "day-off groups", which determine who works on rotating calendar dates. This schedule is set late in each calendar year for the following year. A notable feature of this system is that cycles do not occur on a weekly basis; a typical duty cycle is four days "on" and two days "off". This means that the set of officers working on, say, Mondays in November will not be the same. Further, because operations schedules are made far in advance, officers effectively have no ability to anticipate fluctuations in crime or civilian behavior while on duty, conditional on assigned district and watch. In other words, officers are unable to pre-emptively choose to take days off when they expect crime to spike.

[6] After the initial assignment, an officer may transfer to another district if there are vacancies in that district. The transfer process relies on a bidding process wherein the winning bid goes to the most senior officer. In reality, there are few transfers. For example, between 2010-2016, 67% of officers did not change districts.

3 Identification Strategy

3.1 Summary

We use the CPD's rotational duty calendar to obtain plausibly exogenous variation in the set of officers assigned to work on a particular day. Specifically, we compare all officers working on the same shift but with different levels of misconduct last year. A shift is defined as a unique combination of the day of the week + month-year (e.g., January 2010) + sector (which identifies the geographic area where the officer is working) + watch (which identifies the time of day that the officer is working). Misconduct is measured by the total number of complaints made against the officer in the previous calendar year. Comparing officers on the same shift ensures that we compare the behavior of officers in the same, very narrowly defined working conditions. This design also ensures that officers have the same opportunity to take enforcement action; failure to account for officers' inactivity may lead to biased conclusions (Knox et al., 2020).

The claim is that the level of misconduct is as-good-as-randomly assigned across officers working the same shift because rotations are out of an officer's control. Therefore, the *set of officers* working on the same shift is not determined by the officer's level of misconduct.

3.2 Exogeneity checks

To check whether our identification strategy is valid, we conduct three tests.[7] First, we examine whether officers are able to choose which day of the week they work. We calculate the total number of complaints an officer received in one week and see if it can predict which day of the week the officer works the following week. Table 2 shows that officers with more misconduct are not able to choose which days they work. All of the coefficient estimates are trivial in magnitude and not statistically significant.[8] In addition, the R-squared is very low, ranging from 0.02% to 0.4%. This is consistent with our argument that, since these officers work 4 days on / 2 days off, they are not consistently working the same days each week.

[7] We conduct these tests on patrol officers, which is our analysis sample, rather than on all officers. We replicate these tests on all police officers in Appendix Tables A2, A3, and A4. The results are the same.

[8] The coefficient estimate for Friday is weakly significant at the 10% level, but the magnitude is trivial (0.1 percentage point).

Table 2: Does misconduct predict which day of the week an officer works?

Outcome Variable:	Monday (1)	Tuesday (2)	Wednesday (3)	Thursday (4)	Friday (5)	Saturday (6)	Sunday (7)
Total Complaints Last Week	2.80e-05	0.000312	0.00124	0.000174	-0.00117*	6.28e-05	-0.000647
	(0.000620)	(0.000479)	(0.000789)	(0.000476)	(0.000584)	(0.000855)	(0.000636)
Constant	0.139***	0.142***	0.144***	0.144***	0.145***	0.144***	0.141***
	(2.21e-05)	(1.74e-05)	(3.17e-05)	(1.54e-05)	(2.19e-05)	(3.54e-05)	(2.22e-05)
Observations	3,146,453	3,146,453	3,146,453	3,146,453	3,146,453	3,146,453	3,146,453
R-squared	0.004	0.002	0.002	0.002	0.002	0.002	0.003

Notes: This table reports estimates for the total number of complaints last week on the probability of working on a specific day of the following week. All estimates also include fixed effects for officer, district, and year. Standard errors clustered by officer and district are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Second, we examine whether officers are able to choose which days to work based on expected crime. Although the rotational calendar is out of the officer's control, officers are able to request certain holidays off. Therefore, one potential concern to our analysis is that officers with more misconduct may be choosing to work on days where the crime rate is expected to be higher. This would lead to more potential opportunities for arrests, which leads to more opportunities for complaints made against the officer.

To check this, we calculate the officer's total number of complaints last year and see if it can predict whether they work on holidays the subsequent year that tend to experience the largest number of violent crimes.[9] Table 3 shows that officer misconduct is orthogonal to working on these holidays, suggesting that officers with greater misconduct are neither choosing to nor not choosing to work on these days.[10]

[9] The reason for the asynchronous timing is because officers' work schedules for a given calendar year are set in the fall of the previous year.

[10] The first three holidays are listed in decreasing order of gun violence. St. Patrick's Day and New Year's were included because they typically involve lots of alcohol and, thus, arguably more opportunities for police involvement.

Table 3: Are higher misconduct officers able to choose which days to work based on expected crime?

Outcome Variable:	Memorial Day (1)	Fourth of July (2)	Labor Day (3)	St. Patty's Day (4)	New Year's (5)
Total Complaints Last Year	-0.000800	-0.00189	-0.00401	0.00255	0.00388
	(0.00410)	(0.00315)	(0.00401)	(0.00398)	(0.00287)
Observations	26,207	26,207	26,207	26,207	26,207
Mean Outcome	0.248	0.237	0.250	0.244	0.411

Notes: This table reports estimates for the total number of complaints last year on the probability of working on a specific holiday the following year. New Year's includes working on New Year's Eve and New Year Day. The sample is at the officer-year level. All estimates also control for officer tenure at the start of the year and include fixed effects for officer, lagged modal district, and year. Standard errors clustered by officer and district are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Last, we examine whether misconduct is correlated with an officer's assigned watch. Similar to holiday requests, watch assignments are determined in the previous calendar year. Therefore, we check whether the total number of complaints last year can predict which watch an officer is on this year. Table 4 shows that the estimates are all trivially small in magnitude and not statistically significant, which suggests that officer misconduct is orthogonal to the assigned watch. Taken together, these three tables support our claim that the set of officers who are working on a given day, time, and location—or "shift"—is as good as randomly assigned with respect to misconduct level.

Table 4: Does misconduct predict watch?

Outcome Variable:	Watch 1 (1)	Watch 2 (2)	Watch 3 (3)
Total Complaints Last Year	-0.00118	0.00157	-0.000998
	(0.00274)	(0.00304)	(0.00303)
Observations	18,725	18,725	18,725
Mean Outcome	0.347	0.334	0.442

Notes: This table reports estimates for the total number of complaints last year on the probability of working a given watch the following year. The sample is at the officer-year level. All estimates also control for officer watch last year, officer tenure at the start of the year, officer total arrests two years ago, and include fixed effects for officer, modal district, and year. Standard errors clustered by officer and district are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

In Appendix Tables A2, A3, and A4, we conduct these checks using a categorical measure of misconduct instead of as a linear measure. In Panel A, we compare officers with medium/high misconduct (two or more complaints) to those with low- and no-misconduct (zero or one complaint).

In Panel B, we compare officers with high misconduct (three or more complaints) to the rest. Our findings remain unchanged when using these alternative specifications.

4 Results

4.1 Officers with more misconduct are more likely to make an arrest than those with no misconduct

To examine whether officers with more misconduct are more likely to make an arrest relative to officers with no misconduct, we estimate this regression model:

$$y_{itr} = \beta_0 + \beta_1 Low_{i,r-1} + \beta_2 Medium_{i,r-1} + \beta_3 High_{i,r-1} + X'_{itr}\delta + \varepsilon_{itr} \quad (1)$$

where y_{itr} is the probability that officer i made an arrest on day t in year r. We also estimate the probabilities of making an arrest for an index crime and for a non-index crime. We classify the total number of complaints officer i had the previous year, in year $r-1$, into four buckets: No complaints ("No misconduct"), 1 complaint ("Low misconduct"), 2 complaints ("Medium misconduct"), and 3 or more complaints ("High misconduct"). The reference group is 0 complaints. The term X_{itr} contains controls for officer tenure (included as a quadratic) and a fixed effect for shift. Standard errors are clustered by shift.

Figure 2a depicts the coefficient estimates for β_1, β_2, and β_3. An officer with no misconduct has an 4.6% chance of making at least one arrest on a shift. Those with one complaint, or a low level of misconduct, are 0.64 percentage points more likely to make an arrest and this is statistically significant at the 1% level. Those with two complaints, or a medium level of misconduct, are 1.8 percentage points more likely to make an arrest ($p < 0.01$). Those with three or more complaints, or a high level of misconduct, are 3.7 percentage points more likely to make an arrest ($p < 0.01$). This is an effect size of about 80% of the reference group mean.

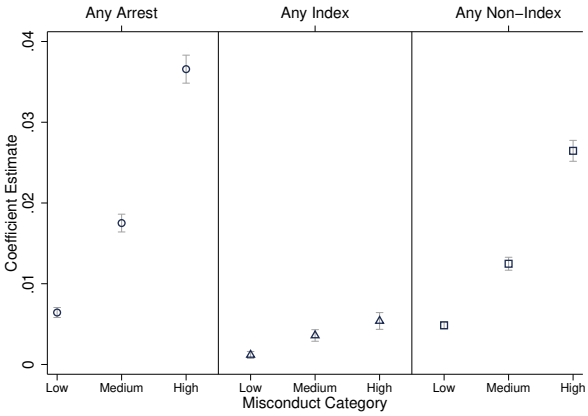
We also conduct t-tests to see if these estimates are statistically significantly different from each other, and they are (see Appendix Table A5). This indicates that officers with high misconduct are more likely to make arrests than those with low- or medium-misconduct.

We also look at the likelihood of making an arrest for an index crime, which are more serious

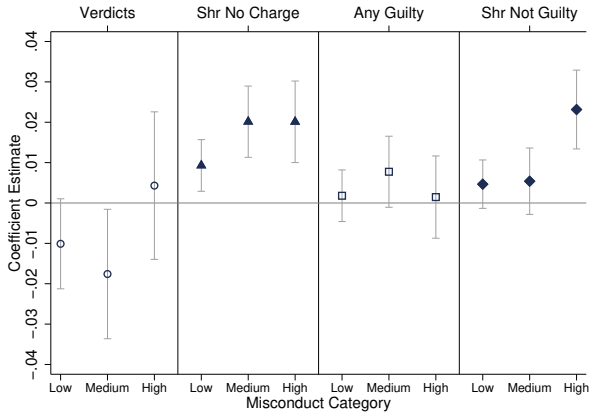
crimes. We see similar results here though on a smaller scale. We also report estimates for the probability of making an arrest for a non-index crime, which are less serious crimes. Again, we see similar results. Further, the magnitudes suggest that the overall arrest results are being driven by arrests for non-index crimes.

Figure 2: Impact of Officer Misconduct on Conviction-less Arrests

(a) Arrest Outcomes



(b) Court Outcomes



Notes: This figure depicts coefficient estimates for β_1, β_2, and β_3 from equation (1) for various outcomes. Figure (a) reports arrest outcomes: probability of an arrest, probability of an arrest for an index-crime, and probability of an arrest for a non-index crime. Figure (b) reports court outcomes: number of verdicts, share of arrests with no charge, probability of an arrest having a guilty outcome, share of an arrest's verdicts with a not guilty outcome. All estimates control for officer tenure and shift. Figure (b) also includes the number of arrests and the number of verdicts as controls. Wings are 95% confidence intervals with clustered standard errors.

Appendix Table A5 reports these estimates in table format. Appendix Table A6 reports estimates for the number of arrests rather than the probability. The results are the same; officers with more misconduct make more arrests than those with no misconduct.

4.2 Officers with more misconduct make more conviction-less arrests than those with no misconduct

To examine whether officers with more misconduct make more conviction-less arrests, we estimate equation (1) on the sample of officers who made at least one arrest and look at four court outcomes: the number of verdicts an arrest has (used as a proxy for charges); the share of arrests that have no charges (since officers may make more than one arrest on a shift); the probability of an arrest having a guilty outcome; and the share of an arrest's verdicts that are not guilty. Because an

arrest may have multiple charges, we estimate the share of verdicts that are not guilty rather than the likelihood of any "not guilty" outcome in order to be more conservative. The regression model is the same as (1), except the term X also includes the total number of arrests officer i made on day t in year r. The regression model for the outcomes "any guilty" and "share not guilty" also include the total number of verdicts from officer i's arrests made on day t in year r.

Figure 2b displays the estimates. Appendix Table A5 reports these estimates in table format. First, we look at the share of arrests that had no charge. Relative to the arrests made by officers with no misconduct, a larger share of arrests made by officers with any misconduct are more likely to result in no charges. The effect sizes are about 4.9% for low-misconduct officers and 10.5% for medium-misconduct and high-misconduct officers (all $p < 0.01$). The estimate for high-misconduct officers is statistically significantly different from the estimate for low-misconduct officers at the 5% level, but the estimates for medium- and high-misconduct officers are not (see Appendix Table A5).

If we examine the probability of any of these charges having a "Guilty" court outcome, we that it largely does not differ across misconduct level. Medium-misconduct officers are 0.77 percentage points more likely to have a guilty outcome, relative to no-misconduct officers, but this estimate is weakly significant at 10%. However, if we look at the share of verdicts that were found "Not Guilty" in court, then we do see differences by misconduct level. Specifically, the share of not guilty outcomes on arrests made by officers with high misconduct is 2.3 percentage points ($p < 0.01$) higher than the share of not guilty outcomes on arrests made by officers with no complaints last year. This is an effect size of 14%. We probe these results to see which arrest-types are driving them and find that the increase in not guilty outcomes is driven by arrests for non-index crimes, which is also what officers with greater misconduct are more likely to make (see Appendix Table A7).

As a robustness check, we examine whether our results hold when we redefine misconduct as complaints about more serious violations of appropriate officer conduct (Appendix Table A8). We classify complaints into three categories: Serious, Failure to Provide Service (FPS), and Personnel Conduct.

"Serious" complaints include allegations of use of force violations, arrest/lockup violations, and search violations. These complaints generally involve potential misconduct by officers during their interactions with civilians. "FPS" complaints, on the other hand, allege a neglect of duty or inade-

quate service/failure to provide service. These complaints are typically filed by individuals who are seeking police assistance or reporting themselves as victims of a crime.

The "Personnel Conduct" category encompasses a broader range of complaints related to the officer's general professional conduct and behavior, such as being absent without permission, being intoxicated, extortion, or the possession/sale of drugs. These complaints signal concerning behavior by the officer rather than during specific incidents involving civilians.[11]

We see similar results when we classify misconduct using the Serious and Personnel Conduct complaint categories. This is not surprising as both signal concerning behavior by the officers, either in their interactions with civilians (Serious) or in their overall professional conduct (Personnel Conduct). In contrast, the estimates using FPS complaints are largely not statistically significant.[12] This difference likely stems from the different nature of FPS complaints, which are typically filed by complainants who are potentially victims of a crime or who seek police help, rather than by complainants who may viewed as criminal by the officers.

Last, we expand the analysis to include arrests made by primary and secondary arresting officers in Appendix Table A9. The results remain the same. In summary, the results across all of these analyses suggest that officers with more misconduct, as measured by complaints, are making a higher number of conviction-less arrests.

5 Additional Analyses

5.1 Officer tenure

In this section, we examine whether the results are driven by officer tenure. This is a plausible hypothesis because more junior officers (those with 5-12 years of experience) are overrepresented in terms of misconduct level (see Figure 1). We control for this relationship in our main regression (equation (1)) by including a quadratic term for officer tenure, but we examine this relationship further in this section. Specifically, we estimate equation (1) separately for each tenure category: < 8 years (25th percentile or below); $[8, 13)$ years (25th-50th percentile); $[13, 18)$ years (50th-75th percentile); and ≥ 18 years (75th percentile or above).

[11] See Appendix Table A.5 in Ba (2018) for a complete list of allegation categories.

[12] Interestingly, the magnitude of the coefficient estimates for the share of no charges and share not guilty are similar for high-misconduct officers identified using FPS complaints compared to those identified using Serious complaints.

The results are in Appendix Table A10. Although relatively more junior officers with high misconduct are more likely to make arrests and non-index arrests, the evidence is more mixed on conviction-less arrests. For example, mid-career officers—those with 13-18 years of experience—with high misconduct have 5 percentage points ($p < 0.1$) more "Not Guilty" outcomes compared to mid-career officers with no misconduct, but the estimate for more senior officers (18 or more years of experience) is not statistically significant although the magnitude is a similar size. Similarly, the most junior officers with high misconduct have 1.7 percentage points ($p < 0.1$) more "Not Guilty" outcomes compared to their peers with no misconduct, but the estimate for officers with 8-13 years of experience is not statistically significant.

5.2 Bad apples

In this section, we examine whether these results are driven by "bad apples"–individual officers whose misconduct may disproportionately account for the patterns we observe. Recent empirical work has documented substantial officer-level heterogeneity in misconduct and its impacts on civilian outcomes (Goncalves and Mello, 2021; Rozema and Schanzenbach, 2019; Chalfin and Kaplan, 2021). For example, Chalfin and Kaplan (2021) find that the top 2% of officers (ranked by total complaints) account for 26% of use of force complaints.

To examine this, we first identify officers who are in the 95th percentile of the rate of complaints over our entire sample. The rate of complaints is defined as the total number of complaints divided by the total number of shifts. Because some officers join the CPD in the middle of our sample period, we make this adjustment to make the number of complaints comparable across officers. We then create quintiles of the total number of shifts and classify officers in the 95th percentile in each quintile. These officers make up about 9% of all officers. About 13% of all officers are ever in the high-misconduct category, and 58% of the officers who are ever in the high-misconduct category are not "bad apples".

The regression model is a modified version of equation 1:

$$
\begin{aligned}
y_{itr} = \beta_0 &+ \beta_1 Low_{i,r-1} + \beta_2 Medium_{i,r-1} + \beta_3 High_{i,r-1} + \beta_4 95th_i \\
&+ \beta_5 \left(Low_{i,r-1} \times 95th\right) + \beta_6 \left(Medium_{i,r-1} \times 95th\right) \\
&+ \beta_7 \left(High_{i,r-1} \times 95th\right) + X'\delta + \varepsilon_{itr}
\end{aligned} \tag{2}
$$

Appendix Figure A3 and Appendix Table A11 report results. We find the same patterns here that we see for the main results on arrest outcomes. High-misconduct officers, regardless of "bad apple" status, are more likely to make an arrest overall and more likely to make an arrest for non-index crimes compared to no-misconduct officers (both $p < 0.01$). The results on court outcomes, however, are mixed. High-misconduct officers who are also bad apples have a higher share of verdicts with not guilty outcomes compared to no-misconduct officers ($p < 0.05$), whereas high-misconduct officers who are not bad apples have a higher share of arrests with no charges compared to no-misconduct officers ($p < 0.05$).

5.3 Partners

In this section, we examine the extent to which high-misconduct officers may influence their partners. The research on peer effects in education documents how individual behavior can be shaped through social learning and behavioral changes (Sacerdote, 2001; Carrell et al., 2018; Lavy et al., 2012; Burke and Sass, 2013). Similarly, high-misconduct officers may influence their partners' behavior through direct interaction and observation, or they may strategically ask their partners to serve as the primary arresting officer on arrest forms to deflect potential scrutiny.

To examine this, we exploit changes in partner assignment due to the CPD rotational duty calendar. We focus on officers who have only one partner in their assigned beat, which make up the majority of shifts[13], and compare an officer's policing behavior when they are assigned to a high-misconduct partner vs. a no-misconduct partner, conditional on working on the same geographic

[13] Specifically, 69% of shifts have 2 officers, 24% have 1 officer, and 4.6% have 3 officers.

beat, watch, day of the week, and month-year. The regression model is as follows:

$$\begin{aligned} y_{itr} = \beta_0 &+ \beta_1 Low_{i,r-1} + \beta_2 Medium_{i,r-1} + \beta_3 High_{i,r-1} + \beta_4 LowP_{it,r-1} \\ &+ \beta_5 MediumP_{it,r-1} + \beta_6 HighP_{it,r-1} + \beta_7 \left(Low_{i,r-1} \times LowP_{it,r-1}\right) \\ &+ \beta_8 \left(Medium_{i,r-1} \times MediumP_{it,r-1}\right) + \beta_9 \left(High_{i,r-1} \times HighP_{it,r-1}\right) \\ &+ X'\delta + \varepsilon_{itr} \end{aligned} \tag{3}$$

where y_{itr}, $Low_{i,r-1}$, $Medium_{i,r-1}$, and $High_{i,r-1}$ are defined as in equation (1). $LowP_{it,r-1}$ is a binary variable equal to 1 if officer i's partner on day t had 1 complaint last year (in year $r-1$) and equal to 0 otherwise. $MediumP_{it,r-1}$ is equal to 1 if officer i's partner on day t had 2 complaints last year. $HighP_{it,r-1}$ is equal to 1 if officer i's partner on day t had 3 or more complaints last year. The term X contains fixed effects for officer and shift. Standard errors are clustered at the officer level.

Appendix Table A12 reports results. In general, most of the coefficient estimates are small in magnitude and/or not statistically significant. Some estimates that are of note are: officers with a high-misconduct partner are 0.27 percentage points ($p < 0.05$) more likely to make an arrest for a non-index crime relative to when they are partnered with a no-misconduct partner (14% of the reference group mean). When a high-misconduct officer is paired with a high-misconduct partner, they are 0.782 percentage points ($p < 0.05$) more likely to make a non-index arrest (72% of the reference group mean). The estimates on "share not guilty" are not statistically significant but the magnitudes are non-trivial. For example, two high-misconduct officers partnered together have 1.6 percentage points (9.6% of the mean) more "not guilty" outcomes compared to a no-misconduct officer pair.

6 Which officers respond most to increased oversight?

In this section, we examine whether increased oversight can affect officer behavior. The policy change is the *Kalven v. City of Chicago* ruling on March 10, 2014 that made police-misconduct records public information under the Illinois Freedom of Information Act (FOIA). Prior to *Kalven v. City of Chicago*, officer misconduct records were not public information. On November 26, 2009,

journalist Jamie Kalven submitted FOIA requests to the Chicago Police Department (CPD) asking for records on officer complaints.[14] On December 8, 2009, CPD denied the requests stating that they were exempt from FOIA disclosure. On December 22, 2009, Kalven filed a lawsuit challenging the protective order.

On March 10, 2014, the Illinois Appellate Court ruled that the police misconduct records were not exempt from FOIA disclosure. However, even after this decision, the Chicago police union (Fraternal Order of Police "FOP") fought to keep these records from being released. In 2014, prompted by FOIA requests from the Chicago Sun-Times and the Chicago Tribune, the city of Chicago attempted to release all officer complaint information going back to 1967.[15] This release was held up by a temporary injunction brought by the Fraternal Order of Police.[16] The injunction allowed only the four most recent years of information to be released, due to a provision in FOP's collective bargaining agreement with the city of Chicago requiring the destruction of records of alleged police misconduct once the record is more than four years old.

FOP then sued CPD for failing to destroy records of police misconduct older than five years.[17] The appellate court vacated the injunction as being against public policy (meaning CPD was able to release misconduct records more than four years old) and also found that destroying these records violated an explicit, well-defined, and dominant public policy. The Illinois Supreme Court agreed.

We theorize that this ruling changed the visibility of officer-specific information on the number and nature of complaints filed against the officer. It was also expected to change policing practices, as evidenced by the Chicago police union's response. The FOP president published a letter on the front page of the union's August 2014 newsletter about the court ruling and how it would affect officers:

> An additional concern of ours is the impact this type of disclosure might have on the reputation of the individual Police Officers and quite frankly, the Chicago Police Department as well. A Police Officer might suffer undue scrutiny from family members, friends, neighbors and even co-workers, when this information is released.

We conduct an event study analysis to examine how an officer's policing behavior changed after

[14] https://casetext.com/case/kalven-v-city-of-chi/
[15] https://caselaw.findlaw.com/court/il-supreme-court/1913070.html
[16] https://www.abajournal.com/news/article/largest_public_database_of_chicago_police_misconduct_debuts
[17] https://casetext.com/case/fraternal-order-of-police-v-the-city-of-chicago

these two events. We restrict the analysis sample to be between September 2013 (6 months before the March 2014 court ruling) and January 2015 (5 months after the August 2014 newsletter). This is to follow the strategy in Rivera and Ba (2023), which also examined this court ruling and its impact on crime rates.

As before, we classify officers into four misconduct levels: no misconduct (0 complaints last year), low misconduct (1 complaint last year), medium misconduct (2 complaints last year), and high misconduct (3 or more complaints last year). Because the event study spans from 2013 to 2015, we use the number of complaints from 2013 to classify officer misconduct levels. This approach ensures that each officer's misconduct level remains constant throughout the entire analysis period.

We estimate this regression model separately for each misconduct group g:

$$r_{it}^{g} = \beta_0^{g} + \beta_1^{g} Event1_t + \beta_2^{g} Event2_t + a_i^{g} + \varepsilon_{it}^{g} \tag{4}$$

where r_{it} is the outcome examined in Section 4 but adjusted for seasonal trends: the probability of making an arrest; the probability of making an index-crime arrest; the probability of making a non-index crime arrest; the number of verdicts; the share of arrests with no charges; the probability of having any guilty verdict; and the share of verdicts that are not guilty. Because there is no comparison group in an event study, we want to make sure that we are not capturing and conflating any seasonal trends in our estimates of the policy changes.[18] $Event1_t$ is an indicator variable equal to 1 if the day t is in April 2014 or later and 0 otherwise. $Event2_t$ is an indicator variable equal to 1 if the day t is in September 2014 or later and 0 otherwise. We exclude the months of the two events (March 2014 and August 2014) from our analysis. The reference period is September 2013 to February 2014, the six-month period before the *Kalven* ruling. The term a_i^g denotes officer fixed effects. Standard errors are clustered by officer.

Table 5 reports results.[19] Across all officers, the number of complaints fell. For most officers, the magnitude of the drop after the court ruling (event 1) is not statistically significantly different from the drop after the FOP newsletter (event 2). However, for high-misconduct officers, complaints

[18] Specifically, we regress the outcome variable on a set of month indicator variables only for the days outside our analysis period (the analysis period is September 1, 2013 to January 31, 2015): $y_{it} = \beta_0 + \delta_m + \varepsilon_{it}$, where δ_m are a set of month indicator variables. The residuals from this auxiliary regression (r_{it}) are used as the dependent variable in equation (4).

[19] For a more detailed analysis of month-to-month changes, please see Appendix Figure A4.

fell by an additional 0.241 percentage points (67%) after the FOP newsletter. These two drops for high-misconduct officers are statistically significantly different from each other ($p < 0.01$).

Regarding the other outcomes, there is a distinction between the court ruling and the union president's letter in the FOP newsletter. Immediately after the court ruling, officers reduce conviction-less arrests by changing their arrest behavior compared to the arrests these same officers made before the court ruling. No-misconduct officers are less likely to make arrests overall, and specifically less likely to make non-index arrests, after the court ruling (both $p < 0.01$). These arrests by no-misconduct officers are also less likely to have no charges filed or have a lower share of "not guilty" verdicts ($p < 0.1$ and $p < 0.01$, respectively) compared to the arrests made by these same officers before the court ruling. Officers with misconduct do not change their likelihood of making an arrest after the court ruling, but the share of arrests with no charges is negative, the probability of any "guilty" outcome is positive, and the share of an arrest's verdicts that are "not guilty" goes down. Most of these estimates are significant at the 10% level.

These results suggest that no-misconduct officers initially reduce conviction-less arrests by reducing the number of arrests. Officers with misconduct appear to be changing their arrest behavior by making fewer conviction-less arrests. The effect among misconduct officers is stronger for low- and medium-misconduct officers and weaker for high-misconduct officers, though the magnitudes of the coefficients are large for high-misconduct officers as well.

After the FOP newsletter, however, we see heterogeneous effects by misconduct level. Arrest likelihoods for no-misconduct officers do not change compared to before the court ruling, but the share of no charges and of "not guilty" verdicts still fall ($p < 0.05$ and $p < 0.01$, respectively). Additionally, the number of verdicts per arrest increases ($p < 0.01$). Low-misconduct officers look more similar to no-misconduct officers than the others, as their probability of arrest decreases ($p < 0.01$), the number of verdicts increases ($p < 0.05$), and the share of no charges and share of not guilty are both negative though not statistically significant. The signs and magnitudes of medium-misconduct officers are similar to low-misconduct officers but none of the court outcomes are statistically significant. These results suggest that these officers are changing the types of arrests they are making after the union president's letter, potentially focusing on more serious incidents that result in more charges.

In contrast, there is suggestive but weak evidence that high-misconduct officers are withdraw-

ing from policing after the union president's letter. For example, the probability of an arrest and specifically for non-index arrests go down ($p < 0.1$ and $p < 0.05$, respectively), while the probability of any "guilty" verdict is negative and the share of no charges and of "not guilty" verdicts are unchanged.

Table 5: Impact of Increased Oversight on Conviction-less Arrests

Outcome Variable:	Complaints (1)	Any Arrest (2)	Any Index (3)	Any Non-Index (4)	Num Verdicts (5)	Shr No Charge (6)	Any Guilty (7)	Shr Not-Guilty (8)
Panel A: No-Misconduct Officers								
Event 1 (Court Ruling)	0.000860***	-0.00311***	-0.000283	-0.00156***	0.00973	-0.0105*	0.00649	-0.0233***
	(8.48e-05)	(0.000600)	(0.000326)	(0.000502)	(0.0101)	(0.00561)	(0.00565)	(0.00562)
Event 2 (FOP Newsletter)	0.000978***	-0.000711	-0.000300	-0.000564	0.0337***	-0.0145**	-0.0119**	-0.0174***
	(7.61e-05)	(0.000680)	(0.000323)	(0.000574)	(0.0101)	(0.00589)	(0.00566)	(0.00567)
Observations	1,030,952	1,030,952	1,030,952	1,030,952	38,347	38,347	38,347	38,347
Reference Group Mean	0.000	0.035	0.013	0.023	1.068	0.200	0.274	0.200
p-value for t-test:								
Event 1 (Court Ruling) = Event 2	0.254	0.000	0.961	0.057	0.022	0.473	0.001	0.242
Panel B: Low-Misconduct Officers								
Event 1 (Court Ruling)	-0.00112***	-0.00166	0.000822	-0.00115	0.0102	-0.0103	0.0259***	-0.0140*
	(0.000211)	(0.00129)	(0.000617)	(0.00110)	(0.0138)	(0.00738)	(0.00719)	(0.00733)
Event 2 (FOP Newsletter)	-0.00123***	-0.00680***	-0.00134**	-0.00566***	0.0304**	-0.0132	-0.00473	-0.0116
	(0.000194)	(0.00130)	(0.000596)	(0.00111)	(0.0150)	(0.00811)	(0.00807)	(0.00781)
Observations	353,581	353,581	353,581	353,581	18,457	18,457	18,457	18,457
Reference Group Mean	0.003	0.053	0.016	0.037	1.098	0.203	0.271	0.204
p-value for t-test:								
Event 1 = Event 2	0.590	0.000	0.001	0.000	0.199	0.725	0.000	0.760
Panel C: Medium-Misconduct Officers								
Event 1 (Court Ruling)	-0.00329***	-0.00173	-0.000281	7.03e-05	-0.00835	-0.0144	0.0178*	-0.0194*
	(0.000444)	(0.00293)	(0.00135)	(0.00262)	(0.0215)	(0.0112)	(0.0105)	(0.0116)
Event 2 (FOP Newsletter)	-0.00269***	-0.00977***	-0.00152	-0.00808***	0.0294	-0.0161	-0.00382	-0.00765
	(0.000491)	(0.00304)	(0.00130)	(0.00286)	(0.0206)	(0.0116)	(0.0111)	(0.0120)
Observations	104,534	104,534	104,534	104,534	8,374	8,374	8,374	8,374
Reference Group Mean	0.006	0.081	0.019	0.063	1.143	0.208	0.280	0.221
p-value for t-test:								
Event 1 = Event 2	0.207	0.006	0.331	0.002	0.049	0.873	0.054	0.337
Panel D: High-Misconduct Officers								
Event 1 (Court Ruling)	-0.00362***	-0.000719	0.000657	0.000396	-0.00249	0.00870	0.0131	-0.0247*
	(0.000940)	(0.00444)	(0.00192)	(0.00405)	(0.0262)	(0.0115)	(0.0142)	(0.0145)
Event 2 (FOP Newsletter)	-0.00603***	-0.00882*	0.00138	-0.0105**	-0.0148	-0.00857	-0.0229	-0.00483
	(0.000905)	(0.00474)	(0.00180)	(0.00462)	(0.0267)	(0.0129)	(0.0151)	(0.0162)
Observations	53,772	53,772	53,772	53,772	5,975	5,975	5,975	5,975
Reference Group Mean	0.010	0.113	0.021	0.093	1.308	0.178	0.314	0.225
p-value for t-test:								
Event 1 = Event 2	0.008	0.062	0.663	0.008	0.650	0.231	0.020	0.195

Notes: This table reports coefficient estimates for β_1^g and β_2^g from equation (4) for various outcomes, separately for each misconduct group g. All outcomes have been adjusted for seasonal trends, and all estimates also include officer fixed effects. Standard errors clustered by officer are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

7 Conclusion

Given the use of an individual's arrest history for many economic and social outcomes, reducing conviction-less arrests is an important policy goal. This paper examines whether police departments and policymakers can reduce conviction-less arrests by targeting officers with greater misconduct, as measured by complaints. Our focus on complaints stems from prior work that finds that civilian

complaints are meaningful indicators of police misconduct (Stoddard et al., 2024). Although past papers have examined force use by the worst offenders, they do not examine arrest behavior or the court outcomes of these arrests (e.g., Rozema and Schanzenbach, 2019; Chalfin and Kaplan, 2021).

Using the Chicago Police Department's rotational duty calendar to obtain plausibly exogenous variation in the set of officers assigned to work on a particular day, we establish an arguably causal relationship between officer misconduct and conviction-less arrests. Officers with high misconduct are 3.7 percentage points (80% of the mean) more likely than officers with no misconduct to make an arrest and 2.7 percentage points (133% of the mean) more likely to make an arrest for less serious crimes. More importantly, the arrests made by officers with high misconduct are 2 percentage points (11% of the mean) less likely to result in charges and are 2.3 percentage points (14% of the mean) more likely to have a "Not Guilty" outcome compared to the arrests made by officers with no misconduct.

Our event study analysis finds that increased transparency through public disclosure of complaint records can influence officer behavior—though with varying effectiveness across misconduct levels. While no- and low-misconduct officers appear more responsive to oversight mechanisms, high-misconduct officers may require more targeted interventions. This result has important policy implications as it demonstrates how transparency requirements may improve policing outcomes at minimal cost. Further, the potential for a nationwide impact is substantial, given that police misconduct records remain confidential in 23 states, have limited availability in 15 states, and are fully public in only 12 states—with many of these 12 still restricting access to unsubstantiated complaints or ongoing investigations.

Our findings also have policy implications for police departments. Despite the predictive power of complaints against officers, many departments do not effectively use them as monitoring tools. One reason may be departmental resource constraints that prevent comprehensive officer monitoring (Black et al., 2017, p. 9). Another issue is that being active on the job and complaints are positively correlated. Therefore, an effective policy is not one that simply seeks to reduce all complaints but rather only those related to improper police behavior. Our findings reveal that increased transparency of officer misconduct has the potential to reduce conviction-less arrests, but mainly among officers with no or low misconduct. A more targeted approach may be necessary to effectively influence high-misconduct officers.

References

Agan, A., J. L. Doleac, and A. Harvey (2023). Misdemeanor prosecution. The Quarterly Journal of Economics 138(3), 1453–1505.

Agan, A., A. Garin, D. Koustas, A. Mas, and C. S. Yang (2024a, September). The labor market impacts of reducing felony convictions. American Economic Review: Insights 6(3), 341–58.

Agan, A. and S. Starr (2018). Ban the box, criminal records, and racial discrimination: A field experiment. The Quarterly Journal of Economics 133(1), 191–235.

Agan, A. Y., A. Garin, D. K. Koustas, A. Mas, and C. Yang (2024b). Can you erase the mark of a criminal record? labor market impacts of criminal record remediation. Technical report, National Bureau of Economic Research.

Ang, D., P. Bencsik, J. M. Bruhn, and E. Derenoncourt (2024). Community engagement with law enforcement after high-profile acts of police violence. Technical report, National Bureau of Economic Research.

Ba, B. (2018). Going the extra mile: The cost of complaint filing, accountability, and law enforcement outcomes in chicago. Working paper.

Ba, B. A., M. Baskar, and R. Mariman (2024). Understanding demand for police alternatives. Technical report, National Bureau of Economic Research.

Bell, M. C. (2021). Next-generation policing research: Three propositions. Journal of Economic Perspectives 35(4), 29–48.

Black, C. et al. (2017). The DOJ investigation of the Chicago Police Department: The complete report by the United States Department of Justice. Simon and Schuster.

Burke, M. A. and T. R. Sass (2013). Classroom peer effects and student achievement. Journal of Labor Economics 31(1), 51–82.

Campbell, R. (2024). What does federal oversight do to policing and public safety? evidence from seattle. Working paper.

Carey, C. A. (2004). No second chance: People with criminal records denied access to public housing. U. Tol. L. Rev. 36, 545.

Carrell, S. E., M. Hoekstra, and E. Kuka (2018). The long-run effects of disruptive peers. American Economic Review 108(11), 3377–3415.

Chalfin, A. and J. Kaplan (2021). How many complaints against police officers can be abated by incapacitating a few "bad apples?". Criminology & Public Policy 20(2), 351–370.

Cheng, C. and W. Long (2018). Improving police services: Evidence from the french quarter task force. Journal of Public Economics 164, 1–18.

Craigie, T.-A. (2020). Ban the box, convictions, and public employment. Economic Inquiry 58(1), 425–445.

Cullen, Z., W. Dobbie, and M. Hoffman (2023). Increasing the demand for workers with a criminal record. The Quarterly Journal of Economics 138(1), 103–150.

Cunningham, J., D. Feir, and R. Gillezeau (2024). Collective bargaining rights, policing, and civilian deaths. Working paper.

Dee, T. S. and J. Pyne (2022). A community response approach to mental health and substance abuse crises reduced crime. Science Advances 8(23), eabm2106.

Doleac, J. L. and B. Hansen (2020). The unintended consequences of "ban the box": Statistical discrimination and employment outcomes when criminal histories are hidden. Journal of Labor Economics 38(2), 321–374.

Goncalves, F. and S. Mello (2021). A few bad apples? racial bias in policing. American Economic Review 111(5), 1406–1441.

Heaton, P., S. Mayson, and M. Stevenson (2017). The downstream consequences of misdemeanor pretrial detention. Stan. L. Rev. 69, 711.

Hirschfield, P. (2009). Another way out: The impact of juvenile arrests on high school dropout. Sociology of Education 82(4), 368–393.

Jackson, O., R. Sullivan, and B. Zhao (2017). Reintegrating the ex-offender population in the us labor market: lessons from the cori reform in massachusetts. Technical report, FRB Boston Research Data Report.

Jain, E. (2015). Arrests as regulation. Stan. L. Rev. 67, 809.

Knox, D., W. Lowe, and J. Mummolo (2020). Administrative Records Mask Racially Biased Policing. American Political Science Review 114, 619–637.

Krishna, T. (2024). 'diversion, a deterrence to delinquency?'the impact of florida's juvenile civil citation program on recidivism. In 2024 APPAM Fall Research Conference. APPAM.

Lavy, V., M. D. Paserman, and A. Schlosser (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. The Economic Journal 122(559), 208–237.

McWilliams, E. R. and B. A. Hunter (2021). The impact of criminal record stigma on quality of life: A test of theoretical pathways. American Journal of Community Psychology 67(1-2), 89–102.

Mueller-Smith, M. and K. T. Schnepel (2021). Diversion in the criminal justice system. The Review of Economic Studies 88(2), 883–936.

Mummolo, J. (2018). Modern police tactics, police-citizen interactions, and the prospects for reform. The Journal of Politics 80(1), 1–15.

Nadel, M. R., G. Pesta, T. Blomberg, W. D. Bales, and M. Greenwald (2018). Civil citation: Diversion or net widening? Journal of Research in Crime and Delinquency 55(2), 278–315.

Ouss, A. and J. Rappaport (2020). Is police behavior getting worse? data selection and the measurement of policing harms. The Journal of Legal Studies 49(1), 153–198.

Prescott, J. and S. B. Starr (2019). Expungement of criminal convictions: An empirical study. Harv. L. Rev. 133, 2460.

Rivera, R. G. and B. A. Ba (2023). The effect of police oversight on crime and misconduct allegations: Evidence from chicago. Review of Economics and Statistics, 1–45.

Rose, E. K. (2021). Does banning the box help ex-offenders get jobs? evaluating the effects of a prominent example. Journal of Labor Economics 39(1), 79–113.

Rozema, K. and M. Schanzenbach (2019). Good cop, bad cop: Using civilian allegations to predict police misconduct. American Economic Journal: Economic Policy 11(2), 225–268.

Rozema, K. and M. Schanzenbach (2023). Does discipline decrease police misconduct? evidence from chicago civilian allegations. American Economic Journal: Applied Economics 15(3), 80–116.

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

Sánchez De La Sierra, R., K. Titeca, H. Xie, A. A. Lameke, and A. J. Malukisa (2024). The real state: Inside the congo's traffic police agency. American Economic Review 114(12), 3976–4014.

Selbin, J., J. McCrary, and J. Epstein (2018). Unmarked? criminal record clearing and employment outcomes. The Journal of Criminal Law and Criminology (1973-) 108(1), 1–72.

Stewart, R. and C. Uggen (2020). Criminal records and college admissions: A modified experimental audit. Criminology 58(1), 156–188.

Stoddard, G., D. J. Fitzpatrick, and J. Ludwig (2024). Predicting police misconduct. Technical report, National Bureau of Economic Research.

Online Appendix for

“The Relationship between Officer Misconduct and Conviction-less Arrests”

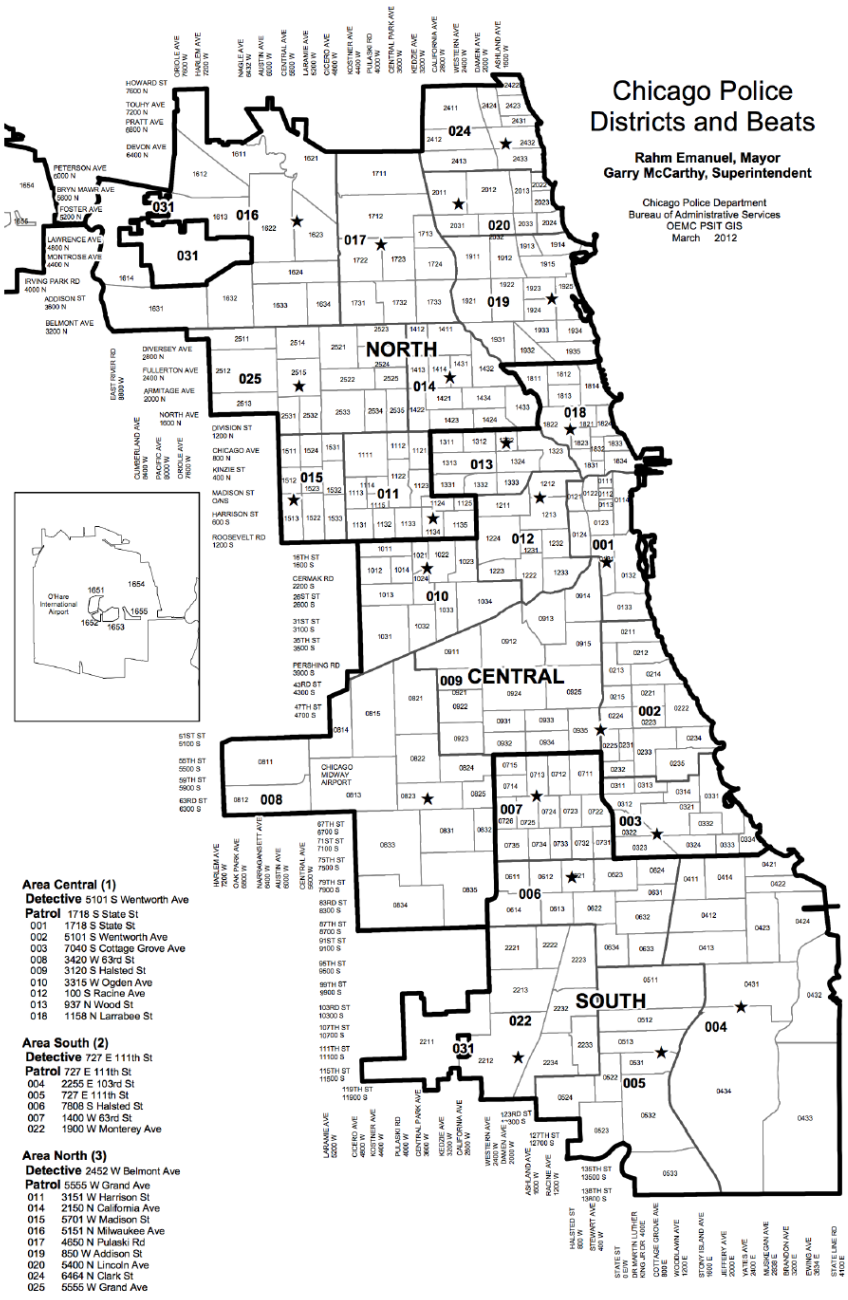
A Additional Figures and Tables

Table A1: Descriptive Statistics: All Officers

		Number of complaints last year:			
	Overall	0	1	2	3+
Unique number of officers	11,606	11,060	5,913	2,690	1,565
Average share of officers	1.000	0.697	0.193	0.067	0.043
Share with any arrest	0.093	0.036	0.065	0.089	0.126
Mean number of arrests	1.219	1.139	1.166	1.201	1.265
	(0.604)	(0.484)	(0.534)	(0.586)	(0.712)
Mean number of verdicts	1.193	1.097	1.129	1.187	1.284
	(0.952)	(0.855)	(0.893)	(0.942)	(1.039)

Notes: This table depicts summary statistics on all Chicago Police Department officers between 2010 and 2016. "Unique number of officers" is the unique number of officers between 2010-2016 in the relevant misconduct category. "Average share of officers" is the average over years. "Share with any arrest" is the share of all officer-shifts between 2010-2016 with any primary arrests. The mean number of arrests and verdicts are conditional on making an arrest. Standard deviations are in parentheses.

Figure A1: Map of Chicago Police Department Districts and Beats



Notes: This map outlines the geographic boundaries of the CPD districts and their corresponding beats in 2012. There were 25 districts in 2012; three districts closed by 2013.

Figure A2: Distribution of Officer Tenure by Watch

Notes: This figure shows the distribution of officer tenure by watch assignments.

Table A2: Does misconduct predict which day of the week an officer works? Alternative Specifications

Outcome Variable:	Monday (1)	Tuesday (2)	Wednesday (3)	Thursday (4)	Friday (5)	Saturday (6)	Sunday (7)
Panel A: Patrol Officer Sample, Medium/High Misconduct vs Not							
One or more Complaints Last Week	-6.22e-05	0.000357	0.00128	0.000240	-0.00115*	-4.75e-05	-0.000625
	(0.000710)	(0.000497)	(0.000813)	(0.000521)	(0.000662)	(0.000855)	(0.000683)
Constant	0.139***	0.142***	0.144***	0.144***	0.145***	0.144***	0.141***
	(2.50e-05)	(1.72e-05)	(3.12e-05)	(1.65e-05)	(2.43e-05)	(3.37e-05)	(2.30e-05)
Observations	3,146,453	3,146,453	3,146,453	3,146,453	3,146,453	3,146,453	3,146,453
R-squared	0.004	0.002	0.002	0.002	0.002	0.002	0.003
Panel B: Patrol Officer Sample, High Misconduct vs Not							
Two or more Complaints Last Week	0.00147	-0.00190	0.00220	0.000686	-0.00402	0.00324	-0.00168
	(0.00331)	(0.00276)	(0.00381)	(0.00255)	(0.00305)	(0.00401)	(0.00321)
Constant	0.139***	0.142***	0.144***	0.144***	0.145***	0.144***	0.141***
	(4.10e-06)	(3.60e-06)	(5.33e-06)	(2.71e-06)	(3.95e-06)	(6.32e-06)	(3.89e-06)
Observations	3,146,453	3,146,453	3,146,453	3,146,453	3,146,453	3,146,453	3,146,453
R-squared	0.004	0.002	0.002	0.002	0.002	0.002	0.003
Panel C: All Officer Sample, Linear Measure of Complaints							
Total Complaints Last Week	0.000445	0.000458	0.000808**	0.000249	-0.000867***	-0.000456	-0.000638**
	(0.000326)	(0.000401)	(0.000346)	(0.000271)	(0.000270)	(0.000384)	(0.000278)
Constant	0.122***	0.144***	0.154***	0.154***	0.153***	0.147***	0.126***
	(7.05e-06)	(1.21e-05)	(1.16e-05)	(7.92e-06)	(8.23e-06)	(1.27e-05)	(5.05e-06)
Observations	8,582,265	8,582,265	8,582,265	8,582,265	8,582,265	8,582,265	8,582,265
R-squared	0.012	0.004	0.002	0.002	0.002	0.003	0.010

Notes: This table reports estimates for the number of complaints last week on the probability of working on a specific day of the week. In Panels A and B, medium-misconduct is defined as 1 complaint last week and high-misconduct is defined as 2 or more complaints last week. All estimates also include fixed effects for officer, district, and year. Standard errors clustered by officer and district are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A3: Are officers with misconduct able to choose which days to work based on expected crime? Alternative Specifications

Outcome Variable:	Memorial Day (1)	Fourth of July (2)	Labor Day (3)	St. Patty's Day (4)	New Year's (5)
Panel A: Patrol Officer Sample, Medium/High Misconduct vs Not					
Two or more Complaints Last Year	0.000274	-0.00491	-0.00523	0.00855	0.0131
	(0.0101)	(0.0110)	(0.00794)	(0.00847)	(0.00993)
Observations	26,207	26,207	26,207	26,207	26,207
Mean Outcome	0.248	0.237	0.250	0.244	0.411
Panel B: Patrol Officer Sample, High Misconduct vs Not					
Three or more Complaints Last Year	0.0108	-0.000354	-0.000588	-0.0111	0.0196
	(0.0200)	(0.0186)	(0.0161)	(0.0212)	(0.0174)
Observations	26,207	26,207	26,207	26,207	26,207
Mean Outcome	0.248	0.237	0.250	0.244	0.411
Panel C: All Officer Sample, Linear Measure of Complaints					
Total Complaints Last Year	-0.00569	-0.00674*	-0.00343	-0.00313	-0.00256
	(0.00386)	(0.00346)	(0.00394)	(0.00349)	(0.00302)
Observations	41,284	41,284	41,284	41,284	41,284
Mean Outcome	0.438	0.422	0.433	0.450	0.665

Notes: This table reports estimates for the number of complaints last year on the probability of working on a specific holiday the following year. New Year's includes working on New Year's Eve and New Year Day. In Panels A and B, medium-misconduct is defined as 2 complaints last year and high-misconduct is defined as 3 or more complaints last year. The sample is at the officer-year level. All estimates also control for officer tenure at the start of the year and include fixed effects for officer, modal district, and year. Standard errors clustered by officer and district are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A4: Does misconduct predict watch for officers? Alternative Specifications

Outcome Variable:	Watch 1 (1)	Watch 2 (2)	Watch 3 (3)
Panel A: Patrol Officer Sample, Medium/High Misconduct vs Not			
Two or more Complaints Last Year	0.00240	0.00613	-0.00237
	(0.00902)	(0.00843)	(0.00755)
Observations	18,725	18,725	18,725
Mean Outcome	0.347	0.334	0.442
Panel B: Patrol Officer Sample, High Misconduct vs Not			
Three or more Complaints Last Year	-0.00229	0.0153	-0.00428
	(0.0105)	(0.0144)	(0.0107)
Observations	18,725	18,725	18,725
Mean Outcome	0.347	0.334	0.442
Panel C: All Officer Sample, Linear Measure of Complaints			
Total Complaints Last Year	-0.00385*	-0.000390	-0.00102
	(0.00210)	(0.00235)	(0.00251)
Observations	30,926	30,926	30,926
Mean Outcome	0.289	0.354	0.393

Notes: This table reports estimates for the number of complaints last year on the probability of working a given watch the following year. Watch 1 is 12AM-8AM, Watch 2 is 8AM-4PM, and Watch 3 is 4PM-12AM. In Panels A and B, medium-misconduct is defined as 2 complaints last year and high-misconduct is defined as 3 or more complaints last year. The sample is at the officer-year level. All estimates also control for officer watch last year, officer tenure at the start of the year, officer total arrests two years ago, and include fixed effects for officer, modal district, and year. Standard errors clustered by officer and district are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A5: Impact of Officer Misconduct on Conviction-less Arrests

Outcome Variable:	Any Arrest (1)	Any Index (2)	Any Non-Index (3)	Num Verdicts (4)	Shr No Charge (5)	Any Guilty (6)	Shr Not-Guilty (7)
Low Misconduct	0.00644***	0.00118***	0.00485***	-0.0101*	0.00930***	0.00180	0.00466
	(0.000317)	(0.000218)	(0.000219)	(0.00569)	(0.00326)	(0.00326)	(0.00306)
Medium Misconduct	0.0175***	0.00359***	0.0125***	-0.0176**	0.0201***	0.00774*	0.00540
	(0.000562)	(0.000365)	(0.000410)	(0.00818)	(0.00450)	(0.00449)	(0.00419)
High Misconduct	0.0366***	0.00539***	0.0265***	0.00431	0.0201***	0.00147	0.0232***
	(0.000884)	(0.000529)	(0.000661)	(0.00932)	(0.00515)	(0.00520)	(0.00498)
Observations	3,209,376	3,209,376	3,209,376	133,447	133,447	133,447	133,447
Reference Group Mean	0.046	0.024	0.020	1.046	0.191	0.266	0.166
p-value of t-test:							
High = Low	0.000	0.000	0.000	0.137	0.043	0.951	0.000
High = Medium	0.000	0.003	0.000	0.047	0.996	0.306	0.002

Notes: This table reports coefficient estimates for β_1, β_2, and β_3 from equation (1) for various outcomes. All estimates also control for officer tenure and shift. Columns 4-7 also include the number of arrests, and columns 6-7 include the number of verdicts. Standard errors clustered by shift are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A6: Impact of Officer Misconduct on Number of Arrests

Outcome Variable:	Num Arrests (1)	Num Index (2)	Num Non-Index (3)
Low Misconduct	0.00734***	0.00120***	0.00544***
	(0.000371)	(0.000242)	(0.000256)
Medium Misconduct	0.0206***	0.00419***	0.0143***
	(0.000675)	(0.000415)	(0.000489)
High Misconduct	0.0442***	0.00629***	0.0310***
	(0.00110)	(0.000608)	(0.000814)
Observations	3,209,376	3,209,376	3,209,376
Reference Group Mean	0.051	0.025	0.022
p-value of t-test:			
High = Low	0.000	0.000	0.000
High = Medium	0.000	0.003	0.000

Notes: This table reports coefficient estimates for β_1, β_2, and β_3 from equation (1) for the number of arrests (column 1), the number of index-crime arrests (column 2), and the number of non-index crime arrests (column 3). All estimates also control for officer tenure and shift. Standard errors clustered by shift are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A7: Impact of Officer Misconduct on Court Outcomes by Arrest Type

Outcome Variable:	Num Verdicts (1)	Shr No Charge (2)	Any Guilty (3)	Shr Not-Guilty (4)
Panel A: Index Arrests				
Low Misconduct	0.00236	-0.000894	0.00281	-6.61e-05
	(0.00911)	(0.00359)	(0.00615)	(0.00558)
Medium Misconduct	-0.0302**	0.00918*	0.0209**	-0.00413
	(0.0138)	(0.00522)	(0.00907)	(0.00823)
High Misconduct	-0.00362	0.00262	-0.00337	0.0145
	(0.0153)	(0.00653)	(0.0111)	(0.0106)
Observations	44,987	44,987	44,987	44,987
Reference Group Mean	1.182	0.063	0.277	0.176
p-value of t-test:				
High = Low	0.709	0.605	0.591	0.181
High = Medium	0.150	0.392	0.064	0.124
Panel B: Non-Index Arrests				
Low Misconduct	0.00843	-0.0107	-0.000622	0.00939*
	(0.0108)	(0.00676)	(0.00543)	(0.00537)
Medium Misconduct	0.0103	-0.00751	-0.0124*	0.0208***
	(0.0151)	(0.00877)	(0.00717)	(0.00709)
High Misconduct	0.0500***	-0.0259***	-0.0121	0.0303***
	(0.0171)	(0.00962)	(0.00812)	(0.00813)
Observations	47,407	47,407	47,407	47,407
Reference Group Mean	0.904	0.339	0.241	0.158
p-value of t-test:				
High = Low	0.017	0.126	0.170	0.013
High = Medium	0.044	0.097	0.973	0.321

Notes: This table reports coefficient estimates for β_1, β_2, and β_3 from equation (1) for court outcomes, separately by arrest-type. The court outcomes are: the number of verdicts (column 1), the share of arrests with no charge (column 2), the probability of an arrest having any guilty outcome (column 3), and the share of an arrest's verdicts with a not guilty outcome (column 4). All estimates also control for officer tenure, shift, and the number of arrests. Columns 3-4 also include the number of verdicts. Standard errors clustered by shift are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A8: Impact of Officer Misconduct on Conviction-less Arrests by Alternative Misconduct Definitions

Outcome Variable:	Any Arrest (1)	Any Index (2)	Any Non-Index (3)	Num Verdicts (4)	Shr No Charge (5)	Any Guilty (6)	Shr Not-Guilty (7)
Panel A: Serious Complaints							
Low Misconduct	0.0145***	0.00276***	0.0106***	-0.0209***	0.0186***	0.00920***	0.00252
	(0.000406)	(0.000264)	(0.000294)	(0.00613)	(0.00348)	(0.00349)	(0.00325)
Medium Misconduct	0.0368***	0.00638***	0.0272***	-0.00691	0.0285***	0.00906*	0.0123**
	(0.000978)	(0.000586)	(0.000751)	(0.0104)	(0.00549)	(0.00550)	(0.00530)
High Misconduct	0.0628***	0.00937***	0.0445***	0.0303**	0.0134*	0.000984	0.0401***
	(0.00167)	(0.000936)	(0.00130)	(0.0130)	(0.00732)	(0.00732)	(0.00715)
Observations	3,209,376	3,209,376	3,209,376	133,447	133,447	133,447	133,447
Reference Group Mean	0.047	0.024	0.020	1.046	0.189	0.264	0.167
p-value of t-test:							
High = Low	0.000	0.000	0.000	0.000	0.495	0.281	0.000
High = Medium	0.000	0.006	0.000	0.018	0.080	0.352	0.001
Panel B: Failure to Provide Service Complaints							
Low Misconduct	0.000951**	0.000353	0.000491*	-0.0190***	0.00822**	-0.00712*	0.00188
	(0.000406)	(0.000276)	(0.000278)	(0.00688)	(0.00389)	(0.00392)	(0.00364)
Medium Misconduct	0.00363***	0.00150**	0.00282***	0.000745	0.00455	0.00660	-0.000911
	(0.00106)	(0.000732)	(0.000726)	(0.0172)	(0.00933)	(0.00957)	(0.00896)
High Misconduct	-0.00200	-0.00521***	0.00154	-0.0552	0.0316	-0.0194	0.0271
	(0.00271)	(0.00176)	(0.00188)	(0.0391)	(0.0249)	(0.0251)	(0.0230)
Observations	3,209,376	3,209,376	3,209,376	133,447	133,447	133,447	133,447
Reference Group Mean	0.052	0.025	0.024	1.052	0.197	0.267	0.170
p-value of t-test:							
High = Low	0.280	0.002	0.578	0.359	0.350	0.628	0.275
High = Medium	0.052	0.000	0.524	0.187	0.304	0.333	0.251
Panel C: Personnel Conduct Complaints							
Low Misconduct	0.00495***	0.000925***	0.00353***	0.00570	-0.000583	0.00312	-0.00172
	(0.000481)	(0.000324)	(0.000335)	(0.00783)	(0.00438)	(0.00443)	(0.00413)
Medium Misconduct	0.0161***	0.00130	0.0116***	-0.00178	0.0132	-0.00864	0.00780
	(0.00167)	(0.00104)	(0.00121)	(0.0206)	(0.0116)	(0.0119)	(0.0119)
High Misconduct	0.0264***	0.0106***	0.0178***	-0.243***	0.105***	-0.0123	0.0495
	(0.00515)	(0.00350)	(0.00393)	(0.0531)	(0.0322)	(0.0346)	(0.0353)
Observations	3,209,376	3,209,376	3,209,376	133,447	133,447	133,447	133,447
Reference Group Mean	0.052	0.025	0.023	1.048	0.197	0.265	0.170
p-value of t-test:							
High = Low	0.001	0.006	0.000	0.000	0.001	0.657	0.150
High = Medium	0.056	0.011	0.133	0.000	0.007	0.920	0.264

Notes: This table reports coefficient estimates for β_1, β_2, and β_3 from equation (1) for various outcomes, separately by alternative misconduct definitions. Panel A defines misconduct using "serious" complaints, which include allegations of use of force violations, arrest/lockup violations, and search violations. Panel B defines misconduct using complaints on failure to provide service. Panel C defines misconduct using complaints about personnel conduct, which relate to the officer's general professional conduct and behavior such as being absent without permission or the possession/sale of drugs. All estimates also control for officer tenure and shift. Columns 4-7 also include the number of arrests, and columns 6-7 include the number of verdicts. Standard errors clustered by shift are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A9: Impact of Officer Misconduct on Conviction-less Arrests by Primary and Secondary Arresting Officers

Outcome Variable:	Any Arrest (1)	Any Index (2)	Any Non-Index (3)	Num Verdicts (4)	Shr No Charge (5)	Any Guilty (6)	Shr Not-Guilty (7)
Low Misconduct	0.00788***	0.000716**	0.00672***	-0.00476	0.00476***	0.000140	0.00253
	(0.000423)	(0.000298)	(0.000297)	(0.00338)	(0.00184)	(0.00187)	(0.00174)
Medium Misconduct	0.0225***	0.00410***	0.0171***	-0.00245	0.0117***	0.00229	0.00405
	(0.000739)	(0.000492)	(0.000546)	(0.00503)	(0.00263)	(0.00269)	(0.00250)
High Misconduct	0.0441***	0.00487***	0.0344***	0.00462	0.0149***	0.00323	0.0145***
	(0.00113)	(0.000706)	(0.000854)	(0.00610)	(0.00321)	(0.00337)	(0.00313)
Observations	3,209,376	3,209,376	3,209,376	298,318	298,318	298,318	298,318
Reference Group Mean	0.090	0.046	0.039	1.078	0.191	0.272	0.167
p-value of t-test:							
High = Low	0.000	0.000	0.000	0.128	0.002	0.368	0.000
High = Medium	0.000	0.337	0.000	0.307	0.372	0.804	0.003

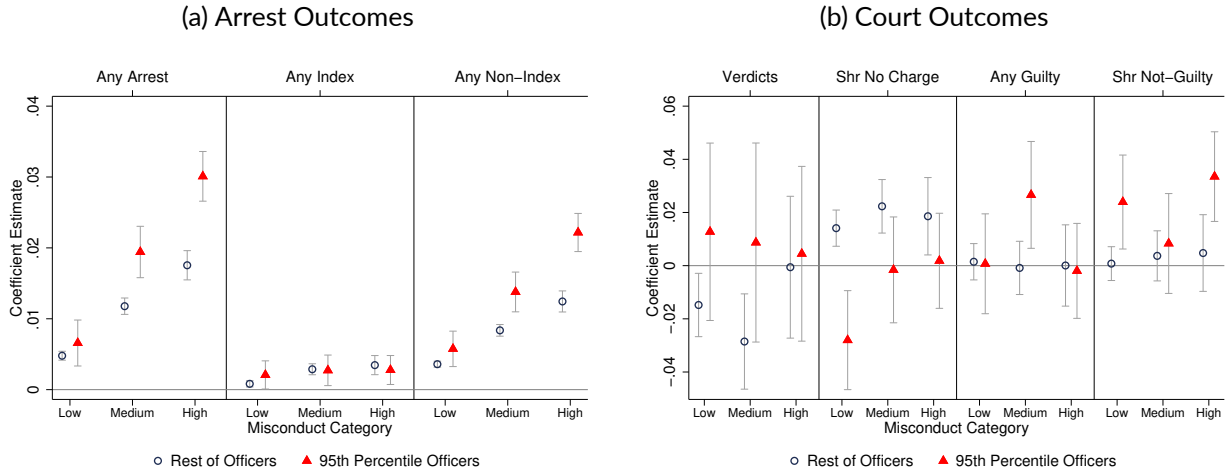
Notes: This table reports coefficient estimates for β_1, β_2, and β_3 from equation (1) for various outcomes on arrests made by primary and secondary arresting officers. All estimates also control for officer tenure and shift. Columns 4-7 also include the number of arrests, and columns 6-7 include the number of verdicts. Standard errors clustered by shift are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A10: Impact of Officer Misconduct on Conviction-less Arrests by Officer Tenure

Outcome Variable:	Any Arrest (1)	Any Index (2)	Any Non-Index (3)	Num Verdicts (4)	Shr No Charge (5)	Any Guilty (6)	Shr Not-Guilty (7)
Panel A: Tenure: < 8 years							
Low Misconduct	0.00911***	0.00175***	0.00636***	-0.00282	0.00490	-0.00128	0.00481
	(0.000667)	(0.000418)	(0.000497)	(0.00935)	(0.00548)	(0.00519)	(0.00502)
Medium Misconduct	0.0139***	0.00264***	0.0102***	0.00638	0.00231	0.00976	-0.00256
	(0.00102)	(0.000628)	(0.000778)	(0.0132)	(0.00740)	(0.00726)	(0.00685)
High Misconduct	0.0349***	0.00402***	0.0245***	0.0249*	-0.00786	0.0109	0.0168**
	(0.00148)	(0.000864)	(0.00113)	(0.0146)	(0.00808)	(0.00821)	(0.00771)
Observations	1,125,475	1,125,475	1,125,475	56,549	56,549	56,549	56,549
Mean Outcome	0.062	0.027	0.031	1.014	0.217	0.254	0.168
p-value of t-test:							
High = Low	0.000	0.011	0.000	0.059	0.119	0.141	0.130
High = Medium	0.000	0.166	0.000	0.268	0.273	0.905	0.033
Panel B: 8-13 years							
Low Misconduct	0.00414***	0.000920**	0.00351***	-0.0150	0.00803	-0.00440	-0.00144
	(0.000631)	(0.000438)	(0.000423)	(0.0160)	(0.00942)	(0.00944)	(0.00871)
Medium Misconduct	0.0225***	0.00480***	0.0154***	-0.0135	0.0170	0.0101	0.0166
	(0.00120)	(0.000774)	(0.000853)	(0.0219)	(0.0125)	(0.0121)	(0.0116)
High Misconduct	0.0399***	0.00580***	0.0315***	0.0187	0.0248*	0.0159	0.0129
	(0.00171)	(0.00104)	(0.00129)	(0.0233)	(0.0140)	(0.0138)	(0.0137)
Observations	851,060	851,060	851,060	20,803	20,803	20,803	20,803
Mean Outcome	0.043	0.023	0.017	1.061	0.183	0.269	0.163
p-value of t-test:							
High = Low	0.000	0.000	0.000	0.166	0.250	0.163	0.311
High = Medium	0.000	0.412	0.000	0.254	0.639	0.726	0.817
Panel C: Tenure 13-18 years							
Low Misconduct	0.00433***	0.000344	0.00382***	-0.0164	0.0184	0.0274**	0.00248
	(0.000641)	(0.000492)	(0.000395)	(0.0215)	(0.0121)	(0.0134)	(0.0122)
Medium Misconduct	0.0117***	0.00156*	0.00907***	-0.0856**	0.0336*	0.0293	0.00254
	(0.00118)	(0.000848)	(0.000789)	(0.0360)	(0.0197)	(0.0218)	(0.0191)
High Misconduct	0.0408***	0.00672***	0.0281***	0.0407	0.0200	-0.00819	0.0504**
	(0.00229)	(0.00149)	(0.00158)	(0.0527)	(0.0236)	(0.0262)	(0.0243)
Observations	737,334	737,334	737,334	11,058	11,058	11,058	11,058
Mean Outcome	0.036	0.022	0.012	1.084	0.155	0.280	0.167
p-value of t-test:							
High = Low	0.000	0.000	0.000	0.286	0.947	0.190	0.055
High = Medium	0.000	0.002	0.000	0.061	0.645	0.252	0.102
Panel D: Tenure: 18 or more yrs							
Low Misconduct	0.00339***	0.00130**	0.00209***	-0.0189	-0.0180	0.0169	0.0154
	(0.000793)	(0.000634)	(0.000471)	(0.0386)	(0.0199)	(0.0223)	(0.0197)
Medium Misconduct	0.0160***	0.00474***	0.0103***	-0.179***	0.0693**	0.0401	-0.00268
	(0.00151)	(0.00114)	(0.000941)	(0.0663)	(0.0304)	(0.0314)	(0.0288)
High Misconduct	0.0175***	0.00724***	0.00992***	0.0168	-0.0153	-0.0631	0.0544
	(0.00236)	(0.00182)	(0.00146)	(0.0557)	(0.0372)	(0.0402)	(0.0434)
Observations	452,310	452,310	452,310	4,898	4,898	4,898	4,898
Mean Outcome	0.031	0.020	0.010	1.083	0.155	0.282	0.163
p-value of t-test:							
High = Low	0.000	0.001	0.000	0.537	0.943	0.054	0.371
High = Medium	0.582	0.230	0.831	0.008	0.052	0.034	0.237

Notes: This table depicts coefficient estimates for β_1, β_2, and β_3 from equation (1) for various outcomes, separately by officer tenure. Panel A reports estimates for officers with less than 8 years of tenure (25th percentile or below). Panel B report estimates for officers with 8-13 years of tenure (25-50th percentile). Panel C reports estimates for officers with 13-18 years of tenure (50-75th percentile). Panel D reports estimates for officers with 18 or more years of tenure (75th percentile or above). All estimates control for officer tenure and shift. Columns 4-7 also include the number of arrests, while Columns 6-7 also include the number of verdicts. Standard errors clustered by shift are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure A3: Are the results driven by "bad apples"?



Notes: This figure depicts coefficient estimates for β_1, β_2, and β_3 from equation (2), separately by "bad apple" status. "Bad apples" are defined as officers in the 95th percentile of the rate of complaints between 2010-2016. For the 95th percentile officers ("bad apples"), the regression model is: $y_{itr} = \beta_0 + \beta_1 Low_{i,r-1} + \beta_2 Medium_{i,r-1} + \beta_3 High_{i,r-1} + \beta_4 Not95th_i + \beta_5 \left(Low_{i,r-1} \times Not95th_i\right) + \beta_6 \left(Medium_{i,r-1} \times Not95th\right) + \beta_7 \left(High_{i,r-1} \times Not95th\right) + X'\delta + \varepsilon_{itr}$. For the rest of the officers, the regression model is equation (2). The estimates are relative to their respective reference groups: officers who are "bad apples" but had 0 complaints last year, and officers who are not "bad apples" and had 0 complaints last year. Wings are 95% confidence intervals with clustered standard errors.

Table A11: Regression results by "bad apple" status

Outcome Variable:	Any Arrest (1)	Any Index (2)	Any Non-Index (3)	Num Verdicts (4)	Shr No Charge (5)	Any Guilty (6)	Shr Not-Guilty (7)
Low Misconduct	0.00479***	0.000822***	0.00359***	-0.0148**	0.0141***	0.00147	0.000772
	(0.000320)	(0.000223)	(0.000218)	(0.00607)	(0.00347)	(0.00348)	(0.00324)
Medium Misconduct	0.0118***	0.00288***	0.00837***	-0.0285***	0.0223***	-0.000847	0.00368
	(0.000590)	(0.000395)	(0.000420)	(0.00914)	(0.00514)	(0.00511)	(0.00481)
High Misconduct	0.0175***	0.00346***	0.0124***	-0.000582	0.0186**	6.64e-05	0.00475
	(0.00105)	(0.000689)	(0.000758)	(0.0136)	(0.00742)	(0.00780)	(0.00736)
95th Percentile of Complaints	0.0320***	0.00556***	0.0231***	0.00531	0.0222***	0.00575	0.00305
	(0.00118)	(0.000728)	(0.000890)	(0.0130)	(0.00714)	(0.00713)	(0.00659)
95th pctile × Low	0.00179	0.00125	0.00216*	0.0275	-0.0421***	-0.000764	0.0232**
	(0.00168)	(0.00104)	(0.00129)	(0.0180)	(0.0100)	(0.0102)	(0.00952)
95th pctile × Medium	0.00764***	-0.000153	0.00542***	0.0372*	-0.0239**	0.0275**	0.00466
	(0.00194)	(0.00117)	(0.00149)	(0.0209)	(0.0113)	(0.0114)	(0.0107)
95th pctile × High	0.0125***	-0.000689	0.00973***	0.00505	-0.0168	-0.00203	0.0287**
	(0.00206)	(0.00124)	(0.00156)	(0.0214)	(0.0116)	(0.0119)	(0.0112)
Observations	3,209,376	3,209,376	3,209,376	133,447	133,447	133,447	133,447
Reference Group Mean	0.045	0.023	0.019	1.043	0.189	0.265	0.166

Notes: This table depicts coefficient estimates for β_1, β_2, and β_3 from equation (2) for various outcomes. "Bad apples" are defined as officers in the 95th percentile of the rate of complaints between 2010-2016. All estimates control for officer tenure and shift. Columns 4-7 also include the number of arrests, while Columns 6-7 also include the number of verdicts. Standard errors clustered by shift are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

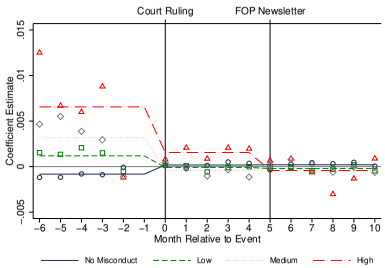
Table A12: Impact of Partner Misconduct on Conviction-less Arrests

Outcome Variable:	Complaints (1)	Any Arrest (2)	Any Index (3)	Any Non-Index (4)	Num Verdicts (5)	Shr No Charge (6)	Any Guilty (7)	Shr Not-Guilty (8)
Own Misconduct								
Low Misconduct	-0.000809***	-0.000301	-0.000583	0.000734	0.0190	-0.0103	0.00528	0.00368
	(0.000112)	(0.000673)	(0.000406)	(0.000500)	(0.0127)	(0.00715)	(0.00648)	(0.00613)
Medium Misconduct	-0.00154***	0.00298**	0.000175	0.00291**	-0.0139	-0.000340	-0.00782	-0.00180
	(0.000221)	(0.00142)	(0.000789)	(0.00115)	(0.0177)	(0.00968)	(0.00930)	(0.00952)
High Misconduct	-0.00190***	-0.000753	-0.000272	0.000355	-0.00893	0.00324	0.00133	0.00384
	(0.000490)	(0.00236)	(0.00114)	(0.00189)	(0.0227)	(0.0139)	(0.0128)	(0.0118)
Partner Misconduct								
Low-Misconduct Partner	0.000116	-0.000788	-0.00122***	0.000353	0.0182	-0.0103	0.00384	-0.000672
	(0.000103)	(0.000601)	(0.000374)	(0.000415)	(0.0112)	(0.00655)	(0.00647)	(0.00600)
Medium-Misconduct Partner	4.04e-05	-0.000228	6.63e-06	0.000615	-0.0284	0.0116	-0.0110	0.0124
	(0.000187)	(0.00103)	(0.000653)	(0.000759)	(0.0177)	(0.0102)	(0.00984)	(0.00975)
High-Misconduct Partner	0.000485	0.000936	-0.00148	0.00267**	-0.0112	-0.00301	-0.00140	-0.00705
	(0.000314)	(0.00171)	(0.000963)	(0.00126)	(0.0235)	(0.0134)	(0.0128)	(0.0126)
Interactions (Own × Partner)								
Low × Low-Partner	-0.000123	0.00188*	0.00101	0.00110	-0.0355*	0.0142	-0.00777	-0.00136
	(0.000192)	(0.00110)	(0.000665)	(0.000827)	(0.0192)	(0.0108)	(0.0107)	(0.00984)
Low × Medium-Partner	0.000519	0.00163	0.000120	0.00151	0.0321	-0.0126	0.0106	-0.0248*
	(0.000338)	(0.00213)	(0.00109)	(0.00174)	(0.0280)	(0.0162)	(0.0148)	(0.0144)
Low × High-Partner	-7.84e-05	0.00327	-2.50e-05	0.00208	-0.0185	0.0318	0.00674	0.0177
	(0.000540)	(0.00300)	(0.00152)	(0.00232)	(0.0377)	(0.0206)	(0.0201)	(0.0189)
Medium × Low-Partner	0.000489	-0.00140	-0.000166	-0.000632	-0.0205	0.0182	0.0120	-0.0206
	(0.000350)	(0.00206)	(0.00113)	(0.00157)	(0.0262)	(0.0141)	(0.0137)	(0.0129)
Medium × Medium-Partner	0.000445	0.000441	-0.000921	0.00126	0.0722**	-0.0187	0.00621	-0.00333
	(0.000477)	(0.00300)	(0.00160)	(0.00234)	(0.0351)	(0.0195)	(0.0187)	(0.0185)
Medium × High-Partner	0.000196	-0.00794**	0.000109	-0.00442	0.0526	-0.00881	-0.00887	0.0180
	(0.000714)	(0.00397)	(0.00204)	(0.00297)	(0.0401)	(0.0220)	(0.0210)	(0.0216)
High × Low-Partner	-8.54e-05	0.00573*	0.00108	0.00331	0.00938	0.00280	-0.0211	0.0240
	(0.000688)	(0.00309)	(0.00184)	(0.00246)	(0.0320)	(0.0184)	(0.0173)	(0.0154)
High × Medium-Partner	0.000925	0.00640	-0.000955	0.00565	0.00913	0.00667	0.00182	-0.0257
	(0.000907)	(0.00424)	(0.00199)	(0.00352)	(0.0338)	(0.0195)	(0.0208)	(0.0209)
High × High-Partner	-0.000349	0.0108**	-0.000283	0.00782**	0.0494	-0.00157	-0.0146	0.0157
	(0.000987)	(0.00466)	(0.00230)	(0.00374)	(0.0443)	(0.0219)	(0.0224)	(0.0214)
Observations	2,213,764	2,213,764	2,213,764	2,213,764	87,234	87,234	87,234	87,234
Reference Group Mean	0.002	0.044	0.022	0.019	1.034	0.200	0.263	0.163

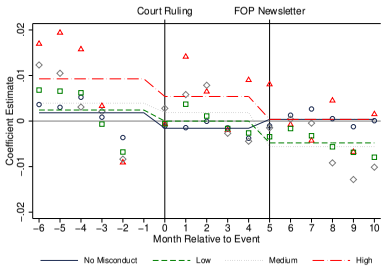
Notes: This table reports coefficient estimates from equation (3) for various outcomes. All estimates include fixed effects for officer and shift. Columns 4-7 also include the number of arrests. Columns 6-7 include the number of verdicts. Standard errors clustered by officer are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Figure A4: Monthly Impact of Increased Oversight on Conviction-less Arrests

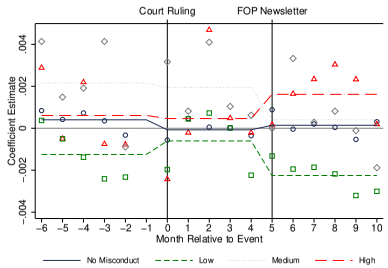
(a) Number of Complaints



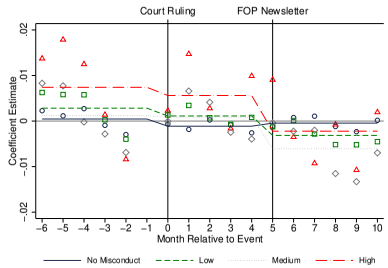
(b) Any Arrest



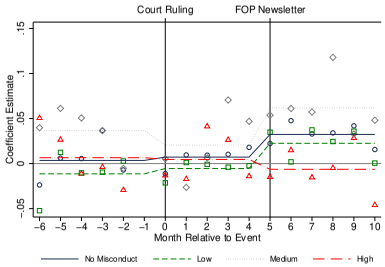
(c) Any Index Arrest



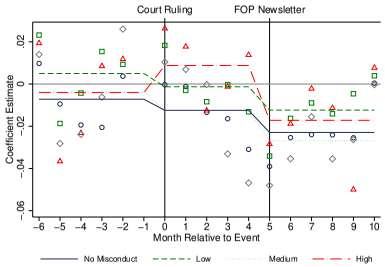
(d) Any Non-Index Arrest



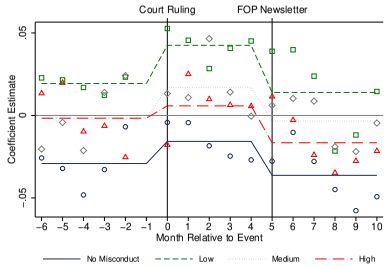
(e) Number of Verdicts



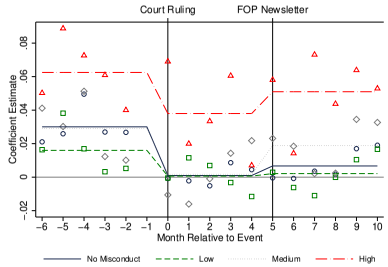
(f) Share with No Charges



(g) Any Guilty Verdict



(h) Share of Not Guilty Verdicts



Notes: This figure depicts estimates for τ_m from the regression model: $r_{it}^g = \beta_0^g + \tau_m^g + a_i^g + \varepsilon_{it}^g$, where τ_m denotes month fixed effects and α_i denotes officer fixed effects. The analysis sample is restricted to 6 months before event 1 (court ruling) and 5 months after event 2 (FOP newsletter). The reference month is February 2014, the month before the *Kalven v. City of Chicago* court ruling. Each symbol denotes a coefficient estimate, while lines are the mean value of the estimates within each of the three time periods.