The Impact of Internal Oversight on Arrest and Use of Force

Maya Mikdash*

Last updated: January 5, 2023

Abstract

While there is much desire for holding police accountable for misconduct, there is little evidence on whether the most common accountability system, internal affairs, impacts police behavior in intended or unintended ways. Using data from a large city where there is conditionally random assignment of officers to 911 calls, I employ regression discontinuity and difference-in-differences methods to distinguish the impact of investigations from confounding factors. Results indicate that increased oversight from internal investigations does not change an officer's likelihood of making an arrest or using force. This is true across different types of allegations, including those that are sustained. Surprisingly, even imposing sanctions after a sustained allegation does not change police behavior, irrespective of the severity of the sanction. This has important policy implications, as it suggests that the current system of internal oversight has no impact on police behavior.

*Texas A&M University; mmikdash@tamu.edu.

1 Introduction

In recent years, there has been significant interest in holding police officers accountable for misconduct, including and especially excessive use of force. While this is driven in large part by concern for those civilians who are mistreated, a growing literature has documented how police violence or misconduct can impact the community more broadly. For example, Ang 2021 and Legewie and Fagan 2019 show that police violence has negative consequences on educational outcomes such as test scores and high school completion. There is also growing evidence that police violence causes the relationship between police and communities to deteriorate, as reflected by reduced civilian cooperation (Ang et al. 2021; Desmond et al. 2016; Zoorob 2020). This distrust of police is also reflected in the fact that only 51 percent of civilians report having confidence in the police (Brenan 2021). Finally, there is concern that police violence and the deterioration of community trust that results can lead to de-policing (Mikdash and Zaiour 2022). Given the voluminous literature documenting how increased police reduces crime,[1] there is a desire to find methods of accountability that do not generate unintended consequences with respect to de-policing.

However, despite the interest in police accountability, there is little evidence to date on how police internal affairs investigations impact police behavior. This is despite the fact that nearly every large police department in the U.S. has an internal affairs unit, which explicitly serves the purpose of "policing the police" by receiving complaints against agency employees, investigating them, and taking the necessary disciplinary actions if an officer is found guilty (on Civil Rights 1981). Police officers perceive these investigations as unjustly handled, and civilians criticize them for exhibiting favoritism and a lack of transparency (Stephens 2016). Importantly, Rozema and Schanzenbach 2019 document that officers who receive the most internal affairs complaints are also those who generate the most in legal damages to the department. What is not clear, however, is whether

[1] See, for example, Levitt 1997; Di Tella and Schargrodsky 2004; Evans and Owens 2007; and Mello 2019; Weisburst 2019.

the investigations themselves, or their resulting findings or sanctions, have any impact on deterring misconduct, or even on generating unintended consequences. This is in large part due to data limitations. Full disciplinary records of police officers are only available in 12 states.[2] Moreover, even when records are available, they typically do not include non-sustained allegations, which make up more than 90 percent of all force complaints ((Reaves and Hickman 2000).

The purpose of this paper is to examine whether internal affairs investigations, and the determinations of guilt and sanctions that sometimes follow, impact police behavior with respect to the likelihood of using force or making an arrest. To do so, I obtained internal affairs investigations data from a US police department through a Freedom of Information Act (FOIA) request.[3] These data allow me to observe every complaint against officers from 2014-2021. In addition, I observe information about the allegation, whether it was sustained or not, and whether the officer was disciplined following the investigation. I link these investigations to 911 calls for service, which enable me to observe officer assignment as well as whether the calls end in an arrest or use of force.

The main advantage of examining this question in this city—other than the availability of the data—is that there is conditional random assignment of officers to calls in this city. In short, call takers are required to dispatch the nearest officer to calls, which means that conditional on police beat and time fixed effects, the assignment of the officer is random. This contrasts with some cities, where dispatchers and/or officers are given discretion in which officers respond to which 911 calls. As a result, there is less concern in this context about whether officers select into different types of interactions as a result of the internal affairs investigation, though I also provide empirical tests documenting the lack of such selection.

In order to distinguish the effect of the investigations from confounding factors, I use two different research designs. The first is a regression discontinuity design with time as

[2] https://project.wnyc.org/disciplinary-records/

[3] Per an agreement with the agency, I cannot publicly identify the city.

the running variable, which compares investigated officers right before and after a complaint was filed. In addition, I also show results from a difference-in-differences model in which I compare the behavior of officers who were and were not investigated, before and after the complaint.

Results from both research designs indicate that internal affair investigations do not affect officer's behavior when dispatched to a call. Estimates enable me to rule out reductions in use of force greater than 25 percent, and rule out reductions (increases) in the likelihood of making an arrest of 4 (6) percent. In addition, I examine effects separately by the quality of the arrest, in order to distinguish between "good" arrests, as proxied by those that are prosecuted, and "bad" arrests, which are not. This provides a direct test as to whether internal affairs investigations have the unintended consequence of de-policing, where officers may not make an arrest even though everyone may agree it is socially desirable to do so. Results indicate there is no effect on either type of arrest.

In addition, I use detailed information about the allegations and the investigations' outcomes to test for differential effects by whether the allegations were initiated by a civilian or an agency employee, and whether they were sustained. Surprisingly, I find no evidence of an effect even when the allegation is sustained. Finally, I examine whether punishing an officer matters, using observed disciplinary actions for most sustained allegations. Again, none of the sanctions affect officer behavior, even when they are as severe as a suspension. In short, results indicate that while internal oversight of police officers does not generate the unintended consequence of reducing police officer effort, it also has no effect on use of force rates.

To my knowledge, this is the first paper to provide evidence on the impact of internal oversight, which is the most common and arguably the most important way in which police officers are held accountable for misconduct, on police arrest and use of force behavior.[4] In

[4]In 2021, only 21 police officers were charged with murder or manslaughter resulting from an on-duty shooting, nationwide. This number was 16 in 2020, 12 in 2019, and 10 in 2018. On average, there are 1000 deadly police-involved shootings per year (Ortiz 2022).

doing so, it contributes to a larger literature on the impact of police reform more generally. Perhaps the most closely related paper is Rozema and Schanzenbach 2020, who examine the impact of sustained civilian allegations on officer misconduct, as measured by unsustained civilian allegations. Results here indicating that internal oversight has no effect on use of force or “bad quality" arrests differ from the findings of Rozema and Schanzenbach 2020, who find evidence of reduced complaints in response to a sustained civilian allegation. More generally, this paper contributes to the literature on the effect of oversight on police effort and crime (e.g., Prendergast 2001; Shi 2009; Premkumar 2019; Prendergast 2021; Ba and Rivera 2019).

There is a general consensus that an increase in oversight due to a public scandal causes “de-policing”, thus increasing crime. Most recently, this has been observed as a result of George Floyd (Mikdash and Zaiour 2022). As noted by Ba and Rivera 2019, however, public scandals also cause a change in civilian behavior—usually manifested in protests and public unrest—which makes it hard to distinguish the effect of increased oversight accompanying scandals from the effect of the change in civilian behavior. Other papers have looked at pattern-or-practice investigations (Chanin and Sheats 2018; Devi and Fryer Jr 2020), GPS-based monitoring strategies (Cheng and Long 2022) and more recently civilian oversight boards and police union memos (Jordan and Kim 2022; Ba and Rivera 2019).

The main contribution of this paper, especially relative to Rozema and Schanzenbach 2020 is that the dispatching protocol in the city allows me to rule out any changes in officer assignment that could potentially affect the outcomes of interest. The main concern here is that following an investigation, officers might self-select into different types of calls, perhaps less serious ones, to avoid undesirable consequences, such as a civilian complaint. In my city, however, call takers dispatch the geographically closest officer within a beat and a shift, creating a setting where officers’ assignments to calls is as-good-as-random. Additionally, using very detailed administrative data on 911 calls, I empirically show that indeed, the results are not driven by an incapacitation effect nor a change in the types of calls they respond to.

These results have important implications for policing and public policy. On the one hand, they indicate that commonly used internal accountability systems have not generated the type of unintended consequences, such as de-policing, that are widely believed to have been caused by scandals and the resulting public demands for accountability. On the other hand, results also demonstrate that the main system of accountability currently used by police departments across the country does not influence police officer behavior in intended ways. This suggests that there may still be work to do with respect to finding an accountability system that deters misconduct while causing minimal unintended consequences.

2 Background

According to the police department's manual on 'personnel policies" that describes the role of the internal affairs division, the division is responsible for "conducting all investigations involving deviations from or violations of the law, the Code of Ethics, Civil Service Rules and Regulations, police department policies and procedures, rules, and verbal or written orders or directives of supervisory personnel." In this paper, I focus on all types of complaints, 90 percent of which are investigated. These exclude "automatic" investigations such as investigations of officer-involved shootings.

Once a complaint is filed against an employee of the department, the internal affairs investigator provides the employee with a written statement of the allegations and information about their rights and responsibilities. In some instances, which are not specified in the manual, the employee is not informed of the allegation until right before the initial review. Similarly, the length of the period between the allegation receipt and the initial review is also not specified. As part of the investigation, the employee is interviewed, and can be subjected to a medical examination, a polygraph, a physical lineup, and/or a submission of financial disclosure statements.

There are four possible dispositions to any complaint: sustained, not sustained/inconclusive, exonerated, or unfounded. A complaint is sustained if "there is enough evidence to establish that the act occurred and that it constituted misconduct". If there is insufficient evidence, it is considered not sustained or inconclusive, and it is unfounded if the investigation discloses that the act did not occur. If the act occurred and it is within the law, the allegation is exonerated. For simplicity, I classify the complaints as either sustained or not sustained, where not sustained includes inconclusive, exonerated, and unfounded allegations. In my sample, the average time between the filing date and the finding date is 60 days. Once an investigation is concluded, the department follows a set of guidelines to determine the penalty level, if any. Possible penalties include written reprimand, training, counseling, reassignment suspension, demotion, or termination.

The amount of discipline depends on the nature of misconduct, the consequences of the violation, and any previous violations by the officer. The guidelines also suggest that they are flexible in determining the hours of suspension if the officer is being suspended, to encourage positive behavior. The maximum cumulative limit of suspension time in any 12-month period is 240 hours (30 days), and any further violation during suspension can result in termination. Finally, any penalty recommendation by the internal affairs division needs to be signed and approved by the Chief of Police, who is the only person who can decide to terminate or suspend an employee after receiving the recommendation. The Chief of Police also retains the right to depart from the guidelines of the disciplinary matrix and penalty table that are provided to determine the level of discipline.

3 Data

The law enforcement agency is from a city that ranks among the largest 30 cities in the U.S. by population, and has a population of over a half million. As a condition of receiving the data, the agreement stipulated that I cannot disclose the name of the city. Data from this city include the 911 calls for service, arrest and use of force data from 2014-2016, and internal affairs investigations from 2014-2021.

3.1 Internal affairs investigations

I obtain information on all internal affair investigations that were initiated by an allegation of misconduct anytime between 2014 to 2021. As Table A1 shows, 74 percent (776 officers) of the officers were investigated at least once between 2014 and 2021. Most of these officers are males (90 percent), and they have fewer years of experience for this police department compared to those who were never investigated in the sample period.

The data allow me to observe the type of the complaint, the result of the investigation (whether it was sustained or not), and the disciplinary action taken, if any. Most importantly, I can observe when the complaint is received, which I define as the treatment date, and full officer information, including their name. Finally, I can observe the date at which they conclude the investigation, or the finding date, for 60 percent of the investigations.

Unfortunately, I do not directly observe who submitted the complaint. However, I am able to infer this based on the type of allegation filed. Any allegations related to use of force, discrimination, unlawful stop, search or entry, and unlawful arrest, are categorized as "civilian" complaints. All other allegations are considered "internal". The latter are complaints that were most likely initiated by an agency employee, such as a supervisor. Examples of internal allegations include "dereliction of duty", "gossiping", "attention to duty", "conduct discrediting to the department", etc.

Table 1 shows summary statistics for the first complaint per officer, by finding. For example, if an officer was treated twice during the sample period, only the first allegation, i.e. the one that is used as the treatment, will be included in this table.[5] Summary statistics show that very few allegations are related to unlawful arrests (4 percent), discrimination (1 percent), or unprofessional conduct (5 percent). Most of the allegations are "internal" according to my classification (74 percent). Finally, around 20 percent are allegations of excessive use of force, and 13 percent are related to an unlawful search or entry.

Out of the 776 allegations, 52 percent are sustained, and almost all the sustained allegations result in some disciplinary action (99 percent). Consistent with anecdotal evidence, only 1 percent of sustained allegations are related to excessive use of force. Finally, the most common disciplinary action taken is counseling (63 percent of sustained allegations) and the least common is demotion (<1 percent).[6]

3.2 911 calls for service

In order to study the effect of internal oversight on officer behavior, I use 911 calls for service from 2014-2016, where I can observe officers' assignment to calls, call characteristics that are recorded by the call taker at the time of the call, and whether the call ended in an arrest or use of force. Call characteristics include a description (domestic violence, assault, robbery, etc. ...), a priority number, day and time, and the exact location.

One important feature of the 911 call dispatch system in this city is that officers do not have discretion over which calls to respond to. In contrary, call takers dispatch the geographically closest officer within a beat and a shift, meaning that ex ante, I should not have selection issues arising from officers choosing which calls to respond to. Since I observe detailed information about each call, I empirically show that indeed, officer assignment does not change following an investigation.

[5] There are 3,170 complaints that were filed during 2014-2021, and 2,107 complaints that were filed between 2014-2016. The majority of the investigated officers (89%) have more than one complaint filed against them between 2014-2021.

[6] I was not able to obtain information about the counseling or the training they undergo as a disciplinary action. Other actions include discussion of record, referred to bureau, or handled at the divisional level.

Finally, I observe the full names of the responding officers. Since officers can be investigated more than once, the treatment date is the date of the first allegation within the sample period. I merge the investigations data with the 911 calls by officer name.

3.3 Outcome variables

The two main outcomes of interest are arrest and use of force. In the arrest data, I observe the offense type (misdemeanor/felony), a description of the charges, and the police department disposition of each arrest. Specifically, I observe whether the arrest was declined for prosecution at the district attorney's office, which account for four percent of total arrests in the sample period. According to the literature, arrests have been used as a measure of police effort or productivity (Mas 2006). However, not all arrests are representative of (good) police effort. For instance, misconduct can be manifested in the form of false or unlawful arrests. These are proxied using arrests that are declined to be prosecuted. In this way, I am able to distinguish between arrests that are arguably "good"—that is, those that are prosecuted—and those that are arguably "bad" (i.e., not prosecuted). I use the police department's disposition of each arrest to identify those that were denied for prosecution as a proxy for police misconduct, and any other arrests as a proxy for police effort.

I link these outcomes to 911 calls for service by a unique incident ID in order to observe call outcomes. Since multiple officers can be dispatched to a given call, the data are at the call-by-officer level. The final data set allows me to observe officer assignment to calls for service, including call information, and treatment status, including whether and when an allegation of misconduct was made against and officer, the finding status, and the disciplinary action taken, if any. It is also important to note that since I only observe the outcome data between 2014-2016, I select the complaints that are filed within that sample period for treatment assignment.[7]

[7]Only 5 percent of the complaints are filed outside the 2014-2016 time window.

4 Empirical strategy

The difficulty in estimating the causal effect of internal oversight on police behavior is that they are nonrandom. For example, officers who are more active are more likely to commit error misconduct, and/or corrupt officers who are more likely to be investigated may be more aggressive, which means they use force at a higher rate than other officers and might make more arrests. Thus, comparing officers who are investigated to those who are not will yield biased estimates.

In order to overcome this issue, I use two empirical strategies: a regression discontinuity in time and a difference-in-differences, with the complaint filing date as discontinuity threshold (i.e. the treatment date), to estimate the effect of internal oversight on police behavior.

4.1 Regression discontinuity design

I estimate a regression discontinuity design in time regression (following Anderson 2014 and Bento et al. 2014). The benefit of using this model is that it allows me to estimate the short run changes in officer behavior caused by the increase in oversight. Since officers are made aware of the investigation and might take part of it, I expect that any impact of the investigation should be observed immediately following a complaint is filed. In particular, I estimate the following equation:

$$Y_{cit} = \alpha_0 + \alpha_1 * 1(Diff_{cit} \geq 0) + \alpha_2 * Diff_{cit} + \alpha_3 * 1(Diff_{cit} \geq 0) * Diff_{cit} + u_{cit} \tag{1}$$

where $Y_{cit} = (Arrest_{cit}; UseofForce_{cit})$ is a vector of the outcome variables. Y_{cit} equals 1 if call "c" at time "t" ends in an arrest or use of force by officer "i", and zero otherwise. The running variable, which represents the days since a complaint was received, is centered at zero, such that t = 0 is the day that the complaint against officer "i" was received. The variable of interest $1(diff_{cit} \geq 0)$ is an indicator variable that takes the value 1 if t$\geq$ 0 and zero otherwise. Thus, α_3 captures the discontinuity in the probability of an arrest or use of force caused by the complaint across time.

Since most complaints are investigated conditional on being made within a reasonable period of time after an incident, α_3 represents the effect of both the presence of a complaint and of that complaint being under investigation. Although I do not observe when the investigation is initiated, all investigations must be concluded within 100 days since receiving the complaint, according to departmental policy. Thus, I define the treatment date, or t=0, as the start date of the investigation.[8] For the main estimates, I use a local linear specification and show that the results are robust to using a quadratic specification in the appendix Table A2.

The identifying assumption of the regression discontinuity in time design, as for regression discontinuity designs more generally, is that all other determinants of the outcome vary smoothly across the threshold. One threat to identification in this context is the possibility that some officers are removed from the field or put on at least part-time desk duty during or after the investigation. To address this concern, I test directly for a discontinuity in the probability of being dispatched to any call after a complaint. Specifically, I create a variable $Dispatched_{it} = 1$ if an officer was dispatched to at least one call at any given day "t". If officers are re-assigned to desk duties after a complaint and during an investigation, I would expect to see a decrease in the probability of being dispatched to calls after the complaint date. A change in the probability of being dispatched indicates a compositional change in the sample.

A related concern is that the presence or conclusion of an investigation will change the *type* of calls to which an officer responds. However, one advantage of studying this particular city is that the dispatch protocol dictates that dispatchers assign the geographically closest officer to a call within a police beat. As a result, conditional on beat and shift, the

[8]Conditional on observing the finding date, investigations are concluded within 60 days from the day the complaint was filed.

assignment to a call is as-good-as-random, and the treatment status should be uncorrelated with the types of calls an officer is dispatched to. This means that it is difficult, and likely impossible, for an officer who remains in the field to alter the type of calls to which she must respond, as a result of the investigation. However, one might also worry that following a complaint against an officer, that officer may be assigned to a different beat where they take different types of calls.

To address these concerns, I do perform three exercises. First, I control for call characteristics to examine whether the main estimates are sensitive to that. Specifically, I control for month-by-year fixed effects, beat fixed effects, call priority, latitude, longitude, call type fixed effects, day of the week, time, and officer fixed effects. If officer assignment is not changing after a complaint, then adding the controls should not affect the main results.

Second, I use all observable call characteristics to estimate predicted arrest and use of force. Specifically, I regress arrest or use of force on beat and time fixed effects, call priority, latitude and longitude, call type, and day of the week. Using the resulting regression results, I predict the likelihood of force and arrest for each call. Then, I test whether predicted arrest or use of force changes after a complaint using Equation 1. If officer assignment is changing due to the complaint, then one would expect to see a discontinuity in the predicted use of force and/or predicted arrest. For example, if an officer is dispatched to less serious calls due to the complaint, then I would see a decrease in the predicted likelihood of using force and/or arrest. Finally, I directly estimate the effect of a complaint on each call characteristic, by estimating Equation 1 for each call characteristic separately, which should yield the same results as the second validity test (effect on predicted values).

I note that while some researchers have expressed concerns about the regression discontinuity design when time is the running variable, those concerns likely present minimal, if any, issues in this particular context. First, Hausman and Rapson 2018 argue that there is often insufficient data in these settings, which forces researchers to rely on observations far away from the cutoff in order to gain power, leading to potential bias. In my setting, there is a large number of treated officers during the sample period (776 officers), which combined with high-frequency data observed at the daily level allow me to avoid using large bandwidths.

A second potential concern is that in a regression discontinuity in time, unobservables correlated with the running variable can have discontinuous effects on the outcomes of interest.[9] I address this point directly in my study by including year-by-month and day-of-week fixed effects to account for potential confounders.

Finally, one would want to use the narrowest possible bandwidth to avoid biases arising from time-varying factors far from the cutoff date. In order to select the optimal bandwidth, I follow Calonico et al. 2020, using a uniform kernel and clustering standard errors at the officer level. This yields optimal bandwidths that vary from 135 to 380 days across the cutoff, depending on the outcome variable. In addition, I also report results using different bandwidths to show that the results are not sensitive to the bandwidth selection.

4.2 Difference-in-differences design

As an alternative approach to estimating the effect of internal investigations on officer behavior, I also implement a difference-in-differences design. This imposes a different identifying assumption than the regression discontinuity. In addition, it addresses any concern that the regression discontinuity approach captures only the short-run effect, whereas difference-in-differences can estimate a longer-run average treatment effect.

In order to implement the difference-in-differences approach, I estimate a two-way fixed effects model where I compare officers who were investigated at least once, to those who were never investigated in my sample period, before and after an investiga-

[9] Hausman and Rapson 2018 give the example of the power plant policy that required them to install emissions control devices. The policy was enacted on Monday, and they were interested in studying the effect of this policy on air pollution. However, air pollution varies discontinuously across the weekend and the beginning of the week.

tion. Specifically, I estimate the following regression:

$$Y_{cit} = \beta_0 + \beta_1 \times PostTreat_{cit} + \gamma_i + \sigma_t + \epsilon_{cit} \tag{2}$$

where β_1 represents the treatment effect, γ_i are officer fixed effects, and σ_t are month-by-year fixed effects.

The identifying assumption behind this model is that absent the complaint, the arresting and use of force behavior by officers who were investigated would have changed similarly compared to other officers who had never been investigated. Although I cannot directly test this assumption, I estimate a dynamic difference-in-differences model to see if there are any differential trends across both groups in the pre-period. Specifically, I estimate the following equation:

$$Y_{cit} = \beta_0 + \sum_{t=-8}^{8} \beta_t \times MonthsPost_{cit} + \gamma_i + \sigma_t + \epsilon_{cit} \tag{3}$$

where $MonthsPost_{cit}$ are indicator variables for months before and after a complaint. Each coefficient, β_t represents the effect over a one month period. If the empirical strategy is valid, I expect to see no divergence in the pre-trends acorss treated and control officers.

Similar to the regression discontinuity design, there is still the concern that after a complaint, they would choose to respond to different types of calls, or their assignment would change – causing selection bias or the results to be driven by a change in their assignment rather than an actual change in their behavior. To address these concerns, I estimate the difference-in-differences using the predicted values – explained above – as a proxy for call type. If investigated officers' assignment is changing after a complaint, then I would expect to see a significant change in predicted outcomes caused by the treatment.

I note that because multiple officers can be dispatched to a given call, the data are at the call-by-officer level. Additionally, the outcome variables are observed only at the call level, rather than the officer level. That is, I do not observe which officer used force or made an arrest. For these reasons, I weight each officer-by-call observation by the inverse of the number of officers dispatched to each call for both designs. For example, if two officers are dispatched to a call, each of the two observations in the data are assigned a probability weight of one-half. Finally, I report standard errors that are clustered at the officer level to allow observations to be correlated across calls for a particular officer.

5 Results – Regression discontinuity

I begin by examining whether internal oversight affects officer assignment. First, I estimate the effect of an investigation on the probability that an officer is dispatched to at least one call on a given day. Results in column 1 of Table 2 show a complaint has a statistically insignificant effect on the probability of being dispatched to any call. To the extent that the probability of being dispatched on a given day is a good proxy of officer assignment, this shows that officer assignment is not changing due to a complaint or an investigation, and the results should not be driven by an incapacitation effect.

Next, I estimate the effect on predicted arrest and predicted use of force, as proxies for call type. I report the results in Table 2. Columns (2) and (3) show that after a complaint, predicted arrest and predicted use of force are not changing. This is also supported with visual evidence in Figure 2, where I plot predicted arrest and predicted use of force as a function of time. As can be seen from the figures, there is no change in both outcomes after $t = 0$, which shows additional support that officer assignment to calls is not changing after a complaint. This indicates that the expected outcome of calls attended by officers, as determined by the observed characteristics of those calls, is unchanged as a result of the investigation.

Finally, I estimate the effect of a complaint on each call characteristic, those used to estimate predicted outcomes, separately. Results are reported in Table 3, where each column represents a separate regression equation for each call characteristic. As shown in the table, none of the coefficients are significant at conventional levels. These estimates are in line with the results from Table 2, where I show that the effect on predicted values is insignificant.

I plot the the raw outcome variables as a function of time in Figure 1, where the x-axis represents the days since the complaint was filed. The solid vertical line represents the date the complaint was filed (i.e., the 'treatment' date), while the dotted vertical line represents the average finding date. Conditional on observing the date at which the investigation is concluded, I estimate that on average, it takes 60 days for a finding to be made after a complaint is filed. Each data point represents the average outcome over a 14-day period. Graphically, both outcomes look constant around the cutoff, which suggests police behavior is not changing as a result of a complaint.

Next, I turn to estimating the effect of internal oversight on arrest and use of force. Formal regression discontinuity estimates from Equation 1 are shown in Table 4. Panel (A) uses a bandwidth of 191 days, which was computed to be the optimal bandwidth using the method proposed by Calonico et al. 2020. Panel (B) reports results from using a 382-day bandwidth (double the optimal bandwidth). The estimate in Column 1 of Panel (A) indicates that after a complaint, there is a 0.009 percentage point decrease in the probability of an arrest, which is statistically indistinguishable from zero at conventional levels. Results are qualitatively unchanged in Column 2 when controlling for call characteristics, officer fixed effects, and day of the week and month-by-year fixed effects.

Importantly, estimates are sufficiently precise to rule out moderate to large effects. For example, in the preferred specification in Column (2), estimates in Panel (A) indicate I can reject that investigations are associated with a reduction in the probability of arrest of more than four percent. Put differently, I can reject effects of the magnitude found by Jordan and Kim 2022 in response to civilian review boards investigations, as well as of the magnitude found in response to police ambushes documented by Sloan 2019.[10] As can be

[10]Jordan and Kim 2022 find that civilian review boards investigations led to a 4.4 percent reduction in arrests, while Sloan 2019 finds an 8 percent decrease in arrests following an ambush, which persists for at

seen in Panel (B), columns (1) and (2), the results are robust to using double the optimal bandwdith.

While total arrests have been used as a measure of police effort in the literature, not all arrests are necessarily indicative of “good" effort by police officers. To address this issue, I attempt to distinguish between “good" arrests—defined as those for which prosecutors choose to file charges—and “bad" arrests, or those in which no charges are filed. Results are reported in columns (3) to (6) of Table 4. Following a complaint, there is a 3 percent (-0.0001 p.p.) decrease in the probability of an arrest that is denied for prosecution (column 4) and less than a 1 percent increase in the probability of other arrests (column 6) (or the “good quality” arrests). Neither of these estimates is statistically significant at conventional levels. The results are robust to controlling for officer fixed effects, month-by-year fixed effects, and call characteristics. This suggests that officers are not responding to an increase in internal oversight by de-policing, though they are also not responding by improving their “good" effort. These results are robust to using twice the optimal bandwidth, as shown in Panel (B) of Table 4. These results suggest that investigations have no effect on either “good" effort, or “bad" effort with respect to making arrests.

Next, I examine the impact of internal affairs investigations on police use of force. Results are shown in Columns 7 and 8 of Table 4. Estimates indicate that after a complaint, there is no change in the likelihood of using force. This is true both across bandwidths (i.e., Panel (A) and Panel (B)), as well as without or with controls (Column 7 and 8). The estimate in Panel (A) of Column 7 indicates that filing a complaint is associated with a statistically insignificant five percent decrease in the likelihood of using force, relative to the mean. Importantly, the 95 percent confidence interval enables me to rule out effects one-third the magnitude of the impact of dispatching officers to different-race neighborhoods (Hoekstra and Sloan 2022). In fact, the precision of these estimates increase when I use double the optimal bandwidth. As can be seen in columns 7 and 8 of Panel (B),

---

least three years after an ambush.

the results are robust compared to the estimates in Panel (A), and they allow me to reject relatively small changes. For instance, I can reject increases (decreases) that are bigger than 16 percent (12 percent).

6 Results – Difference-in-differences

Before estimating the two-way fixed effects model, I plot the coefficients from Equation 3 in Figure 3. To the extent I observe no divergence in the pre-trends across both groups in the pre-investigation period, it provides some comfort that the two groups would also be unlikely to diverge post-investigation, except for the impact of the investigation. Figure 3 shows that before $t = 0$, there were no significant differences nor diverging trends across officers who were treated and those who were not. After the complaint date, both outcomes seem to be unchanging. This is consistent with the findings from the regression discontinuity design, as it suggests that internal investigations did not impact officer behavior in any way.

I estimate the difference-in-differences model using Equation 2 and report the results in Table 5 for both outcomes. Table 5 shows that after a complaint, the probability of an arrest increases by a statistically insignificant two percent. Similarly, the effect on use of force is very close to zero, in line with the results in Table 4. In comparison to the main RD estimates from Equation 1, the estimates are very similar. For both outcomes, although the point estimates are larger in magnitude when using the difference-in-differences model, both models suggest that internal oversight does not have a significant effect on arrest and use of force. Moreover, the estimates from the difference-in-differences model are not statistically different from those from the regression discontinuity design.

As with the regression discontinuity design, one threat to identification is if officers were dispatched to different calls after the investigation versus before, compared to other officers. In order to confirm this type of selection is not driving the results, I test for changes in call composition directly. First, I use all observed call characteristics to predict arrest and use of force. These include call hour, day of the week, month-by-year fixed effects, call type fixed effects, beat-by-shift fixed effects, and call priority. Then, I estimate the effect of a complaint on the predicted outcomes using equation Equation 2, and I plot the coefficients with their 95% confidence intervals in Figure 3. Results in Figure 3 indicate that being subject to an investigation does not impact the composition of calls to which an officer is dispatched.

Overall, the results in section 5 and section 6 show that increased oversight due to an internal affairs investigation of a complaint does not affect officer behavior when dispatched to a call. Internal oversight does not cause a decrease in the likelihood of using force or the likelihood of a "bad" arrest, but it also does not cause de-policing. Using both methods, I can reject very small changes in arrest, and relatively small changes in use of force.

7 Heterogeneity

7.1 Complaint type

While the results above demonstrate that investigations do not have any impact on arrest or use of force outcomes on average, it is possible that the behavioral response depends on the complaint type, its seriousness, and who the complainant is. For instance, allegations that are sustained, or that are followed by disciplinary action, may well induce a larger behavioral response. Similarly, it is possible that allegations by civilians may induce larger or smaller effects than allegations by someone in the police department.

Results by complaint type for arrests and use of force are shown in Panels (A) and (B) of Table 6, respectively. Columns 1 and 2 show the effect of any sustained complaint, columns 3 and 4 show the effect of a sustained complaint that was made by a civilian, and columns 5 and 6 show the effect of a sustained internal complaint. As can be seen from column 1, sustained complaints do not have an effect on either arrest nor use of force. This result is robust to adding officer fixed effects, month-by-year fixed effects, and call characteristics. For example, column 2 shows that after a sustained complaint, the likelihood that a 911 call for service results in an arrest increases by less than a 0.1 p.p., which is a 1 percent increase relative to the control mean. This estimate is insignificant at conventional levels. This is also true for use of force. The results in panel (B) of column 1 show that after a sustained complaint, the likelihood of using any force increases by decreases by less than 1 percent, relative to the control mean, and this decrease is insignificant with and without including all controls (Panel B, column 2).

As for civilian complaints, column (3) shows that after a sustained civilian complaint, there is an imprecise 9.5 percent decrease in the likelihood of arrest. Adding officer fixed effects, month-by-year fixed effects, and call characteristics to increase precision, column (4) shows suggestive evidence that a sustained civilian arrest causes a 17 percent decrease in the likelihood of an arrest, conditional on being dispatched to a call. As for use of force, a sustained civilian complaint has no effect on using force, as shown in columns (3) and (4) of Panel B. Finally, column (5) shows that a sustained internal complaint has an insignificant effect on arrest. However, when I add full controls in column (6), there is suggestive evidence that a sustained internal complaint increases the likelihood of an arrest by 7 percent. Similar to a civilian complaint, it does not have an effect on using force (columns (5) and (6)).

Overall, results from Table 6 indicate that even sustained allegations do not affect the arrest and use of force behavior of officers, regardless of who the complainant is. Out of the twelve estimates reported in the table, none are significant at the 5 percent level, and only two, in opposite directions, are significant at the 10 percent level.

7.2 Discipline type

As shown in Table 1, most officers with a sustained complaint end up being disciplined (98%). However, disciplinary actions vary by severity, the rarest and most severe of which can be demotions.[11] As a result, it is possible that while there is no effect of sustained investigations on average, perhaps investigations that are followed by severe penalties do impact police behavior.

Results in Table 7 show the regression discontinuity estimates by the type of disciplinary action taken. In particular, in Columns (1) - (4) I show results for suspensions, counseling, training, and written reprimands. Results for both arrests and use of force, in Panels A and B, respectively, indicate that there is no evidence of effects for any of these disciplinary actions. For example, the estimate in Column (1) of Panel A indicates that the most severe penalty—suspension—is associated with a 24 percent *increase* in arrests and a 10 percent *increase* in use of force. Thus, while it is possible that even more severe penalties—such as demotions, only two of which are observed in my data—would generate either de-policing or reductions in use of force, evidence in Table 7 indicates that commonly used disciplinary actions do not.

8 Robustness Checks

8.1 Bandwidth selection

One potential concern for any regression discontinuity design is the extent to which the results depend on bandwidth selection. Additionally, as alluded to earlier, Hausman and Rapson 2018 argue that it is especially important to focus on small bandwidths when using time as the running variable. As a result, I estimate discontinuities using ten different bandwidths, including the one I use for the main specification, and report the coefficient

[11] I observe only two demotions in my data.

estimates with the 95% confidence intervals for both outcome variables in Figure A1. The smallest bandwidth I use is 50 days, and the largest one is 500 days pre- and post- treatment. Although the estimates are less precise when I use a narrower bandwidth of 50 days, the figure shows that the results are not sensitive to bandwidth selection for both outcomes of interest. Independent of the selected bandwidth, the effect of a complaint on both arrest and use of force is centered around zero.

8.2 Quadratic specification

Following Gelman and Imbens 2019, I use a local quadratic function of the time variable as a robustness check. As can be seen from panels (A) and (B) of Table A2, the results are robust compared to the linear specification estimates in for both outcomes. For example, column 2 shows that an investigation causes a statistically insignificant 2 percent (5 percent) increase in the probability of an arrest (using force) when using a quadratic function, which are not statistically different from the main estimates in column (1).

8.3 Donut regression discontinuity design

I estimate two donut regression discontinuity models to account for any anticipation effect. For example, officers might anticipate that they will be investigated for misconduct prior to the treatment date, so they decide to decrease their interaction with civilians by arresting less people or using less force. For robustness, I estimate a donut RD where I omit the treatment date (Donut 1), and a donut RD where I omit the last week right before treatment, including the treatment date (Donut 2). Columns (3) and (4) of Table A2 suggest that the results are robust to omitting observations around the cutoff date for both outcomes. This is also consistent with the graphical evidence in Figure 1, where both outcome variables seem to be steady around the treatment date.

8.4 Using finding date

The main estimates from the regression discontinuity design reflect the immediate response to being under an investigation. Thus, one potential concern is that an investigation might not have an effect until the complaint is actually sustained. In order to examine whether sustaining a complaint, rather than being under investigation, has an effect, I use the finding date as the treatment date (t=0), i.e. the date at which the complaint was sustained. Conditional on observing the finding date (I only observe it for 60% of the sample), I estimate the main equation (Equation 1) using the finding date as the cutoff date. This means that $1(\text{after} \geq 0)$ is equal to 1 on or after the day at which the complaint was sustained.

I report the results in Table A3. Similar to the results in Table 6, sustaining a complaint does not have any effect on the probability of arrest or use of force. Again, there is some suggestive evidence that the likelihood of arrest decreases following a sustained civilian complaint (17 percent), but this effect is not robust to adding controls (column 4). Similarly, the effect of a sustained internal complaint is insignificant once I use the finding date as the treatment date. The effect on use of force remains insignificant across different types of complaints.

In summary, the results of this section show that the estimates are not sensitive to bandwidth selection, assumptions regarding the relationship between the outcome variables and the running variable, or the choice of the treatment date. As a result, I conclude that internal oversight due to the investigation of a complaint does not change officer behavior in intended, or unintended, ways.

9 Conclusion

In this paper, I study the effect of internal affair investigations on police behavior, as measured by arrest and use of force. I do so using two different research designs. Results from both a regression discontinuity design and a difference-in-differences analysis indicate that internal affairs investigations have no impact on police behavior with respect to arrest or use of force. I find no evidence that this form of oversight, which is the most common and arguably the most important system of police accountability in the country either causes de-policing or reduction in misconduct or use of force. This is true for any type of allegation, regardless if they were sustained or not, and regardless whether they resulted in a punishment, such as a suspension or a written reprimand.

The findings of this paper conflict with much of the existing literature on police oversight. On one hand, increased oversight because of a public scandal or a pattern-or-practice investigation cause de-policing (e.g., Mikdash and Zaiour 2022; Cheng and Long 2022; Devi and Fryer Jr 2020). On another hand, other types of oversight, usually those not involving heightened public attention, can improve police behavior and productivity without causing de-policing, such as police union memos (Ba and Rivera 2019), civilian review boards (Jordan and Kim 2022),and GPS monitoring (Cheng and Long 2018).

Compared to these papers, I can reject effects that are one-fifth to one-sixth the magnitude they find. For example, Cheng and Long 2022 show that high-profile police killings lead to a 26.4 percent decrease in arrests over a sample of 60 large US cities. This is in line with Shi 2009, where increased oversight due to the Cincinnati riots, combined with a civil-rights investigation and increased media attention, led to a 36 percent decrease in misdemeanor arrests. In contrast, civilian review boards investigations decrease complaints by about 15 percent (Jordan and Kim 2022), and sustained civilian allegations decrease future civilian allegations by 40 percent (Rozema and Schanzenbach 2020). Although I do not use the same measure of misconduct, I see no evidence of a decrease in use of force or a decrease in “bad” arrests. Relative to Jordan and Kim 2022 and Rozema and Schanzenbach 2020, I find no evidence of a decrease in misconduct.

The results of this paper have important implications for self-governance in policing. Primarily, it is clear that the main accountability system used by the police does not lead to reductions in use of force or in arrests that are not prosecuted. It is difficult to know exactly why this is the case, but there can be multiple possibilities, such as insufficient penalties, in-group bias within the internal affairs units, or union involvement, etc.[12] Finally, discussions around police oversight suggest a tradeoff between holding officers accountable, their productivity, and their safety. The results here show that internal affairs investigations do not affect police misconduct, but they also do not cause de-policing.

[12]Evidence regarding police union contracts and misconduct is mixed (Cunningham et al. 2021; Goncalves 2020).

References

Anderson, Michael L, “Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion,” *American Economic Review*, 2014, *104* (9), 2763–96.

Ang, Desmond, “The effects of police violence on inner city students,” *The Quarterly Journal of Economics*, 2021, *136* (1), 115–168.

_ , Panka Bencsik, Jesse Bruhn, and Ellora Derenoncourt, “Police violence reduces civilian cooperation and engagement with law enforcement,” https://scholar.harvard.edu/files/ang/files/abbd_crimereporting.pdf 2021.

Ba, Bocar A and Roman Rivera, “The effect of police oversight on crime and allegations of misconduct: Evidence from chicago,” *U of Penn, Inst for Law & Econ Research Paper*, 2019, (19-42).

Bento, Antonio, Daniel Kaffine, Kevin Roth, and Matthew Zaragoza-Watkins, “The effects of regulation in the presence of multiple unpriced externalities: Evidence from the transportation sector,” *American Economic Journal: Economic Policy*, 2014, *6* (3), 1–29.

Brenan, Megan, “Americans’ Confidence in Major U.S. Institutions Dips,” https://news.gallup.com/poll/352316/americans-confidence-major-institutions-dips.aspx/ 2021.

Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell, “Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs,” *The Econometrics Journal*, 2020, *23* (2), 192–210.

Chanin, Joshua and Brittany Sheats, “Depolicing as dissent shirking: Examining the effects of pattern or practice misconduct reform on police behavior,” *Criminal Justice Review*, 2018, *43* (2), 105–126.

Cheng, Cheng and Wei Long, “Improving police services: Evidence from the French quarter task force,” *Journal of Public Economics*, 2018, *164*, 1–18.

_ and _ , “The effect of highly publicized police killings on policing: Evidence from large US cities,” *Journal of Public Economics*, 2022, *206*, 104557.

Cunningham, Jamein, Donna Feir, and Rob Gillezeau, “Collective bargaining rights, policing, and civilian deaths,” 2021.

Desmond, Matthew, Andrew V Papachristos, and David S Kirk, “Police violence and citizen crime reporting in the black community,” *American sociological review*, 2016, *81* (5), 857–876.

Devi, Tanaya and Roland G Fryer Jr, “Policing the police: The impact of" pattern-or-practice" investigations on crime,” Technical Report, National Bureau of Economic Research 2020.

Evans, William N and Emily G Owens, "COPS and crime," *Journal of Public Economics*, 2007, *1* (91), 181–201.

Gelman, Andrew and Guido Imbens, "Why high-order polynomials should not be used in regression discontinuity designs," *Journal of Business & Economic Statistics*, 2019, *37* (3), 447–456.

Goncalves, Felipe, "Do police unions increase misconduct," Technical Report, Working paper 2020.

Hausman, Catherine and David S Rapson, "Regression discontinuity in time: Considerations for empirical applications," *Annual Review of Resource Economics*, 2018, *10*, 533–552.

Hoekstra, Mark and CarlyWill Sloan, "Does race matter for police use of force? Evidence from 911 calls," *American Economic Review*, 2022, *112* (3), 827–60.

Jordan, Andrew and Taeho Kim, "The Effects of Investigating Police Misconduct," *Available at SSRN 4099052*, 2022.

Legewie, Joscha and Jeffrey Fagan, "Aggressive policing and the educational performance of minority youth," *American Sociological Review*, 2019, *84* (2), 220–247.

Levitt, Steven D, "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," *American Economic Review*, June 1997, *87* (3), 270–290.

Mas, Alexandre, "Pay, reference points, and police performance," *The Quarterly Journal of Economics*, 2006, *121* (3), 783–821.

Mello, Steven, "More COPS, less crime," *Journal of public economics*, 2019, *172*, 174–200.

Mikdash, Maya and Reem Zaiour, "Does (All) Police Violence Cause De-policing? Evidence from George Floyd and Police Shootings in Minneapolis," in "AEA Papers and Proceedings," Vol. 112 2022, pp. 170–73.

on Civil Rights, United States Commission, *Who is Guarding the Guardians?: A Report on Police Practices: a Report of the United States Commission on Civil Rights*, The Commission, 1981.

Ortiz, Erik, "More officers were charged in fatal police shootings in 2021. Not everyone sees progress.," https://www.nbcnews.com/news/us-news/officers-charged-fatal-police-shootings-2021-not-everyone-sees-progres-rcna12799 2022.

Premkumar, Deepak, "Public Scrutiny and Police Effort: Evidence from Arrests and Crime After High-Profile Police Killings," *Available at SSRN 3715223*, 2019.

Prendergast, Canice, "Selection and oversight in the public sector, with the Los Angeles Police Department as an example," 2001.

— , "'Drive and Wave': The Response to LAPD Police Reforms after Rampart," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2021, (2021-25).

Reaves, Brian A and Matthew J Hickman, *Law Enforcement Management and Administrative Statistics, 2000: Data for Individual State and Local Agencies with 1 00 or more Ojficers*, Washington: US Department of Justice, 2000.

Rozema, Kyle and Max M Schanzenbach, "Does Discipline Decrease Police Misconduct? Evidence from Chicago Civilian Allegations," *Evidence from Chicago Civilian Allegations (August 7, 2020)*, 2020.

— **and Max Schanzenbach**, "Good Cop, Bad Cop: Using Civilian Allegations to Predict Police Misconduct," *American Economic Journal: Economic Policy*, 2019, *11* (2), 225–268.

Shi, Lan, "The limit of oversight in policing: Evidence from the 2001 Cincinnati riot," *Journal of Public Economics*, 2009, *93* (1-2), 99–113.

Sloan, CarlyWill, "The Effect of Violence Against Police on Policing Behavior," 2019.

Stephens, Darrel W, "Police discipline: A case for change," 2016.

Tella, Rafael Di and Ernesto Schargrodsky, "Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack," *American Economic Review*, 2004, *94* (1), 115–133.

Weisburst, Emily K, "Safety in police numbers: Evidence of police effectiveness from federal COPS grant applications," *American Law and Economics Review*, 2019, *21* (1), 81–109.

Zoorob, Michael, "Do police brutality stories reduce 911 calls? Reassessing an important criminological finding," *American sociological review*, 2020, *85* (1), 176–183.

10 Tables and Figures

Table 1: Complaints Characteristics

	(1) All Allegations	(2) Sustained	(3) Not Sustained
Allegation			
Unlawful Arrest/Detention	0.0361	0.00988	0.0647
	(0.187)	(0.0990)	(0.246)
Unlawful Search/Entry	0.128	0.106	0.151
	(0.334)	(0.308)	(0.358)
Excessive Use of Force	0.198	0.0123	0.402
	(0.399)	(0.111)	(0.491)
Discrimination	0.0129	0	0.0270
	(0.113)	(0)	(0.162)
Unprofessional/Verbal	0.0451	0.0148	0.0782
	(0.208)	(0.121)	(0.269)
Non-civilian Allegation	0.744	0.862	0.615
	(0.437)	(0.346)	(0.487)
Action Taken			
Disciplined	0.518	0.985	0.00809
	(0.500)	(0.121)	(0.0897)
Demotion	0.00258	0.00494	0
	(0.0507)	(0.0702)	(0)
Counseling	0.332	0.637	0
	(0.471)	(0.481)	(0)
Training	0.0838	0.143	0.0189
	(0.277)	(0.351)	(0.136)
Written Reprimand	0.0438	0.0840	0
	(0.205)	(0.278)	(0)
Suspension	0.0554	0.106	0
	(0.229)	(0.308)	(0)
Other Action	0.0219	0.0370	0.00539
	(0.146)	(0.189)	(0.0733)
Observations	776	405	371

Standard deviations in parentheses

Notes: This table shows the summary statistics for the first complaint that has ever been filed against any officer within the 2014-2016 sample period. Column (1) shows the characteristics of all allegations, column (2) shows the characteristics of sustained complaints, while column (3) shows the characteristics of the complaints that were not sustained.

Table 2: Effect of Internal Oversight on Officer Assignment

	(1) Dispatched	(2) Predicted Arrest	(3) Predicted UOF
After Complaint	-0.00658	0.000228	-0.0000686
	(0.00571)	(0.00104)	(0.000105)
N	153324	168835	148354
Control Mean	0.245	0.0569	0.00200
Bandwidth	147.9	129.1	110.5

Standard errors in parentheses

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

Notes: This table shows the effect of a complaint on officer assignment using three main variables. I generate a variable "dispatched" that represents the probability that officer "i" is dispatched to at least 1 call at time "t", where time is measured in days. Next, using call characteristics, I predict the probability of arrest and use of force for each call "c". Specifically, I regress arrest and use of force on call description fixed effects, day of the week fixed effects, month-by-year fixed effects, beat-by-shift fixed effects, priority of the call, and call hour to estimate the predicted values. Using Equation 1, I estimate the effect of a complaint on these three outcomes separately, specifying an optimal bandwidth using a uniform kernel. I also report the control mean, which is the average of the outcomes in the pre-period. Standard errors are clustered at the officer level and reported in parentheses.

Table 3: Correlation between Internal Oversight and Call Characteristics

	(1) Day of Week	(2) Dispatched Hour	(3) Shift	(4) Call Type	(5) Latitude	(6) Longitude	(7) Percent Black
After Complaint	0.00653	0.179	0.00450	0.552	-0.000653	-0.00356	-0.000643
	(0.0522)	(0.176)	(0.0198)	(0.814)	(0.000709)	(0.00295)	(0.000409)
N	235882	199568	192001	185959	205951	231696	197675
Control Mean	3.113	13.01	0.963	140.3	31.07	-211.3	0.0354
Bandwidth	201.5	160.2	152.1	146.5	167.3	196.5	159

Standard errors in parentheses

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

Notes: This table reports the coefficient on variable $1(Diff_{cit} \geq 0)$ from Equation 1, where I estimate the effect of a complaint on each call characteristic separately, as a validity check. Each column represents a separate regression estimate. Shift is a variable that takes the value 0 if the call happens between 6 pm and midnight, the value 1 if it is between midnight and 7 am, and the value 2 if it is between 8 am and 4 pm. Call type is a categorical variable that represents different types of calls. I use optimal bandwidth estimations and report the control mean for each variable, i.e. the average outcome in the pre-period. Standard errors are clustered at the officer level and reported in parentheses.

Table 4: Effect of Internal Oversight on Arrest and Use of Force

	Arrest, by disposition							
	Any Arrest		Denied for Prosecution		Other		**Use of Force**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: 1× Optimal bandwidth								
After Complaint	-0.0000852	0.000376	-0.0000861	-0.000103	-0.000371	0.000342	-0.0000954	-0.000111
	(0.00167)	(0.00148)	(0.000394)	(0.000407)	(0.00156)	(0.00138)	(0.000198)	(0.000208)
N	227388	227388	207835	207835	237879	237879	175098	175098
Control Mean	0.0568	0.0568	0.00305	0.00305	0.0538	0.0538	0.00201	0.00201
ATE(%)	-0.150	0.662	-2.825	-3.368	-0.691	0.637	-4.742	-5.535
Bandwidth	191.3	191.3	169.8	169.8	203.6	203.6	135.1	135.1
Panel B: 2× Optimal bandwidth								
After Complaint	-0.000757	0.000224	-0.000122	-0.000164	-0.000149	0.000608	0.000128	0.0000331
	(0.00135)	(0.00117)	(0.000329)	(0.000367)	(0.00126)	(0.00111)	(0.000134)	(0.000145)
N	371566	371566	342282	342282	388146	388146	291532	291532
Control Mean	0.0568	0.0568	0.00305	0.00305	0.0538	0.0538	0.00201	0.00201
ATE(%)	-1.332	0.394	-4.014	-5.368	-0.277	1.130	6.346	1.644
Bandwidth	382.7	382.7	339.5	339.5	407.2	407.2	270.1	270.1
Officer FE		Y		Y		Y		Y
Month× Year FE		Y		Y		Y		Y
Call Characteristics		Y		Y		Y		Y

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

Notes: This table shows the RDiT estimates from Equation 1 with arrest and use of force as outcomes. The first two columns show the effect of a complaint on any type of arrest, columns (3) and (4) show the effect on arrests that were denied for prosecution, while columns (5) and (6) show the effect on all other arrests (i.e. arrests that were not denied for prosecution, along with arrests with other disposition types, such as release on bond, transferred to another party, indicted, etc. ...). Additionally, columns (7) and (8) report the effect on use of force of any type. The odd columns (1), (3), (5), and (7) represent the results excluding any controls, while the even columns (2), (4), (6), and (8) show the results using officer fixed effects, month-by-year fixed effects, and all call characteristics, including call description, day of the week, call hour, priority of the call, Longitude and Latitude. In panel (A), I use the optimal bandwidth estimated separately for each outcome following Calonico et al. 2020 (rdbwselect command in Stata). For robustness, I report the results using 2× the optimal bandwidth in panel (B). The control mean represents the average outcomes in the pre-period. For all estimates, standard errors are clustered at the officer level and reported in parentheses.

Table 5: Difference-in-differences estimates

	Arrest		Use of Force	
	(1)	(2)	(3)	(4)
After Complaint	0.00181	0.00126	0.0000893	0.0000570
	(0.00112)	(0.000895)	(0.000107)	(0.000104)
N	711595	711595	711595	711595
Control Mean	0.0577	0.0577	0.00232	0.00232
ATE(%)	3.131	2.190	3.846	2.456
Month$\times FE$	Y	Y	Y	Y
Officer FE	Y	Y	Y	Y
Full Controls		Y		Y

Standard errors in parentheses

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

Notes: This table shows the difference-in-differences estimates from Equation 2, where I compare investigated officers to those who are never investigated, before and after an investigation. Columns (1) and (2) show the effect on arrest, while columns (3) and (4) show the effect on use of force. All columns include month-by-year fixed effects and officer fixed effects. Columns (2) and (4) include all call characteristics, including day-of-week fixed effects, call hour, call type, priority, shift, Latitude and Longitude. Standard errors are clustered at the officer level.

Table 6: Effect of Sustained Complaints on Arrest and Use of Force

	Sustained					
	Any Complaint		Civilian Complaint		Internal Complaint	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Arrest						
After Complaint	0.00124	0.000778	-0.00579	-0.0103*	0.00278	0.00361*
	(0.00225)	(0.00199)	(0.00581)	(0.00582)	(0.00233)	(0.00206)
N	126598	126598	26392	26392	103339	103339
Control Mean	0.0530	0.0530	0.0607	0.0607	0.0516	0.0516
ATE(%)	2.349	1.468	-9.541	-16.98	5.390	6.991
BW	172.9	172.9	233.9	233.9	168	168
Panel B: Use of Force						
After Complaint	-0.0000714	-0.000295	-0.0000677	-0.000221	-0.0000736	-0.000362
	(0.000193)	(0.000230)	(0.000337)	(0.000508)	(0.000226)	(0.000264)
N	126598	126598	24696	24696	99539	99539
Control Mean	0.0530	0.0530	0.0609	0.0609	0.0517	0.0517
ATE(%)	-0.135	-0.558	-0.111	-0.363	-0.142	-0.700
Bandwidth	172.4	172.4	212.1	212.1	160.9	160.9
Officer FE		Y		Y		Y
Month× Year FE		Y		Y		Y
Call Characteristics		Y		Y		Y

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

Notes: This table shows the effect of a sustained complaint on arrest and use of force. I report the coefficient on $1(Diff_{cit}) \geq 0$ using complaint filing date as t=0. Columns (1) and (2) show the effect of any sustained complaint, columns (3) and (4) show the effect of a sustained civilian complaint, while columns (5) and (6) show the effect of a sustained internal complaint. I classify complaints into "civilian" and "internal" using the allegation type. Civilian complaints include allegations of excessive use of force, discrimination, unlawful arrest, stop, search, or entry. Internal complaints include allegations of violating the code of conduct, gossiping, failing to file a report, conduct discrediting to the department, failure to appear in court, or violation of safety practices. Panel (A) shows the effect on arrest, while panel (B) shows the effect on use of force. Odd columns include officer fixed effects, month-by-year fixed effects, and call characteristics controls. Standard errors are clustered at the officer level and reported in parentheses.

Table 7: Effect of Disciplinary Actions on Arrest and Use of Force

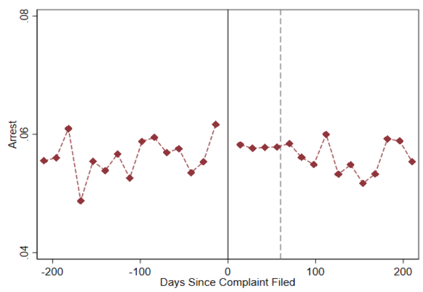
	(1) Suspension	(2) Counseling	(3) Training	(4) Written Reprimand
Panel A: Arrest				
After Complaint	0.0123	-0.000993	0.00175	-0.0119
	(0.0106)	(0.00230)	(0.00518)	(0.00877)
N	9343	80132	22617	6465
Control Mean	0.0513	0.0529	0.0573	0.0570
ATE(%)	23.90	-1.876	3.061	-20.84
BW	161.5	170.3	177.6	87
Panel B: Use of Force				
After Complaint	0.0000490	-0.0000304	0.000205	0.000717
	(0.000690)	(0.000201)	(0.000307)	(0.000474)
N	10384	122817	27037	11095
Control Mean	0.0511	0.0525	0.0576	0.0506
ATE(%)	0.0959	-0.0580	0.356	1.416
Bandwidth	187.1	296.3	225.6	172.1
Officer FE	Y	Y	Y	Y
Month× Year FE	Y	Y	Y	Y
Call Characteristics	Y	Y	Y	Y

Standard errors in parentheses

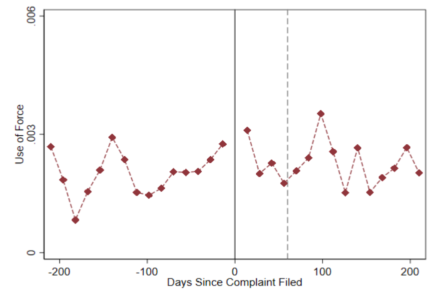
* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: This table shows the effect of different types of disciplinary actions on arrest and use of force. I report the coefficient on $1(Diff_{cit} \geq 0)$ using complaint filing date as t=0. Each column represents the effect of a unique discipline type. Column (1) shows the effect of a suspension, column (2) shows the effect of counseling, column (3) shows the effect of training, while column (4) shows the effect of a written reprimand. These include All columns include officer fixed effects, month-by-year fixed effects, and call characteristics controls. Standard errors are clustered at the officer level and reported in parentheses.

Figure 1: Effect of Internal Oversight on Arrest and Use of Force



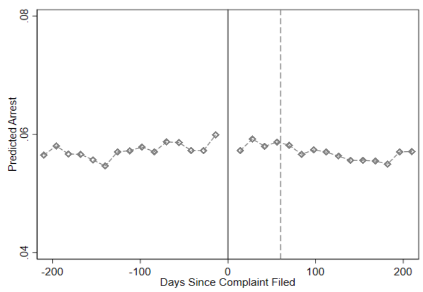
(a) Arrest



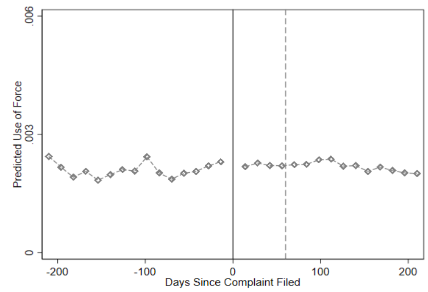
(b) Use of Force

Notes: Figures (a) and (b) show the raw data for both outcomes over time. Each point represent the average outcome over a 14-day period. Time is normalized with respect to the filing date of the complaint, such that t=0 represent the date at which a complaint is filed, and this is represented by the black vertical line. The gray dashed line represents the average time at which a finding is reached. Conditional on observing the finding date, the average finding date is 60 days after a complaint has been filed.

Figure 2: Effect of Internal Oversight on Predicted Outcomes



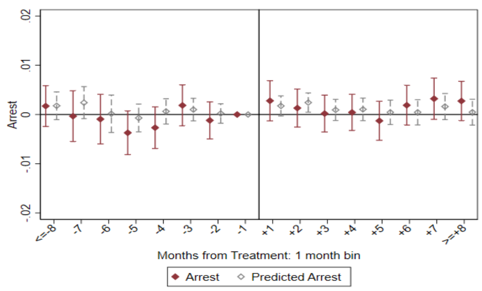
(a) Predicted Arrest



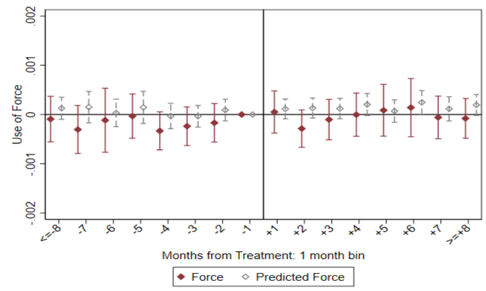
(b) Predicted Use of Force

Notes: Figures (a) and (b) show the predicted values for both outcomes over time. I use call characteristics to calculate predicted outcomes for each call. I regress arrest and use of force on call description fixed effects, day of the week fixed effects, month-by-year fixed effects, beat-by-shift fixed effects, priority of the call, and call hour to estimate the predicted values. Each point represent the average outcome over a 14-day period. Time is normalized with respect to the filing date of the complaint, such that t=0 represent the date at which a complaint is filed, and this is represented by the black vertical line. The gray dashed line represents the average time at which a finding is reached. Conditional on observing the finding date, the average finding date is 60 days after a complaint has been filed.

Figure 3: Dynamic Difference-in-differences



(a) Arrest



(b) Use of Force

Notes: Figures (a) and (b) represent the dynamic difference-in-differences estimates from Equation 3. For both estimates, I use bins that are one month long, and t = -1 is the excluded period. The treatment date, or the date at which the complaint is filed is t = 0, and is represented by the black vertical line. I report the coefficients and the 95% confidence intervals for all estimates. The maroon dots represent the effect on real outcomes, while the gray ones represent the effect on predicted values. Standard errors are clustered at the officer level for all estimations.

11 Online Appendix

Table A1: Officer Characteristics

	(1) Entire Sample	(2) Investigated	(3) Never Investigated
Investigated (at least once)	0.736	1	0
	(0.441)	(0)	(0)
Female	0.117	0.0954	0.176
	(0.321)	(0.294)	(0.382)
Year Hired	2002.9	2004.7	1997.8
	(8.252)	(7.578)	(7.948)
Year Left	2017.2	2017.3	2016.8
	(2.101)	(1.960)	(2.344)
Retired or Terminated	0.149	0.137	0.183
	(0.356)	(0.344)	(0.388)
Observations	1054	776	278

Standard deviations in parentheses

Notes: This table shows the summary statistics, including mean, standard deviation, and number of observations, for all the officers in my sample. Column (1) shows average characteristics for the entire sample, column (2) shows the average characteristics for those who were investigated at least once within the sample period, while column (3) represents the sample of officers who were never investigated within the sample period.

Table A2: Robustness Check 1 — Using quadratic and donut RD estimations

	(1) Main	(2) Quadratic	(3) Donut 1	(4) Donut 2
Panel A: Arrest				
After Complaint	0.000376	0.00125	-0.000372	0.000268
	(0.00148)	(0.00208)	(0.00150)	(0.00154)
N	227388	227388	226420	220882
Control Mean	0.0567	0.0567	0.0567	0.0564
ATE(%)	0.664	2.210	-0.656	0.475
BW	191.3	191.3	191.3	191.3
Panel B: Use of Force				
After Complaint	-0.000111	0.000111	-0.000212	-0.000171
	(0.000208)	(0.000338)	(0.000205)	(0.000215)
N	175098	175098	174130	168592
Control Mean	0.00212	0.00212	0.00212	0.00204
ATE(%)	-5.244	5.245	-9.970	-8.398
Bandwidth	135.1	135.1	135.1	135.1
Officer FE	Y	Y	Y	Y
Month× Year FE	Y	Y	Y	Y
Call Characteristics	Y	Y	Y	Y
Exclude	-	-	t=0	-7≤ t ≤ 0

Standard errors in parentheses

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

Notes: This table shows the results of three robustness checks. In column (1), I report the main estimates from Table 4. Column (2) shows the estimates assuming a quadratic relationship between the outcome variables and the running variable. In columns (3) and (4), I estimate a donut RD where I exclude t=0 and -7 ≥t ≥0, respectively. Panels (A) and (B) show the effect on arrest and use of force, respectively, and standard errors are clustered at the officer level.

Table A3: Robustness Check 2 — Using finding date as the treatment date

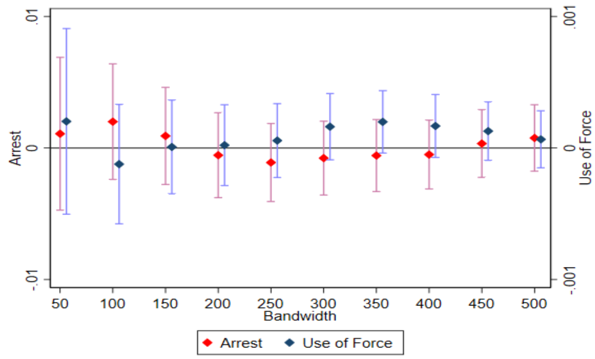
	Sustained					
	Any Complaint		Civilian Complaint		Internal Complaint	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Arrest						
After Finding	-0.00160	-0.000810	-0.0103*	-0.00751	0.000524	0.000785
	(0.00191)	(0.00162)	(0.00579)	(0.00460)	(0.00200)	(0.00172)
N	145159	145159	21089	21089	130396	130396
Control Mean	0.0530	0.0530	0.0617	0.0617	0.0517	0.0517
ATE(%)	-3.012	-1.530	-16.67	-12.18	1.014	1.519
Bandwidth	198.8	198.8	165.3	165.3	219.8	219.8
Panel B: Use of Force						
After Finding	-0.0000258	-0.0000430	-0.000364	-0.000415	0.000119	0.0000764
	(0.000167)	(0.000166)	(0.000449)	(0.000476)	(0.000173)	(0.000170)
N	152454	152454	24352	24352	110369	110369
Control Mean	0.0529	0.0529	0.0592	0.0592	0.0517	0.0517
ATE(%)	-0.0488	-0.0813	-0.615	-0.701	0.230	0.148
Bandwidth	211.1	211.1	196.3	196.3	178.9	178.9
Officer FE		Y		Y		Y
Month×FE		Y		Y		Y
Call Characteristics		Y		Y		Y

Standard errors in parentheses

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

Notes: This table shows the effect of sustained complaints on the outcomes of interest. Instead of using the filing date as the treatment date, I use the finding date as t = 0 (finding date is only available for 60% of the sample). Columns (1) and (2) show the effect of any sustained complaint, columns (3) and (4) show the effect of a sustained civilian complaint, while columns (5) and (6) show the effect of a sustained internal complaint. I classify complaints into “civilian” and “internal” using the allegation type. Civilian complaints include allegations of excessive use of force, discrimination, unlawful arrest, stop, search, or entry. Internal complaints include allegations of violating the code of conduct, gossiping, failing to file a report, conduct discrediting to the department, failure to appear in court, or violation of safety practices. Panel (A) shows the effect on arrest, while panel (B) shows the effect on use of force. Odd columns include officer fixed effects, month-by-year fixed effects, and call characteristics controls. Standard errors are clustered at the officer level and reported in parentheses.

Figure A1: Effect of Internal Oversight on Arrest and Use of Force by Bandwidth



Notes: In this graph, I report the coefficient for $1(Diff_{cit} \geq 0)$ from Equation 1 for arrest and use of force using different bandwidths. The red and blue dots represent the effect on arrest and use of force, respectively, while the lines represent the 95% confidence intervals. The x-axis represent the bandwidth used, where 50 days on each side of the cutoff is the narrowest bandwidth, and 500 days is the largest.