Good Cop, Bad Cop:
Using Civilian Allegations to Predict Police Misconduct

By KYLE ROZEMA AND MAX SCHANZENBACH[*]

In response to high-profile cases of police misconduct, reformers are calling for greater use of civilian allegations in identifying potential problem officers. This paper applies an Empirical Bayes framework to data on civilian allegations and civil rights litigation in Chicago to assess the predictive value of civilian allegations for serious future misconduct. We find a strong relationship between allegations and future civil rights litigation, especially for the very worst officers. The worst one percent of officers, as measured by civilian allegations, generate almost five times the number of payouts and over four times the total damage payouts in civil rights litigation. These findings suggest that intervention efforts could be fruitfully concentrated among a relatively small group.

I. Introduction

Police officers are tasked with a very complicated undertaking: they must make arrests and prevent crimes without violating the rights of civilians, under subjective standards such as "reasonable force" and "probable cause." Many of their decisions are imbued with a great deal of discretion, which gives rise to opportunities for abuse of power. For many reasons, including recent high-profile incidents involving police shootings and deaths in custody, preventing police officer misconduct is now at the top of the public policy agenda.

Those who supervise police officers also confront a complicated undertaking: they must prevent police officer misconduct, while still giving police officers the incentives and flexibility to fight crime. Detection of police officer misconduct by supervisors is difficult because it is costly to monitor police interactions with civilians. To facilitate detection of police misconduct, most large police departments have an administrative process through which civilians can bring allegations of police misconduct.

There is little empirical evidence on the viability of civilian allegations to identify problem officers. Police officer organizations and some criminal justice scholars have questioned the use of civilian allegations, arguing that the rate at which officers receive civilian allegations largely reflects officer productivity (Worden et al., 2012; Lersch, 2002). Moreover, state and local regulations and union contracts often create an administrative process highly favorable to police officers (e.g., requiring civilians to swear out an affidavit, tightly regulating the investigation, and subjecting discipline to lengthy appeal and arbitration processes).[1] Nonetheless, pressure to make greater use of civilian allegations is growing. For example, commentators have pointed out that Chicago police officer Jason Van Dyke, recently indicted for the shooting death of a young African-American man named Laquan MacDonald, had a long history of civilian allegations, including twenty allegations in the five years leading up to the shooting (Complaint to DOJ, 2016).

This article assesses the potential for civilian allegations to predict police officer misconduct using recently released data on over 50,000 civilian allegations

---

[1] See for example Illinois's Uniform Peace Officer's Disciplinary Act, 50 ILCS 725/1 et seq.

Electronic copy available at: https://ssrn.com/abstract=2866696

of police officer misconduct in Chicago. We use empirical Bayes estimation procedure to construct a "shrunken" measure of officer-level civilian allegations that (1) controls for officer assignment and officer characteristics, and (2) accounts for the reliability of the allegations by shrinking noisier estimates toward zero. This approach has been utilized in other settings, including the teacher value-added literature (e.g., Kane and Staiger, 2008; Chetty et al., 2014, 2015; Bacher-Hicks, et al., 2014).

The estimation strategy requires that we have adequately controlled for officer and assignment characteristics. There are two main sources of potential bias in the shrunken allegations: (1) risk created by the officer's assignment (such as high-crime, high-interaction areas), and (2) risk created by the officer's own work ethic (productivity).

We more formally assess the potential for bias in three ways. First, we test for bias in the shrunken allegations by assessing whether shrunken allegations predict the number of allegations an officer receives in other time periods and other districts. Our key approach here uses a quasi-experimental designed based on changes in officer district assignment. This test for bias was set forth in Chetty et al. (2014) in the context of teacher value-added. If the controls are adequate, the allegations an officer receives after switching districts, relative to the new district average, should be predicted by the shrunken allegation estimated outside the time period of the change. If officer allegations in other districts are poor predictors of their performance when they switch districts, then the controls do not adequately account for officer assignment. We find that allegations made against an officer by civilians after switching districts are well-predicted by the shrunken allegation estimated from other periods. This result suggests that officer shrunken allegations are not a product of their district assignment.

Second, we examine the relationship between civilian allegations and reports of officer misconduct that are made by supervisors or that arise from off-duty behavior ("non-civilian allegations"). It is doubtful that negative behavior toward supervisors, failure to show for work or respond to calls, and anti-social off-duty behaviors reflect officer productivity. We find that civilian and non-civilian allegations are highly correlated, which suggests that civilian allegations do not solely reflect productivity.

2

Finally, we assess the sensitivity of the relationship between shrunken allegations and litigation to the set of controls used in estimating the shrunken allegation. We also assess the predictive power of civilian allegations above and beyond that of non-civilian allegations. The results are stable to the set of controls used, and shrunken civilian allegations predict litigation even when controlling for non-civilian allegations.

We then test the power of shrunken civilian allegations to predict serious misconduct as measured by civil rights litigation. Because of attorney incentives stemming from contingency fee arrangements and the legal obstacles to civil rights litigation, such litigation filters for the most serious incidents of officer misconduct. We pay particular attention to payouts, which provide a measure of the seriousness of the harm inflicted or egregiousness of the officer's conduct. Notably, we observe not only awards at trial but also payment amounts resulting from settlements.[2] Thus, we test whether civilian allegations, which usually do not involve serious harm, can predict relatively less frequent but more serious civil rights litigation and associated payouts. These effects on damages further validate our allegations-based measure because large damages require serious injury or egregious conduct that is not likely to be generated by interactions alone.

We find a strong non-linear relationship between shrunken allegations and future civil rights litigation. For the officers in the bottom 80 percent of shrunken allegations, shrunken allegations are unrelated to civil rights litigation. The relationship between shrunken allegations and litigation spikes for worst 5 percent of officers and, when we consider damages, spikes further for the worst 1 percent of officers. The worst 1 percent of officers, as measured by civilian allegations, generate almost five times the number of payouts and four times the total damage payouts in civil rights litigation. This non-linear relationship suggests that intervention efforts could be fruitfully concentrated among a relatively small group of officers. Moreover, it suggests that officers with a moderate number of allegations are at no greater risk of committing serious misconduct than officers who receive no allegations.

---

[2] The collective bargaining agreements between Chicago police officers and the City of Chicago provides for indemnification of on-duty police officers, and Chicago's public spending is a public record.

In a final analysis, we present evidence on a policy debate regarding the requirement that civilians swear an affidavit as part of the investigation process. An affidavit is a sworn statement and those swearing are warned of the consequences of perjury. Roughly 55% of allegations in Chicago are dismissed for failure by the allegor to swear an affidavit. We find that allegations lacking an affidavit have the same predictive power as affidavit-based allegations. As discussed below, these findings raise concerns about the value of affidavit requirements.

The remainder of the paper is organized as follows. Part II discusses policing, the civilian allegations process in Chicago, and civil rights ligation. Our analysis then proceeds in four steps. In Section III, we develop a shrunken measure of officer type based on his or her allegations. In Section IV, we introduce the data. In Section V, we assess the adequacy of our controls by testing whether the measures perform similarly over time and across districts. In Section VI, we test whether civilian allegations predict other measures of officer misconduct, in particular allegations levelled by supervisors, allegations resulting from off-duty misconduct, and, most importantly, civil rights litigation and associated payouts. Section VII concludes.

II. Institutional Background

The interactions between supervisors and police officers, and between police officers and the community, are governed by a complicated web of laws and institutions. Between police officers and their supervisors are state laws, union contracts, and human resource (or "internal affairs") departments. Between police officers and the civilians are criminal laws, Constitutional rights, civil remedies for rights violations, and in most large cities an administrative review process, either independent of or within the police department, through which civilians may file allegations of police misconduct. While civil litigation can result in payments to civilians, the administrative process for which civilians file allegations does not. Civilians pursue the administrative route for entirely non-monetary motives. To understand the possible interactions between civilian allegation processes, civil law claims, and police incentives, we review the basic

4

incentive structure of the Chicago Police Department and the civil and administrative actions that civilians may pursue.

A. Civilian Allegations and Policing

Police officers usually act outside the presence of supervisors while engaging in multiple interactions with civilians almost every day.[3] Monitoring these interactions is, even with advances in technology, a significant challenge. In addition, state laws, city regulations, and police union contracts at times greatly restrain the power of officials to monitor and discipline police officers (Harmon, 2012). Disciplinary actions are usually subject to an appeals process that may end in binding arbitration. Investigation techniques into misconduct are also highly regulated. The Chicago Fraternal Order of Police Contract (2012) specifies, for instance, that no officer can be questioned until 24 hours have elapsed since a shooting and that an officer cannot be disciplined if recordings later contradict his or her statements to investigators.[4] The contract also provides that the details of misconduct investigations should be destroyed periodically.

In addition to such restrictions, police officer compensation is quite rigid. Compensation is often determined under collective bargaining agreements or civil service statutes that specify pay. In Chicago, salaries are entirely determined by rank and years of service. For example, the 2012 collective bargaining agreement with the Fraternal Order of Police provides three (increasing) salary grades corresponding to police officer, detective, and sergeant, with pay differentials within those grades based on years of service (Chicago Fraternal Order of Police Contract, 2012). Even if rigid pay were not required by law or contract, fixed police compensation may be the best payment structure because of well-known problems of incentive pay in multitask environments such as policing (Holmstrom and Milgrom, 1991). Much of the personnel economics literature focuses on agency problems in the employment relationship and the potential for incentive pay to improve outcomes,[5] but this is, of course, not a feasible option for police officers (e.g., pay per arrest).

---

[3] Walker and Katz (2012) estimate that there are 1,100 use-of-force incidents per day in the United States, and 43 million significant police-civilian encounters every year.

[4] Unless the tape was not available to investigators at the time.

[5] See Lazear and Oyer (2012) for a thorough review.

5

Although discipline and pay are highly regulated, police departments may create performance incentives via promotions in rank, a concept well-developed in the literature on tournaments (Lazear and Rosen, 1981; Green and Stokey, 1983; Nalebuff and Stiglitz, 1983). However, promotion in Chicago is primarily determined by exam scores (Police Department Hiring Plan for Sworn Titles, 2011). Subjective factors, such as performance evaluations and discipline history, are only considered for those who score above a pre-determined threshold. Unsustained civilian allegations (almost 98% of total) are not considered in the promotion process.

In light of Chicago's promotion policies and fixed compensation within rank, personnel tools such as termination, suspensions, reassignment, reductions in rank, or diminishment of promotion possibilities are necessary to properly align police officer incentives. Even without such incentives, police officers with a taste for misconduct will indulge in it if misconduct is not costly to them. Penalizing officers for misconduct, however, requires effective monitoring. Without cameras accurately documenting all actions of police officers and a complete review of the footage, civilian allegations will be a necessary feature of monitoring.[6]

The possibility of using civilian allegations to identify and prevent police officer misconduct is not a new idea. Almost all large police departments have standardized administrative procedures under which civilians may file allegations against police officers. Investigations of these allegations take place under a variety of procedures that vary greatly from city to city (Finn, 2001), but the process is usually slow. In Chicago the average time to process and complete an investigation of a civilian allegation is about one year, after which there may be a lengthy appeals process for any resulting discipline; in New York City the process is only slightly faster and officers are also rarely disciplined (Kane and White, 2012).

Some departments use civilian allegations as part of so-called "Early Intervention" or "Early Warning" programs (Shultz, 2015; Walker and Katz, 2012; Walker et al., 2001). A recent expert review of Chicago policing recommended the implementation of a meaningful early intervention system

---

[6] Even the use of monitoring equipment in Chicago has run into problems because of equipment failures. Roughly eighty percent of audio fails to record, which the department has acknowledged is often caused intentionally by officers disabling equipment (Chicago Tribune, 2015).

6

(Police Accountability Task Force, 2016). Some evidence suggests that such programs reduce civilian allegations (Walker et al., 2001), and a study of police officers in New York City found that officers who were terminated for cause (a rare event) tended to have more civilian allegations, but the study did not adjust for risk factors in officer assignment (Kane and White, 2009). However, some have argued that intervention programs or punishments deter effective policing (Worden et al., 2013). Other evidence suggests that the most active officers, in terms of arrests and stops, are the ones most likely to draw civilian allegations (Brandl et al., 2001; Lersch, 2002; Terrill and McCluskey, 2002). However, arrests and ticketing are not necessarily appropriate measures of officer productivity.

We address concerns over differences in risk exposure and officer productivity by (1) conditioning on district-time fixed effects and officer characteristics, and (2) linking civilian allegations to non-civilian allegations and to civil litigation. As explained in greater detail below, there are institutional arguments for why civil rights litigation is a good external metric for serious misconduct. Allegations arising from other sources, such as from supervisors, likewise provide some external validation and are unrelated to officer productivity.

B. Chicago's Civilian Allegation Process

The Independent Police Review Authority (IPRA) has initial responsibility for intake and processing of civilian allegations.[7] The agency and its employees operate outside of the police department, and the head of the agency is appointed by the mayor. Either the IPRA itself or another agency will investigate the allegation. Allegations do not only arise from on-duty interactions with civilians. Other police officers and supervisors may make allegations as well through a separate, internal process. We call these "internal allegations." In addition, off-

---

[7] Although there were some structural changes when the IPRA replaced the Office of Professional Standards in 2007, the allegation process remained largely the same across the two agencies (Safer et al., 2014). Certain allegations may be assigned to other agencies for investigation, such as the Internal Affairs Division of the police department. The Chicago Police Board adjudicates allegations against officers brought by the Superintendent of Police if discipline sought is dismissal or more than 30 days suspension. Also, the Board hears appeals from the IPRA disciplinary process if the Chief Administrator of the IPRA and the Superintendent of Police cannot agree on a disciplinary action.

duty behavior that is criminal or "unbecoming" may come to the attention of the IPRA, and we call these "off-duty" allegations.

At the conclusion of the investigation, allegations are either "sustained," "not sustained," or the officer is "exonerated." A civilian allegation that is "not sustained" indicates that there was some, but not sufficient, evidence of misconduct. This can occur for two reasons. First, the facts did not establish misconduct by clear and convincing evidence. Second, many allegations are ultimately closed because of an unwillingness on the part of a witness or alleger to cooperate with the investigation or make a sworn statement by affidavit.

In our data, only 2.4% of civilian allegations are "sustained." Even if sustained, appeals from any disciplinary recommendations may be subject to arbitration or appealed through an administrative review process. Officers are rarely disciplined as a result of civilian allegations.[8]

In order to proceed to an investigation, a civilian making an allegation must make a sworn statement by affidavit. The requirement of a sworn affidavit from the accuser has come under criticism for chilling the allegation process. As part of the affidavit process, the accuser is warned of potential criminal liability for perjury if he or she swears to something they know not to be true.[9] Recent policy proposals have suggested relaxing the affidavit requirement (Police Accountability Task Force 2016). Of course, the justification behind the affidavit requirement is that it filters out false allegations and such a requirement could improve the signal contained in the allegation process. Below we test for whether sworn or unsworn civilian allegations have different predictive power. We find no difference between them.

C. Civil Litigation against Police Officers

---

[8] If an allegation is "sustained" by the IPRA, the IPRA only recommends a punishment to the Superintendent of police. There are no clear guidelines for what discipline is appropriate, a fact highlighted in the external evaluations of the IPRA (Safer et al., 2014; Police Accountability Task Force, 2016). If the IPRA and the Superintendent cannot agree on a punishment, the Chicago Police Board resolves the dispute. If termination is recommended, the recommendation must proceed to Police Board review.

[9] Under Illinois law, investigations into allegations made by civilians against the police can proceed only if there is a sworn affidavit. The law further directs that "[a]ny complaint, having been supported by a sworn affidavit, and having been found, in total or in part, to contain knowingly false material information, shall be presented to the appropriate State's Attorney for a determination of prosecution." 50 ILCS 725/3.8.

An individual whose constitutional rights were violated by a police officer may bring a civil action against the officer under federal or state law. The most popular cause of action is the "Section 1983" claim brought under 42 U.S.C. 1983, which allows a private right of action against a state or local official who "under color of law" subjects anyone to "deprivation of any rights, privileges, or immunity secured by the Constitution and laws." Such actions encompass the use of excessive force, unlawful search and seizure, or wrongful arrest by a police officer.

There are a number of barriers to Section 1983 actions against police officers that filter out marginal or frivolous cases. First, police officers enjoy "qualified immunity," which means that in order to be found liable, police actions must have been both objectively unreasonable and violate a clearly established statute or right. "[Q]ualified immunity is an immunity from suit rather than a mere defense to liability."[10] Thus, Section 1983 cases can be dismissed on the threshold determination of whether qualified immunity applies.

Second, the burden of proof is on the plaintiff. To win an excessive force claim, for instance, the plaintiff must show that the officer objectively used unreasonable force under the circumstances, and it would have been clear to a reasonable police officer that the alleged conduct violated a statute or Constitutional right.[11] Because there is no clear legal rule that defines reasonable force, the amount of force required to be "unreasonable" is a fact-intensive inquiry. The reasonableness standard incorporates the fact that "officers are often forced to make split-second judgments" in "tense, uncertain, and rapidly evolving" situations.[12] In addition, some police misconduct, such as racist verbal abuse, does not implicate Constitutional rights,[13] and not all violations of Constitutional rights lead to damages.[14] Moreover, Chicago in particular is noted for its vigorous defense of 1983 suits, including several incidents of allegedly

---

[10] Pearson v. Callahan, 555 U.S. 223, 231 (2009).

[11] Saucier v. Katz, 533 U.S. 194 (2001).

[12] Graham v. Connor, 490 U.S 386 (1989), at 936-97.

[13] For example, in *Freeman v. Arpaio*, 125 F.3d 732, (9th Cir. 1997), a federal circuit court held that verbal abuse directed at religious and ethnic background does not give rise to a cognizable constitutional violation.

[14] For example, in *Corpus v. Bennett,* 430 F.3d 912 (8th Cir. 2005), a federal circuit court held that a strike to the head that did not result in injury only merited nominal damages of $1.

9

wrongfully withholding evidence that resulted in judicial sanctions (Weiss 2017).[15]

Finally, the financial incentives of attorneys representing plaintiffs will select for higher damage cases because of the prevalence of contingency fee arrangements. Although attorney's fees are potentially available for prevailing parties in Section 1983 actions under federal law, such fees are often difficult to collect and the definition of "prevailing" is not always straightforward. Without a significant damages award, courts have been generally reluctant to find that the plaintiff prevailed (Schwartz, 2007). As one observer has written, "plaintiffs with low damages or those seeking injunctive relief have no remedy under section 1983, because no lawyer will take their cases to court" (Reingold, 2008, p. 47). In sum, Section 1983 claims face a variety of hurdles and are brought by attorneys when there is a prospect of a significant recovery.

Because we examine the effect of allegations on litigation, it is important to consider whether civilian allegations can be mechanically related to litigation in a way that might raise identification concerns. In particular, cases in which a payout occurs are investigated by the IPRA as a matter of policy. We removed these cases from the data, but their inclusion did not materially change the results.

To further address possible simultaneity, the estimation strategy discussed below uses previous history of allegations to predict future litigation, removing any contemporaneous link between allegations and litigation. However, a question remains whether an officer's history of allegations could influence the future course of litigation. For example, identification concerns would arise if lawyers target officers with large numbers of allegations or if prior allegations are admissible as evidence. However, the number and extent of allegations is unlikely to influence the course of litigation because unsustained allegations were not generally discoverable and, if discovered, are not generally admissible in a civil trial. First, prior to the public disclosure of the allegations data in late 2015, the public did not have access to police disciplinary records. Thus, lawyers could not target police officers who had lengthy histories of allegations or disciplines when deciding to file a suit because that information would not be known at filing. This

---

[15] The number of lawsuits brought and settled against police officers in Chicago between 2009 and 2014 did not greatly vary over time. Thus, it is unlikely that policy changes regarding city settlements affect our results. See Appendix Figure A1.

boilerplate
Electronic copy available at: https://ssrn.com/abstract=2866696

implies that, in our sample period, the decision of whether to file a lawsuit was unlikely to be based on the number of allegations an officer has received.

Second, after filing a lawsuit, rules of evidence limit the discovery and admissibility of allegations. Chicago's attorneys vigorously fought against discovery of allegation histories and such discovery was often granted on only a limited basis and mainly when the allegations bore directly on the conduct at issue (St. Clair et al., 2016). Even if discovered, allegations are unlikely to be admissible. Under the Federal Rules of Evidence, evidence of prior bad acts is not generally admissible to show a defendant's propensity to commit an act, and exceptions to this rule in Section 1983 cases are made under a complicated multi-part test. That test all but requires the allegation to have been (1) based on very similar facts to the case at hand, and (2) to have been sustained, which allegations rarely are.[16] In short, institutional barriers prevent a feedback loop between allegations and litigation.

Another possible complication occurs if people simultaneously pursue allegations and civil lawsuits over the same incident. Most lawyers would not permit a client to use the allegation process because doing so requires a sworn statement that may then become admissible in court. Moreover, damages are not available in the allegation process, so claims giving rise to damages would not benefit, but could only be harmed, by the allegation process.

III. Conceptual Framework

We estimate a measure of officer type from civilian allegations using Bayes estimation, which treats individual officer parameters as random effects and adjusts them for differences in signal strength. This approach has been used in other settings, including the teacher value-added literature (e.g., McClellan and Staiger, 1999; Kane and Staiger, 2008; Bacher-Hicks, et al., 2014; Chetty et al., 2014, 2015).

---

[16] In *Okai v. Verfuth,* 275 F.3d 606, (7th Cir. 2001), an appeals court in applying the test for the exception held that *"actual evidence,* in the form of sustained complaints or potential witness testimony...."* should make admissibility of disciplinary records much more likely. Okai at 611 (emphasis in original). The Seventh Circuit Court of Appeals includes Chicago.

The Bayes estimates have two components. First, it estimates officer residuals after controlling for observable characteristics of police officers and their environment. Second, it adjusts the officer residuals to reflect its reliability by shrinking noisier estimates toward zero. Equations (1) and (2) set out the framework.

$$A_{ijt} = \beta \mathbf{X}_{ijt} + v_{ijt} \tag{1}$$

$$\text{where} \qquad v_{ijt} = \mu_j + \epsilon_{ijt} \tag{2}$$

In Equation (1), the dependent variable A_{ijt} is the number of civilian allegations generated from assignment i performed by officer j in time t, and $\mathbf{X}_{ijt}$ is the set of controls (discussed below).

The error term v_{ijt} is decomposed in Equation (2) into two components: a permanent component for officer type μ_j and an idiosyncratic component ϵ_{ijt}. Officer shrunken allegations are estimated as follows. First, Equation (1) is estimated by maximum likelihood and then each officer's mean residual is calculated $(\bar{v}_j)$.[17] Next, the mean officer residual is multiplied by an estimate of its reliability $(\hat{R}_j)$ to form the Bayes estimate of officer misconduct $\hat{\mu}_j$ as follows:

$$\hat{\mu}_j = \bar{v}_j \hat{R}_j \tag{3}$$

where

$$\hat{R}_j = \frac{\text{Var}(\mu_j)}{\text{Var}(v_{ijt})} \tag{4}$$

where $\bar{v}_j$ is the mean officer residual estimated from Equation (1) and $\hat{R}_j$ is the shrinkage factor that reflects the reliability of the mean residual $\bar{v}_j$ as an estimate of μ_j.

We explore different sets of control variables that account for the risk posed by the officer's assignment as well as officer characteristics such as age and

---

[17] The empirical Bayes estimator is implemented in Stata using the mixed command. For a description of empirical Bayes estimation using the mixed command, see Rabe-Hesketh & Skrondal (2012), p.111-112.

12

experience. Our preferred specification controls for district-time fixed effects, thus absorbing variation across and within district over time. The shrunken allegation is therefore an estimate of an officer's relative allegations over time. Moreover, we include a set of indicator variables that account for officer age, year of birth, and time on the force. This accounts for officer experience and cohort effects.

In separate specifications, we control for district and year effects but include a lagged control for the unit's average allegations per officer as well as the unit's prior variance. As shown below, the results are little changed across the sets of controls.

The shrunken allegations provide an accurate measure of officer type only if our controls adequately account for factors other than misconduct that cause differences in allegation rates between officers. Bias could arise from two sources. First, the number and nature of civilian interactions inherent in an officer's environment creates risks that our controls may not sufficiently account for. In particular, assignments within district are not observed and may carry different risk. Moreover, officers may sort across districts or within districts based on their own characteristics.

To assess this potential source of bias, below we test whether the allegations of officers after changing districts is predicted by the shrunken allegation estimated from outside the period of the change. If officers who change district have consistently different allegations in their new district, then our unit-time controls are not fully accounting for differences in officer environment or there is significant officer-district match quality.

Second, an officer's own productivity and conscientiousness may change the number and nature their interactions with civilians, creating a higher risk but one that is not attributable to actual misconduct. To assess this issue, we test for whether civilian allegations are correlated with internal and off-duty allegations, which are not based on civilian interactions. The grounds for such non-civilian allegations, as discussed in greater detail below, are not correlated with productivity. The results suggest that we adequately account for the risk of officer environment and that the shrunken allegations reflect officer conduct not solely associated with officer productivity.

13

With shrunken allegations estimated and tested for bias, we proceed to the main analysis. Relying on civilian allegations from 2002 to 2008 to estimate shrunken allegations, we evaluate the power of shrunken allegations to predict misconduct from 2009 to 2014. The use of separate time periods breaks the potential for simultaneity between shrunken allegations and our other measures of misconduct.

Officers are observed for different lengths of time between 2009 and 2014 so have different exposure to litigation in that period. We therefore construct an exposure-adjusted measure of officer misconduct ("misconduct propensity"), M_j, by multiplying the shrunken allegation by the number of periods an officer was an on the force between 2009 and 2014, E_j, according to Equation (5).

$$\widehat{M}_j = \hat{\mu}_j \times E_j \qquad (5)$$

We then estimate the relationship between misconduct propensity $\widehat{M}_j$ and involvement in civil rights litigation and associated payouts according to Equation (6):

$$Y_j = \alpha + \beta \widehat{M}_j + \zeta_j \qquad (6)$$

where Y_j is an outcome of interest described below for officer j over time period 2009 to 2014. The reported results will use a flexible form of M_j to account for the possibility that the relationship is non-monotonic. In particular, we will report the results by dividing up shrunken allegations into deciles and breaking out officers in the top 5 percent and 1 percent of the distribution.

It is worth emphasizing that the exposure measure E_j is calculated based on the same time period as the litigation outcomes Y_j, whereas the intensity measure of misconduct $\hat{\mu}_j$ is based on past time periods. The coefficient of interest, β, estimates the relationship between the exposure-adjusted shrunken allegation and the outcomes.

IV. Data and Descriptive Statistics

14

To conduct our analysis, we link personnel, allegation, and litigation data from four sources: (A) a panel of personnel data on the district or unit to which police officers were assigned on any given day while they were on the force; (B) a panel of roughly 50,000 civilian allegations of police officer misconduct, as well as 28,000 internal allegations and 5,000 off-duty allegations; (C) a panel of the universe of federal and state lawsuits in which Chicago police officers are named; (D) a panel of the universe of lawsuit payments made on behalf of these officers by the City of Chicago (officers are indemnified by the City).

A. Personnel Data

We obtained personnel records on all Chicago police officers by filing a Freedom of Information Act (FOIA) request with City of Chicago Department of Human Resources. The bi-annual personnel data include each officer's name, position title (rank), and original hire date. These records are available from 2002 to 2014. We use the officers first and last name and the original hire date to link records with the other datasets. We set the unit of analysis in estimating the shrunken allegations at the same six-month window (January to June; July to December) to match the structure of the personnel files. We limit our analysis to officers at four levels of rank: officer, officer in special unit, detective, and sergeant. This is because higher-level officers, such as captains and other supervisors, are often named in lawsuits as a matter of course, which means there is little information content in a lawsuit against a supervisor. On average from 2002 to 2014, there were roughly 12,000 police officers, detectives, and sergeants in Chicago each year.

We first match the personnel records to data on the district to which each police officer was assigned at the start of each period. There are 25 police districts in Chicago defined geographically, as well as some city-wide units such as the narcotics and gang task forces. Because the city-wide units have a relatively small number of officers and we use district-time fixed effects, we group officers in the city-wide units together.[18]

---

[18] The results are consistent when we restrict the sample to officers in the geographically defined police units, as well as when we group only the smallest city-wide units together and treat larger units as separate.

B. Civilian Allegations Data

Chicago Police Department records of officer misconduct allegations were not available to the public until December 2015. In 2014, a court decision *Kalven v. Chicago*[19] opened to the public a partial register of allegations made against police officers (from 2001 to 2006 and from 2011 to 2015), which became available to the public in December 2015. A second round of data was released in October 2016 based on a freedom of information request from the Chicago Tribune.[20] We use allegations made against Chicago police officers from 2002 to 2014 from the second round of the data release, which we obtained from the Citizens Police Data Project at the Invisible Institute, an organization that serves as a clearing house for the data.[21] Allegation records contain the identity of the officer in question, the date in which the incident for the civilian allegation occurred, and the type of allegation (e.g., verbal abuse, wrongful arrest, or excessive force). We exclude the over 3,000 allegations that arose because of civil litigation, which is indicated as a type of allegation.

The data include the type of conduct that gave rise to the allegation, ranging from allegations about off-duty behavior to the use of excessive force. Table 1 provides a summary of the type and number of allegations as coded by the city. We identify "civilian allegations" as those that relate to interactions between civilians and on-duty police officers, such as "First Amendment/Wrongful Arrest" and "Excessive Force." We separate the remaining non-civilian allegations into "internal allegations," which are allegations that arise from reports from supervisors or fellow officers, and "off-duty allegations," most of which are generated by off-duty behavior. The most common internal allegations are "neglect of duty," "failure to make a report," and "failure to provide service."[22] The remaining offenses include insubordination and failure of a drug or alcohol test. The "off-duty" allegations include behavior that is often criminal in nature, such as drunk driving and domestic violence, but also include repeated traffic violations.

---

[19] 7 N.E.3d 741 (2014).
[20] For a description, see Richards et al. (2016).
[21] http://invisible.institute/police-data/
[22] "Failure to provide service" is also an allegation made by civilians, but we classify it as an internal complaint because it signifies a failure of performance of duties.

16

Using geospatial data on Chicago's police districts available from the City of Chicago, Figure 1 plots the mean per-officer average number of annual civilian allegations by police district. With the city-wide annual average rate of 0.36 civilian allegations per officer over the sample period, annual rates vary substantially across districts, from as low as in 0.17 allegations per-officer in District 19 to as high as 0.62 in District 11. This substantial variation motivates our risk-adjusted approach to estimating the shrunken civilian allegations. Crime rates vary across police districts and within police districts over time, and civilian-police interactions, and the frequency and intensity of those interactions, vary over time, across geographic areas, and even within a geographic area over time. Officers in each district-time period may face different inherent risks of having an allegation made against them.

Figure 2 reports the distributions of the shrunken civilian allegations from Equation (3) (Panel A) and the distribution of the misconduct propensity from Equation (5) (Panel B). Both distributions are approximately log-normal, and are very similar in shape. The first panel of Table 2 provides descriptive statistics of civilian allegations at the officer-six-month period level from 2002 to 2008. The mean number of civilian allegations is 0.16 per period. The variance of predicted civilian allegations for each officer is 21% of the total variance of allegations across officers. The second panel of Table 2 provides descriptive statistics of allegations from 2008 to 2014, which will be the outcome variables. The mean number of civilian allegations is 0.11 per period. Internal allegations are about half as frequent as civilian allegations, and off-duty allegations are about one-tenth as frequent as civilian allegations.

C. Data on Civil Litigation Filings

Using Bloomberg Law, we obtained the universe of federal and state civil actions from 2009 to 2014 where Chicago or a city department was listed as a party. For each civil action, the Bloomberg data contains the filing date of the lawsuit and the names of each defendant in the suit, which we use in combination with officer names to link up to the personnel data. Bloomberg also codes the data by cause of action. We identified 5,809 instances between 2009 and 2014 in which a Chicago police officer was listed as a party to a federal Section 1983

action or a state action based on a deprivation of Constitutional rights.[23] From 2009 to 2014, officers are named in an average of 0.23 lawsuits. Note that 17 percent of officers are named as defendants in civil rights lawsuits, but some officers are named in multiple lawsuits.

D. Data on Civil Litigation Payments

We obtained the universe of actual payment amounts to civil rights plaintiffs made by the City of Chicago from 2009 to 2014.[24] The collective bargaining agreements between police officers and the City of Chicago provide for indemnification of on-duty police officers in civil suits, and all payments made by the city in litigation matters are public.[25]

The payment data does not list the officers that were named in the litigation or the filing date of the lawsuit. However, the data contains the civil action number, which we link to the civil litigation data from Bloomberg, thus identifying the officer defendants in the litigation.[26] Between 2009 and 2014, we identified 2,099 occasions that an officer was part of a lawsuit that resulted in a payment. The average officer is named in 0.14 lawsuits for which a payment was

---

[23] To match officers from the personnel data to officers named in litigation in the Bloomberg data, we followed four-step process. First, we exclude officers from the personnel data who do not have a unique name. Second, we used a fuzzy match record linkage process to link all officers in the personnel data to all named defendants in the Bloomberg data. This created matches with a probabilistic assessment of the similarity of the character strings of the officer names in both datasets. Third, we required a probabilistic match rate to be at least 0.9. This resulted in 19,770 potential officer matches. In a final step, a research assistant manually assessed the officer names and scored the potential name matched on a standard 4 step scale (not a match, maybe a match, very likely a match, and definitely a match). We retained matches that the research assistant identified to be "very likely a match" and "definitely a match". Using a lower threshold introduces measurement error and thus change the size of the point estimates, but the sign and significance hold under the different definition we use for the match quality.

[24] Figure A1 in the Appendix shows officer involvement in lawsuits over time. The number of officers named in a lawsuit, named in a lawsuit that resulted in a payment, and named in a lawsuit that resulted in a payment above $100,000 is stable over time.

[25] Chicago discloses the data at http://www.cityofchicago.org/city/en/depts/dol.html.

[26] We created crosswalks for both federal and state cases. Federal and state cases can be distinguished by the civil action number. For instance, federal cases follow the format of XX-CV-XXXX, and state cases formats include XX-M1-XXXXX and XX-L-XXXX. To construct the federal crosswalk, we first acquired all the federal actions coded as Section 1983 claims or prisoner claims in the Northern District of Illinois from Bloomberg Law, and all state court dockets for cases in the Illinois Circuit Court for Cook County in which the City of Chicago or Chicago Police were named as a party. Section 1983 and prisoner claims receive numeric codes of 440 and 550 from the Federal Judicial Center.

18

made. Note that 11 percent of officers are involved in a case for which a payment was made, but some officers are involved in multiple lawsuits with payments. The average payout per officer from 2009 to 2014 was $40,519.

Chicago's total payments in police officer misconduct cases have averaged nearly $50 million annually from 2009 to 2014. Figure 3 shows the distribution of payments that were matched to the personnel records, where the x-axis is in natural log scale because the long right tail of payment amounts. Conditional on a payment, the median payment is $40,000 and the average is $329,322. Settlements are clustered below $100,000, likely due to the requirement that the City Council approve settlements in excess of $100,000 (Iris, 2014). The average officer is named in 0.03 lawsuits with a payout of at least $100,000. Because of this unique procedural feature of large settlements, we study them as an additional outcome.

V. Testing for Bias in the Shrunken Allegations

In this section, we test for the adequacy of the controls by assessing how well shrunken civilian allegations predict civilian allegations outside the time period in which the shrunken allegations were estimated. First, we assess whether shrunken allegations predict future civilian allegations. Second, we assess whether shrunken allegations predict civilian allegations after officers switch districts. If the controls adequately account for assignment risk, these relationships should be one-to-one.

A. Predicting Future Civilian Allegations from Shrunken Allegations

To assess whether civilian allegations predict future allegations, we regress officer period average civilian allegations from July 2008 to 2014 on officer period average shrunken allegations estimated from 2002 to June 2008. If the model is correctly specified, the relationship should be one-to-one.

However, two factors prevent the relationship from being exactly one to one. First, the allegation rate is bounded by zero, and many officers receive zero or only a few allegations. Because we anticipate a flatter relationship for officers

19

on the left side of the distribution, we allow a kink point at the median in the display of the results. Second, there was a secular decline in allegations beginning in about 2011. Average officer allegations from 2002 to 2008 are almost identical to the average allegations from July 2008 to 2010, but fall thereafter by about half.

Figure 4 reports two binned scatterplots of the relationship between shrunken allegations and allegations from July 2008 to 2010 (Panel A) and then in the entire post-period (Panel B). Each bin represents five percent of the sample, and we additionally break out the top one percent to assess whether the relationship continues to hold in the extreme right tail. In both panels, there is a strong and precisely estimated relationship between shrunken allegations and allegations in the second period. The relationship is approximately linear for the top half of shrunken allegation distribution but is flat for the bottom half. The fit line is much steeper in the Panel A than Panel B, reflecting the decline in the allegation rate beginning in 2011.

In Panel A of Figure 4, the coefficient of shrunken allegations on the actual allegation rate from 2008 through 2010 is 0.90. For officers above the median in shrunken allegations, the coefficient on shrunken allegations is 1.02 with a standard error of 0.03. In either case, the relationship is close to the one-to-one relationship, which is consistent with correct specification.

Panel B of Figure 4 includes the whole post-period on the y-axis. Here, the coefficient of shrunken allegations on the actual allegation rate is now 0.53 for the whole distribution and 0.59 for officers above the median in shrunken allegations. As in Panel A, the relationship is precisely estimated, but the coefficients are now well below one. However, the weaker relationship is almost fully explained by the secular decline in allegations in the second period. Officers in the top 50 percent of the shrunken allegation distribution had an average of 0.27 allegations from 2002 to 2008 but an average of 0.18 allegations from 2008 to 2014. Thus, given the one-third decline in the allegation rate, the coefficient estimate of 0.59 is consistent with correct specification.

20

In summary, the relationship between shrunken allegations and future allegations is close to one-to-one for officers in the top half of the shrunken allegations distribution.[27] We also note that the flat relationship for officers with a shrunken allegation below the median is similar to the non-linearities evident in the litigation results below. We discuss the implications of this flat relationship in greater detail in the results section.

B. Test for Bias Using Variation from Officers Switching Districts

We now assess whether shrunken allegations predict civilian allegations after officers switch districts. The motivation for this quasi-experimental test is to assess whether shrunken allegations estimated from outside a window of officer entry into or departure from a district predict actual allegations within the event window. We note here, however, that additional controls (officer age, experience, rank) affect the relationships between shrunken allegations and outcomes very little (see below).

If the model is correctly specified, officer shrunken allegations from outside the event should have a 1:1 correspondence with allegations within the event window. By contrast, if officer allegations in other districts are poor predictors of allegations when officers switch districts, then the model is misspecified either because selection is not adequately controlled for or because there is a significant officer-district match effect. Similar quasi-experimental analyses have been used in the context of teacher value-added (Chetty et al., 2014).

In addition, we report an event study focused on outlier officers. We do so for two reasons. First, the event study on outliers tests for the adequacy of our controls by assessing whether high or low-allegation officers have the effect

---

[27] A final concern is that officers with allegations could exit the force over time. To test for allegations-biased attrition, we estimate the relationship between) shrunken allegations from 2002 to 2008 and (1) whether an officer is on the police force at any time between 2009 and 2014, and (2) whether an officer on the police force in 2009 exits by 2014. We find no evidence of that shrunken allegations are related to attrition (see Appendix Table A1 for results).

predicted when they enter or leave a district.[28] Second, in our litigation results we find a non-linear relationship between shrunken civilian allegations and litigation that is concentrated in the right tail of the distribution. Accordingly, the right tail of the distribution is of special interest.

B.1 Entry of Officers

We define entering officers as those who enter a new district from a previous district and then remain in the new district for at least three periods. We define non-switching officers as officers who were in the district before the switching officers entered and remained in the district for at least three periods.[29] We estimate a "leave out" shrunken civilian allegation for each of the switching and non-switching officers based only on the allegations that occur outside the event window.[30] In total, there are 780 events.

We then construct (1) the difference in the mean allegations for the switching officers and the non-switching officers in the event window, and (2) the difference between the mean shrunken allegation for the switching officers and the non-switching officers outside the event window (the leave-out measure). A large difference in (1) indicates that officers who switched into a district received more allegations on average than non-switching officers in the district in the three periods after the switch. A large difference in (2) indicates that officers who switched into a district received more allegations outside of the event time window than non-switching officers in the district (controlling for risk factors and adjusting for noise). Panel A of Figure 5 presents a binned-scatterplot with (1) on the y-axis and (2) on the x-axis. Panel A of Table 3 reports regression results. The coefficient is 0.98 without controls and 1.01 with district and time fixed effects, and the standard error in both columns is 0.17. These results are consistent with the adequacy of our controls.

---

[28] Here we treat switching events as going to or from either geographic districts or city-wide units.
[29] We do not count newly hired officers as switching, and we also require officers to have been in a previous district for at least two periods.
[30] We estimate the leave-out shrunken allegation separately for each event, using all others that are not in the event in question as controls. To generate the leave-out shrunken allegations, we estimate the shrunken allegation after setting the allegations for all officers in a district-time to missing and the 2 periods before and after the district-time in question.

22

Next, we implement another version of the switchers' tests that includes a first-difference within a district during an expanded event window. The approach here differs because the event window has both a pre-event period and a post-event period and draws on variation within a district over time. The pre-event period is t=-1, t=-2. The post-event period is t=0, t=1, and t=2. We use the same definition of switching and non-switching officers. The leave out shrunken allegations are estimated as described above. We calculate (1) the difference in the mean shrunken allegation of officers in a district-time event from before the event to after the event. The mean shrunken allegations of officers before the entry event excludes the switching officers and the mean shrunken allegations of officers after the event includes the switching officers. We next calculate (2) the difference in the mean allegations of officers in a district-time event in the two periods before the event (in the event window) and the three periods after the event (in the event window). The mean allegations of officers before the entry event excludes the switching officers and the mean allegations of officers after the event includes the switching officers. Thus, this test estimates the within district change in expected and actual allegations around the entry event. Panel A of Figure A4 in the Appendix presents a binned-scatterplot. The regression results are reported in Columns 3 and 4 of Panel A of Table 3. We estimate coefficients of 1.17 and 1.24 with standard errors of 0.25 and 0.37, respectively. The results are statistically similar to the previous estimates, but the standard errors more than double. However, the larger standard errors are expected because the differencing within district increases measurement error.

B.2 Departures of Officers

We define a departure event as one or more officers departing a district in a given time period. We define non-switching officers and estimate shrunken allegations on a leave out basis as before. In total, there are 1461 departure events. We again take two approaches. First, we estimate the relationship between officers' mean per-period allegations in the event window and the officer's shrunken allegation from outside the event window. Panel B of Figure 5 presents

23

a binned-scatterplot of the results. Columns 1 and 2 of Panel B of Table 3 reports regression results. We find that the departure tests result in precisely estimated coefficients of 1.05 without controls and 1.06 with controls.

Second, we expand the event window to include both a pre-event period and a post-event period and draw on variation within a district over time.[31] We then calculate (1) the difference in the mean shrunken allegation of officers in a district-time event from before the event to after the event. The mean shrunken allegations of officers before the departure event includes the switching officers and the mean shrunken allegations of officers after the event excludes the switching officers. We next calculate (2) the difference in the mean allegations of officers in a district-time event in the two periods before the event (in the event window) and the three periods after the event (in the event window). The mean allegations of officers before the departure event includes the switching officers and the mean allegations of officers after the departure event excludes the switching officers.

Columns 3 and 4 of Panel B of Table 3 reports regression results. Panel B of Figure A4 in the Appendix presents a binned-scatterplot of the results. The first difference estimates increase to 3.21 and 2.78 and the standard errors increase to 0.88 and 1.21. In this case, the results are so imprecise that we cannot draw any inferences. Again, we ascribe the loss of precision to the increase in measurement error resulting from the differencing within district in the second approach. In either case, however, there is no evidence of misspecification.

B.3 *Officers in the Tail of the Distribution*

Figure 6 reports the results of an event study analysis. For each entry event, we calculate the mean shrunken allegation for the cohort of non-switching officers. For each cohort of officers entering a district, we then calculate the difference in the mean shrunken allegation of the entering cohort and the non-switching cohort. We then define an entering high (low) misconduct officer

---

[31] The number of events decreases from 1461 in the levels departure event to 775 in this analysis because of the additional event period and the need to observe shrunken allegations in the pre-event period for departing officers.

cohort as one who is in the top (bottom) 5 percent of this differential. The figures plot the mean civilian allegations for the cohort of non-switching officers (solid line) and the cohort of non-switching officers plus entering or departing high-misconduct officers (dashed line).

Panel A plots the event study for the entry of high misconduct officers, and Panel B plots the event study for departures of high misconduct officers. In both cases, average district allegations respond as predicted, by increasing (decreasing) when high misconduct officers enter (leave). In both panels, the difference between the cohort of non-movers and the entire group (entrants plus non-movers) is statistically significant at the one percent level.

Panels C and D plot the event study for the entry and departure of low misconduct officers, and provide evidence that low misconduct officers only slightly influence the mean number of allegations in districts they enter or leave. This is not surprising because there is a zero lower bound for good behavior, and there are not large differences between officers in the middle and lower end of the distribution of allegations (see Figure 2).

One concern with the event study approach in Figure 6 is that officers switching districts could be different than officers not switching districts. If so, then we could just be picking up a systematic difference in officers who switch districts. To overcome this concern, we study all switching officer at the individual level. For this analysis, we retain all switching events, and include an indicator for any switching officer that takes a value of 1 for each period t, t+1, and t+2 for each entering officer. We then interact this switching indictor with indictors for the high and low misconduct officers. We then estimate the following regression.

$$A_{jt} = \alpha + \beta S_j + \gamma S_j \times H_j + \theta S_j \times L_j + \sigma_{dt} + \zeta_{jt} \qquad (8)$$

where A_{jt} is the number of civilian allegations for officer i in district d in time t. S_j is the indicator for whether officer j is a switching officer. H_j and L_j are indicators for whether the switching officer is a high or low misconduct officer

25

relative to officers in district they are switching to or from (top and bottom 5 percent). σ_{dt} are district-time fixed effects.

The results are reported in Table 4. Panel A reports the results for officer entrances. The main effect on the indicator for entering officers indicates that officers who enter a district have slightly more allegations than non-switching officers. In Column 3, which uses district-time fixed effects, entering officers have 0.025 more allegations relative to non-switching officers. This is effect is small relative to the sample average of 0.18 allegations. By contrast, entering high misconduct officers have an additional 0.803 allegations after entry, or roughly four times the average number of allegations of all the non-switching and entering officers. We find evidence that low misconduct officers indeed receive fewer allegations (on average 0.210 fewer, which is significant at the one-percent level).

Panel B reports the results of the departure events. Consistent with the finding that entering officers have more allegations than the non-switching officers, we find that departing officers have more allegations in the two periods before departing than other officers in that district and time period. High misconduct departing officers have on mean 0.767 more allegations in the two periods before departing than other officers in the district and time. We also find evidence that low misconduct departing officers indeed receive fewer allegations in the periods before departing than other officers in the district and time period, but the effect is again roughly one-fifth the size of the bad officer effect.

B.4 Conclusion

In all of the above tests, we find no evidence of bias from district-level or within-district assignment. Shrunken allegations in the first period predict actual allegations in the second period, particularly for the top-half of the shrunken allegation distribution. In the switching tests, allegations change in manner that is well-predicted by differences in the shrunken allegations between the cohorts of switching officers and those officers who did not switch. When the worst officers enter or depart a district, allegations change in the direction expected. In short,

officers who are bad in one district are also bad in another district, and officers who were bad in the past are also bad in the future.

VI. Results

This section explores the power of shrunken civilian allegations to predict future serious misconduct. The analysis proceeds in three parts. In Section V.A, we examine the power of shrunken civilian allegations to predict internal and off-duty allegations. Internal allegations usually reflect a failure to perform essential police functions (see Table 1 and discussion). Thus, we are interested in internal allegations in part to establish that civilian allegations are not solely driven by productivity. By contrast, most off-duty allegations result from criminal misconduct or "conduct unbecoming" such as socializing with a known felon, and thus reflect misconduct that is criminal or unethical but unrelated to on-the-job behavior. In Section V.B, we examine the power of shrunken civilian allegations to predict involvement in civil rights litigation and associated payouts. As discussed above, involvement in litigation and associated payouts are likely to reflect serious underlying misconduct. In Section V.C, we extend the analysis by assessing whether the predictive power of allegations supported by an affidavit is different than that of allegations unsupported by an affidavit.

A. Civilian Allegations and Non-Civilian Allegations

We assess the ability of shrunken civilian allegations to predict future non-civilian allegations in the split sample. Figure 7 reports two binned scatterplots in which shrunken civilian allegations from 2002 to June 2008 are on the x-axis. On the y-axis, Panels A and B report internal and off-duty allegations from July 2008 to 2014. The y-axis is reported relative to the average internal and off-duty allegations (e.g., the y-axis at one is at the average, and the y-axis at two is at twice the average). In both panels, we find a precisely estimated positive correlation between first period shrunken civilian allegations and second period non-civilian allegations. The relationship is approximately linear. Compared to the average officer, the very worst one-percent of officers in shrunken civilian

27

allegations receive roughly two times and four times as many internal and off-duty allegations.

B. Civilian Allegations and Civil Rights Litigation

Next, we explore whether shrunken civilian allegations predict serious police officer misconduct as measured by civil rights litigation. We use four outcome measures of an officer's litigation outcomes in the second period: (1) the number of civil rights lawsuits in which an officer was named, (2) the number of civil rights lawsuits that involved a payout (either through an award or a settlement), (3) the number of civil rights lawsuits that involved a payment that the City Council had to approve (at least $100,000), and (4) the total amount of payments in cases that involved the officer.

Figure 8 displays the non-parametric results of Equation 6, grouping officers according to their misconduct propensity into deciles. We also break the top decile of misconduct propensity into three categories and separately report the 90th to 95th percentile, 95th to 99th, and the top 1% of officers. The x-axis indicates the mean misconduct propensity for a given decile, and the y-axis is the point estimate from Equation (6)'s regression of the litigation outcomes on group indicators. Officers in the lowest decile are the reference group and each panel reports the sample mean for comparison purposes.

Panels A and B of Figure 8 provide the results for the number of lawsuits an officer was named in and the number that resulted in a payment.[32] The estimates over the misconduct propensity distribution in Panels A and B track each other quite closely. The number of lawsuits and lawsuits for which a payment is made are flat across the first seven misconduct propensity deciles. There is a slight increase in lawsuits at the eighth decile and a larger, statistically significant increase at the ninth decile. Officers in the ninth decile receive about 0.1 more lawsuits or lawsuits with payments than those in the first seven deciles. The change in the top decile is more dramatic. Indeed, the worst five percent of

---

[32] We also estimated a linear probability model with an indicator variable for "any lawsuit" and "any lawsuit with a payment" and found very similar results.

28

officers receive roughly 0.4 more lawsuits and lawsuits with payouts than those in the bottom seven deciles. There is no statistical difference between the worst one percent of officers and those in the 95th through 99th percentiles.

Panel C of Figure 8 reports the results for the number of lawsuits that result in a payout of $100,000 or more, necessitating the approval of the settlement by the city council. Involvement in such a lawsuit is quite rare, with only 3 percent of officers being named in such a suit from 2009 to 2014. Because payments of over $100,000 involve additional review by city officials, they are indicators of particularly serious misconduct. The first nine deciles are not meaningfully different from each other. However, large differences emerge after the 90th percentile. The worst one percent of officers average roughly 0.1 more lawsuit payments that were approved by the city council.

Panels D through F of Figure 8 report results for total payments using OLS in levels, OLS in natural logs, and a quantile regression at the 95th percentile. We use different approaches because we are not only interested in the average payout, but also the distribution of damages and the power of allegations to predict large payouts. Quantile regressions are particularly helpful because there is also a long right tail in the damages distribution (see Figure 2). We choose to report quantile regressions using the 95th percentile because only eleven percent of officers have a lawsuit with a payout from 2009 to 2014, so the 95th percentile represents roughly the midpoint of positive damages.

Panel D reports the results for payouts in levels estimated by OLS. Here, the estimates are fairly noisy, likely because of the high variance in payouts levels by officer. Nonetheless, the same pattern from Panels A, B, and C is again evident. The point estimates on misconduct propensity are roughly flat for the first eight deciles but increase for the very worst officers. Moving from the 95th through 99th percentile to the top percentile approximately doubles the point estimate of misconduct propensity, from $63,209 to $139,000. Moreover, the worst 10 percent have statistically more damages compared to the bottom 90 percent of officers ($57,247 more; $p<0.01$) (not shown). When damages are

29

estimated in natural log form in Panel E,[33] the estimates are more precise, and the same basic pattern emerges, except that now there is a statistical difference in the eighth decile and the magnitude of the point estimate on the worst one percent, while larger than that of the 95th to 99th percentile, is smaller compared to the results in levels in Panel C.

The quantile estimates in Panel F show dramatic differences in damages across the distribution of misconduct propensity. The sample 95th percentile in damages is $55,000, and the conditional 95th percentiles for the bottom eight deciles are little different from each other. By contrast, there is a large spike in the conditional 95th percentile of payouts for those at the very top of misconduct propensity. The 95th percentile of damages is $370,000 for officers in the 95th to 99th percentile of misconduct propensity and is $1,165,582 for the very worst 1% of officers. For comparison purposes, Figure A2 in the Appendix reports quantile results estimated at the 91st, 93rd, and 97th percentile.

The use of different controls to estimate shrunken allegations makes little difference to the estimated effect of the misconduct propensity in predicting future litigation outcomes. Figure A3 in the Appendix reports results when the shrunken allegation is estimated (1) controlling for district and time fixed effects and lagged district mean allegations and lagged district variance in allegations (the solid line with circle markers), and (2) in our preferred specification but with the addition of shrunken internal and off-duty allegations (the dash line with square markers). When internal and off-duty allegations are controlled for, there is only a small attenuation in the point estimates. This suggests that civilian allegations have significant value in predicting serious misconduct independent of the information contained in internal and off-duty allegations.

Table 5 reports regression results analogous to Figure 7. Panel A of Table 5 reports a linear specification, Panel B reports the linear specification after controlling for the misconduct propensity of both internal allegations and off-duty allegations from the period 2002-2008, and Panel C reports a linear specification separately for misconduct propensity less than zero and misconduct propensity

---

[33] We take the ln(payout + 1) to account for zero payouts.

30

greater than or equal to zero. The regression results report a strong and precisely-estimated relationship between misconduct propensity and each of the various litigation measures. In addition, a stronger relationship is clearly observed for officers in the top half of the distribution of allegations in Panel C.

C. Affidavit Requirement

As discussed in Section II(B), Chicago requires civilians making allegations to swear out an affidavit concerning the facts alleged. Almost 55% of investigations are dropped for failure of the complainant to swear out an affidavit, but thus far we have treated allegations lacking an affidavit the same as allegations with an affidavit. Recent reform proposals in Chicago have called for modifying the affidavit requirement by allowing investigators to pursue the investigation if they have reasonable grounds for doing so, regardless of whether an affidavit is sworn (Police Accountability Task Force, 2016). Some argue that the affidavit requirement, with its warning of penalties for perjury, deters meritorious allegations from being pursued. After all, the allegation process is only an administrative procedure and cannot yield a payment for the alleger. On the other hand, the affidavit requirement may filter out baseless allegations, improving the signal contained in allegations and shielding officers from administrative harassment.

Here we shed some light on the extent to which the affidavit requirement deters meritorious allegations by testing whether the predictive power of allegations with and without affidavits is different. We first test whether allegations dismissed for lack of an affidavit have systematically different predictive power in civil litigation than affidavit-based allegations. We next assess the sensitivity of the litigation results to the weighting of allegations with and without affidavits. The intuition behind these tests is as follows. On the one hand, if requiring an affidavit largely prevents meritless allegations from being filed, then allegations that are not supported by an affidavit should be less predictive of litigation. On the other hand, if requiring an affidavit mostly deters meritorious allegations from being pursued, then allegations that are not

31

supported by an affidavit would be as predictive of litigation as allegations with an affidavit.

We have to adjust the sample time frame to perform this analysis. Prior to 2007, the data do not record "lack of an affidavit" as a reason for dismissal of a civilian allegation. Thus, we split the remaining time frame equally, estimating the shrunken allegations based on data from 2007 to the first half of 2010 and using data on litigation from the second half of 2010 through 2014.[34] We then separately estimate shrunken allegations for allegations with and without affidavits. Figure A5 demonstrates that the distribution of shrunken affidavit and non-affidavit allegations are similar. Figure A6 in the Appendix demonstrates that shrunken civilian allegations with and without affidavits predict future civilian allegations (analogous to Figure 4).

We next re-estimate Equation 6 for the litigation outcomes. Figure A7 in the Appendix reports the non-parametric results. The non-parametric results in Figure A7 suggest there is no meaningful different between allegations with an affidavit (solid line) and allegations without an affidavit (dashed line) for most of the outcomes. Because of the decrease in sample size in both civilian allegations and litigation outcomes, and because the number of civilian allegations is split roughly in half between those with and without an affidavit, the precision of the estimates in Figure A7 is decreased compared to those in Figure 9 (the OLS results in levels are particularly noisy). Qualitatively, the results in Figure A7 parallel the results in Figure 9. There is a flat relationship between shrunken allegations (with or without affidavits) and civil litigation for the first eight deciles but a pronounced increase in the top decile. More importantly, the results are similar whether we rely on allegations with or without an affidavit. If anything, allegations without an affidavit are even more correlated with future litigation, but the difference is not statistically significant. Appendix Table A2 is analogous to Table 5 and provides further confirmation that there are no

---

[34] We obtain similar but noisier results when we use data on civilian allegations from 2007 to 2009 and data on litigation from 2009 to 2015.

statistically significant differences between the predictive power of affidavit and non-affidavit based allegations.

In a final step, we assess the sensitivity of the main results in Figure 9 to the weighting of allegations with and without affidavits. We further describe the methods and present the results in the Appendix. The results suggest that civil litigation is best explained by a shrunken allegation estimated from a roughly equal weighting of allegations with and without affidavits. In other words, allegations not supported by an affidavit predict future misconduct in a similar manner as affidavits supported by an affidavit (see Appendix Figures A8 and A9 and accompanying text). Our findings are consistent with the view that meritorious allegations are deterred by the affidavit requirement.

VII. Discussion and Conclusion

In this article, we have tested the power of civilian allegations of police misconduct to predict future serious officer misconduct as measured by civil rights litigation. We found that empirical Bayes measures of civilian allegations predict officer involvement in civil rights litigation and associated payouts. This effect was most evident for the worst five percent of officers, and even more so for the worst one percent. By our estimates, removing the worst one percent of officers (120 in total) from regular civilian contact—either by reassignment or termination—and replacing them with an average officer would have saved Chicago over $6 million in payouts over the years 2009 to 2014 (not including legal fees). Of course, not all serious misconduct is reflected in damage awards, and the underlying misconduct this litigation represents has implications for police-community relations and police effectiveness.

Our results have several important policy implications. First, the non-linear relationship between civilian allegations and civil rights litigation implies that intervention efforts could be fruitfully concentrated among a relatively small group of officers. The evidence suggested that only officers with abnormally high allegations should be scrutinized. Officers who have risk-adjusted allegations below the 80 to 90th percentiles are little different than officers who receive no allegations. An approach targeted at the very worst officers, who are easily distinguished from their peers, would decrease the likelihood of over-deterrence.

33

Second, the results show that the mere presence of excessive civilian allegations, whether or not investigated or sustained, can be used to identify officers who are at high risk of perpetrating serious misconduct and creating significant liability for the city. Our results do not address how intervention should proceed, but it could take various forms, from dismissal to early intervention programs that try to counsel or reassign potential problem officers before a serious incident occurs. For example, Officer Jason Van Dyke, indicted for murder in the shooting death of Laquan MacDonald and for whom Chicago has paid $5 million in damages, was in the worst 3 percent of officers in the shrunken allegation distribution before the shooting. Based on civilian allegations prior to the shooting, our results identified Officer Van Dyke as a problem officer.

A related avenue of future research would be to assess whether officer characteristics observable to police departments at the hiring stage, such as education or prior experience, are associated with later civilian allegations. Officer characteristics observable to police departments at the hiring stage (e.g., education) might not be able to predict the rare events like large payments made in civil rights litigation, but they might predict the more frequent events of civilian allegations. Such information could inform hiring standards. Unfortunately, we lacked the human resource data to perform that analysis in this study.

Finally, the results suggest that the investigation of civilian allegations should be taken more seriously. At present, many cities devote few resources to investigating civilian allegations and make little use of the results of an investigation. A more serious investigatory process would have the benefit of screening out frivolous allegations, improving the signal quality contained in allegations. Currently in Chicago and other major cities, civilian allegations are investigated slowly, few are sustained, and fewer still result in sanctions or disciplinary action. Our results, of course, do not provide guidance on how investigations should be conducted or what sanctions are warranted. However, the institutional disregard for civilian allegations is at odds with the finding that civilian allegations can predict serious misconduct.

34

35

References

Agreement between Chicago Fraternal Order of Police of Chicago Lodge no. 7 and the City of Chicago 2012-2017, available at http://www.chicagofop.org/contract/.

Brandl, S., M. Stroshine, and J. Frank (2001). Who are the Complaint Prone Officers? An Examination of the Relationship between Police Officers' Attributes, Arrest Activity, Assignment, and Citizens' Complaints about Excessive Force. *Journal of Criminal Justice* 29, 521–529.

Chicago Police Department Hiring Plan for Sworn Titles (2011), available at https://www.cityofchicago.org/city/en/depts/dhr/supp_info/hiring_plans.html.

Finn, P. (2001). Citizen Review of Police: Approaches and Implementation. *Institute of Justice, U.S. Department of Justice*

Finney, L. (2015). The Four-Decade Fight for Chicago Police Reform. *Chicago Tribune November 27, 2015*

Green, J. and N. Stokey (1983). A Comparison of Tournaments and Contracts. *Journal of Political Economy* 91 (3), 349–364.

Harmon, R. (2012). The Problem of Policing. *Michigan Law Review* 110 (5), 761–818.

Holmstrom, B. and P. Milgrom (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization* 7, 24–52.

Harris, C. (2009). Exploring the Relationship between Experience and Problem Behaviors: A Longitudinal Analysis of Officers from a Large Cohort. *Police Quarterly*, 12:192–213.

Harris, C. (2010). Problem Officers? Analyzing Problem Behavior Patterns from a Large Cohort. *Journal of Criminal Justice*, 38:216–225.

Harris, C. and R. Worden (2012). The Effect of Sanctions on Police Misconduct. *Crime & Delinquency.*

Kane, Robert J., and Michael D. White. "Bad cops." *Criminology & Public Policy* 8.4 (2009): 737-769.

Kane, R. and M. White (2012). *Jammed Up: Bad Cops, Police Misconduct, and the New York City Police Department*. New York University Press.

Koenker, R. and K. Hallock (2001). Quantile Regression. *Journal of Economic Perspectives* 15(4), 143-156.

Lazear, Edward P., and Paul Oyer (2012). Personnel Economics. *Handbook of Organizational Economics*, Chapter 3.

Lazear, E. and S. Rosen (1981). Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy* 89 (5), 841–864.

Lersch, K. (1998). Predicting Citizen Race in Allegations of Misconduct Against the Police. *Journal of Criminal Justice* 26: 87–97.

Lersch, K. (2002). Are Citizen Complaints Just Another Measure of Officer Productivity? An Analysis of Citizen Complaints and Officer Activity Measures. *Police Practice and Research* 3, 135–147.

Nalebuff, B. and J. Stiglitz (1983). Prizes and Incentives: Towards a General Theory of Compensation and Competition. *Bell Journal of Economics* 14 (1), 21–43.

Oyer, P. and S. Schaefer (2011). Personnel Economics: Hiring and Incentives. Volume 4, Part B, Chapter 20 of Handbook of Labor Economics, pp. 1769–1823. Elsevier.

Police Accountability Task Force (2016). Recommendations for Reform: Restoring Trust between the Chicago Police and the Communities They Serve, available at https://chicagopatf.org/wp-content/uploads/2016/04/PATF_Final_Report_4_13_16-1.pdf.

Rabe-Hesketh, S. and A. Skrondal (2012). *Multilevel and Longitudinal Modeling Using Stata, Volume 1: Continuous Responses, Third Edition.* Stata Press.

Reingold, P. (2008). Requiem for Section 1983. *Duke Journal of Constitutional Law and Public Policy* 3 (1), 1–47.

Safer, R., K. Khemani, and J. O'Keefe (2014). Preventing and Disciplining Police Misconduct: An Independent Review and Recommendations Concern Chicago's Police Disciplinary System.

Schmadeke, S., D. Heinzmann, and D. Hinkel (2015). Chicago Police Video Hindered by Errors, Omissions. *Chicago Tribune December 23, 2015*

Schwartz, M. (2006). Fundamentals of Section 1983 in Ross, Mary M, and Edwin P. Voss. *Sword and Shield: A Practical Approach to Section 1983 Litigation*. Chicago, Ill: Section of State and Local Government Law, American Bar Association, 2006.

Sen, S. (2010). Enforcing Police Accountability through Civilian Oversight, SAGE Publications.

Shultz, A. (2015). Early Warning Systems: What's New? What's Working? *CNA Analysis and Solutions.*

St. Clair, S., J. Coen and J. Smith Richards (2016), City Hall often Fights to Keep Files Secret in Police Abuse Suits. *Chicago Tribune, June 24, 2016*.

Sword and Shield: A Practical Approach to Section 1983 Litigation.

Terrill, W., and J. McCluskey (2002). Citizen Complaints and Problem Officers: Examining Officer Behavior. *Journal of Criminal Justice* 30(2): 143-155.

Walker, S., P. Alpert, and D. Kenney (2001). Early Warning Systems: Responding to the Problem Police Officer. *U.S. Department of Justice, Office of Justice Programs, National Institute of Justice.*

Walker, S., and C. Katz (2012). *Police in America*. McGraw-Hill.

Weiss, Debra Cassens (2017) "Aldermen urge federal review of Chicago law department, sanctioned eight times in police suits." *ABA Journal, January 17, 2017*.

Worden, R., C. Harris, and S. McLean (2014). Risk Assessment and Risk Management in Policing. *Policing: An International Journal of Police Strategies & Management* 37(2): 239-258.

Worden, R., M. Kim, C. Harris, M. Pratte, S. Dorn, and S. Hyland (2012). Intervention with Problem Officers: An Outcome Evaluation of an EIS Intervention. *Criminal Justice and Behavior* 40 (4), 409–437.

Worrall, J. (2002). If You Build It, They Will Come: Consequences of Improved Citizen Complaint Review Procedures. *Crime & Delinquency* 48(3): 355-379.

Van Brunt, A., S. Bedi, L. Bowman, and C. Futterman (2016), Complaint Submitted to the United States Department of Justice Documenting the Role of the Independent Police Review Authority in Perpetuating a Code of Silence and Culture of Violence in the Chicago Police Department.

Figures

Figure 1: Mean Number of Annual Allegations Per Officer by Chicago Police District

Annual Allegations
Per Officer
(.6,.7]
(.5,.6]
(.4,.5]
(.3,.4]
(.2,.3]
[.1,.2]
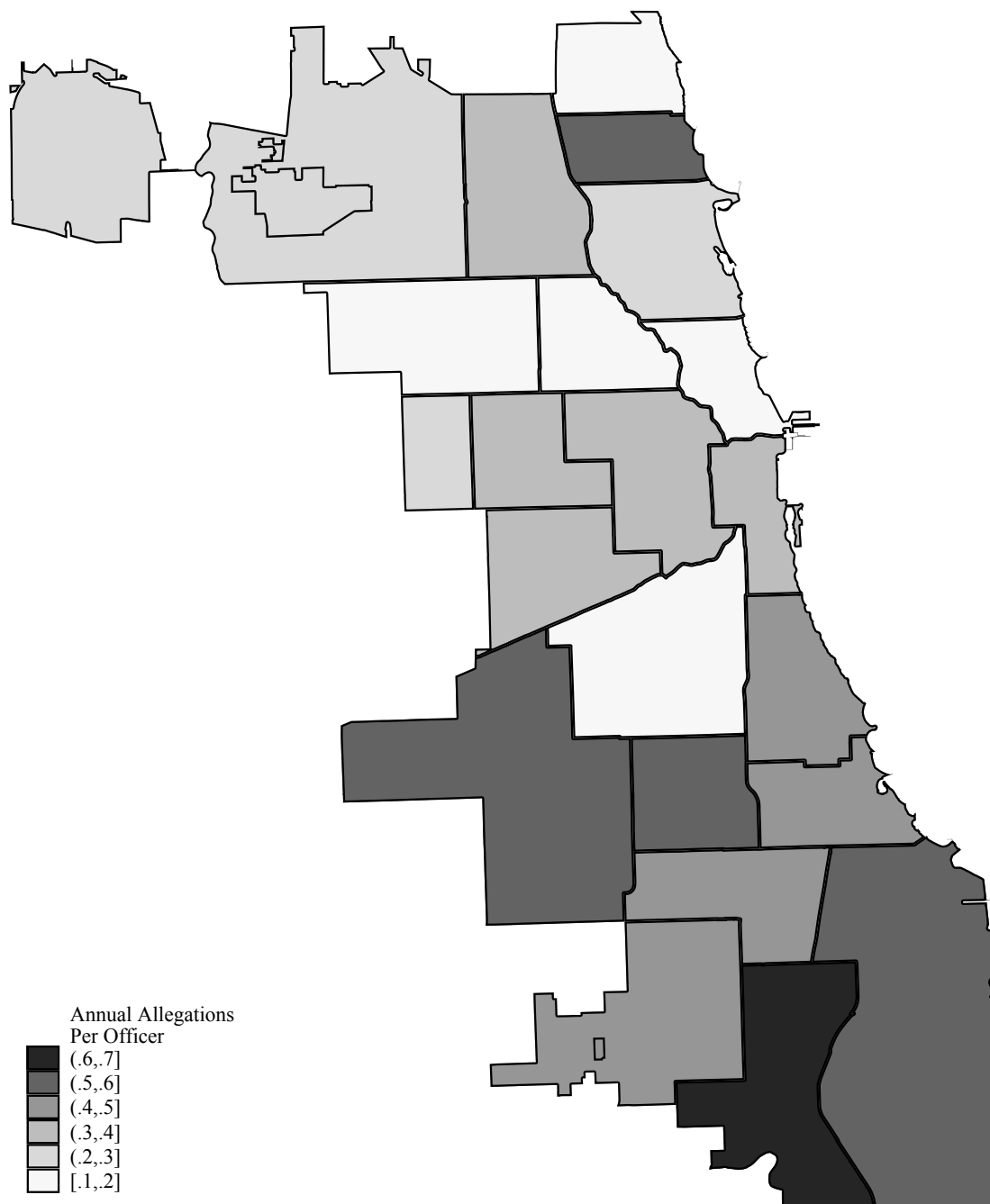
Notes: The figure reports the mean per-officer average number of annual civilian allegations by police district from 2002 to 2014.

Figure 2: Bayes Estimates: Intensity and Exposure-Adjusted Measures

A. Shrunken Allegation Distribution
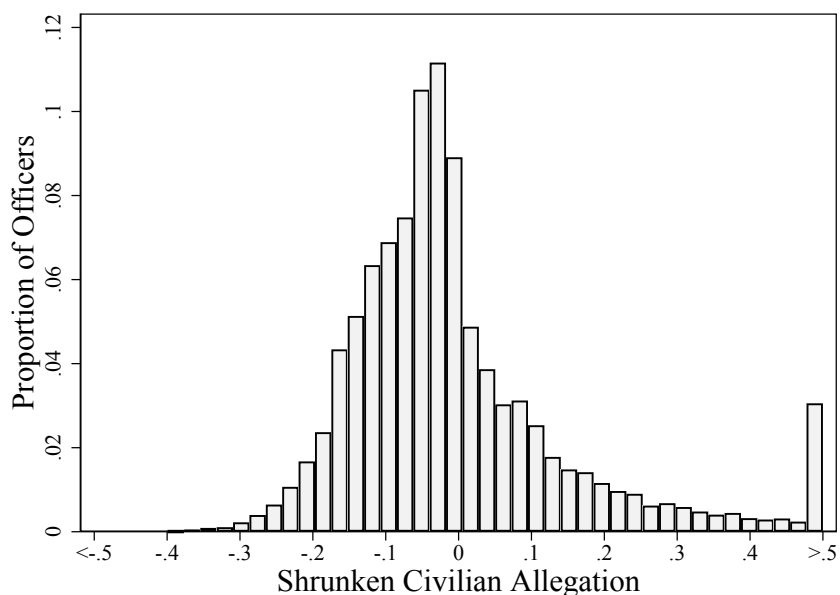
B. Misconduct Propensity Distribution
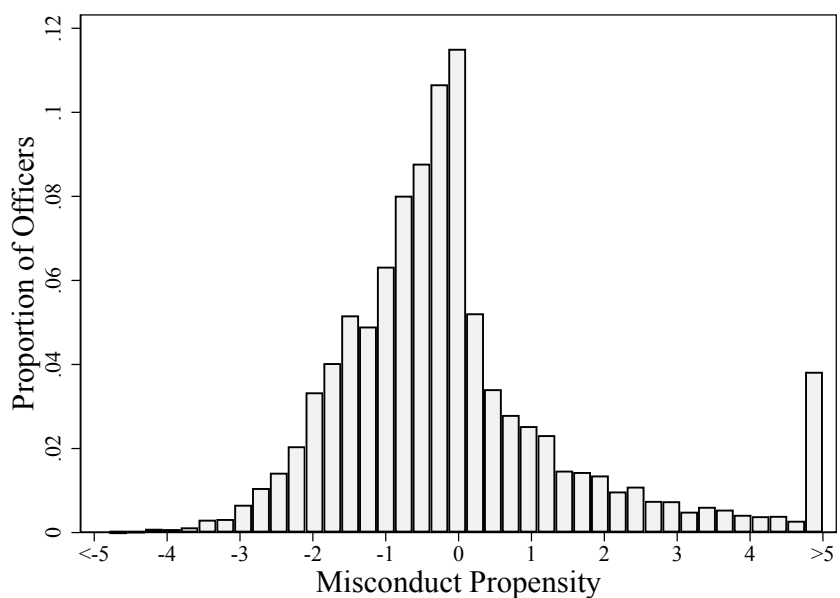
Notes: The figures plot distributions of the Bayes estimates of officer civilian allegations that shrink noisy estimates toward zero. Panel A reports the distribution of the officer-level "shrunken" civilian allegations from 2002 to 2008. The shrunken allegations are estimated controlling for officer assignment-time fixed effects and officer characteristics. Panel B reports the distribution of the exposure-adjusted measure of shrunken allegations ("misconduct propensity"). Because we observe officers for different lengths of time periods between 2009 to 2015, officers have different exposure to litigation risk in that period. The misconduct propensity is an exposure-adjusted measure of officer misconduct which is calculated by multiplying the shrunken allegation by the number of periods an officer was an on the force between 2009 and 2015.

Figure 3: Payout Amounts
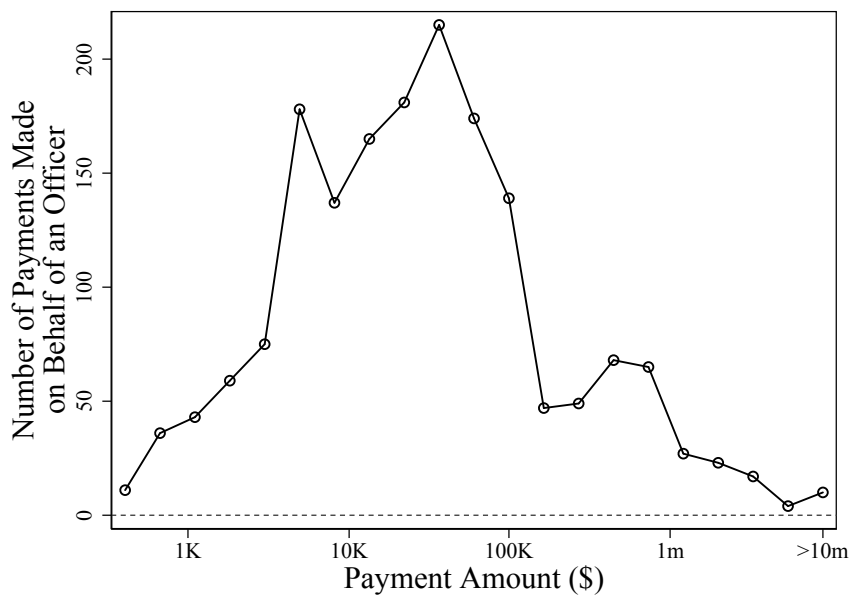
Notes: The figure plots the distribution of payments made by the City of Chicago for civil rights lawsuits in which a police officers was a named defendant between 2009 and 2014. Payments from lawsuits were matched by officer name to personnel records. The figure reports the number of payments made on behalf of an officer in 20 bins. The payment amount on the x-axis is the average payment in the bin (the x-axis in natural log scale).

Figure 4: The Relationship between First-Period Shrunken Civilian Allegations and Second-Period Civilian Allegations

A. Civilian Allegations from 2008 to 2010
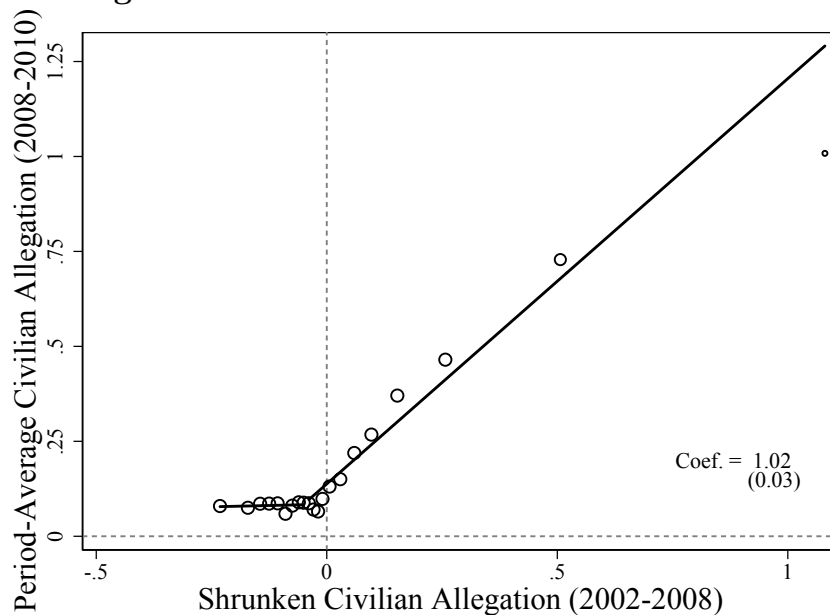
B. Civilian Allegations from 2008 to 2014
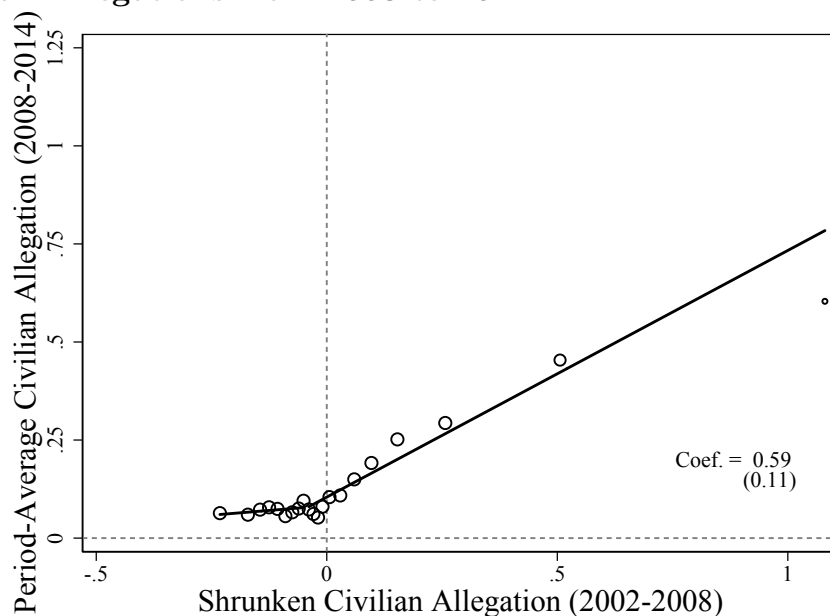
Notes: The figure is a binned scatterplot of the relationship between shrunken civilian allegations estimated relying on data from 2002 to June 2008 and civilian allegations from July 2008 to 2010 (Panel A) and civilian allegations from July 2008 to 2014 (Panel B). The shrunken civilian allegations are from Bayes estimates that control for officer assignment and officer characteristics and account for the reliability of the allegations by shrinking noisy estimates toward zero. Reported coefficients reflect the slope of the line above the median. The overall coefficients are 0.90 for Panel A and 0.53 for Panel B.

Figure 5: Effect of Leave-Out Shrunken Civilian Allegations on Actual Civilian Allegations when Officers Enter Districts

A. Differences in Mean Allegations After Arrival
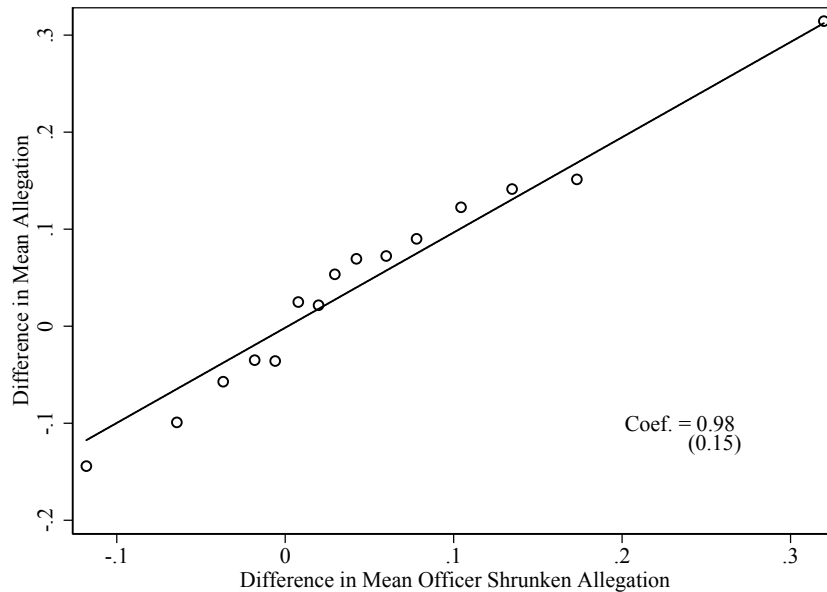
B. Allegations After Departure
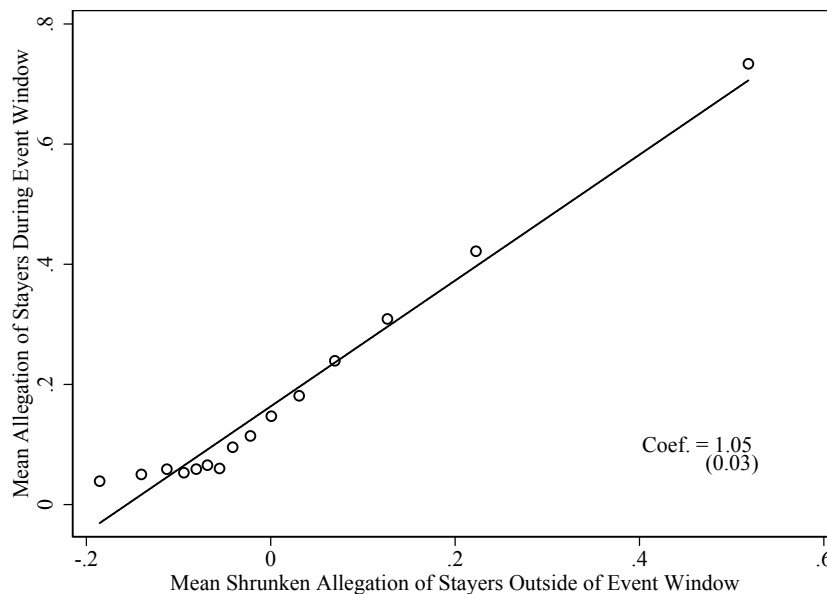
Notes: The figures are binned scatterplots to test the adequacy of our controls relying on police officers entering or departing a district. Entering officers are those who enter a new district and then remain in that district for at least three periods. Non-switching officers are in the district for 2 periods before officers switch into the district and remain in a district for at least 2 periods after. A shrunken "leave-out" measure of the civilian allegations is estimated for each officer time-period based only on the allegations that occur outside the event. We estimate the leave-out shrunken allegation separately for each event, using all others that are not in the event as controls. In Panel A, the x-axis is the difference between the average shrunken allegation for the switching and the non-switching officers outside the event window and the y-axis is the difference in the average allegations for the switching and non-switching officers during the event window. There are 780 events in which officers switch into a district. In Panel B, the x-axis is the shrunken allegations of non-switching officers in a district-time event before the departure of officers and the y-axis is their mean allegation after the departure. There are 1461 events in which officers switch out of a district. See text for more details.

Figure 6: Effects of Outlier Officer Entry and Departure on Civilian Allegations

A. High Misconduct Officer Entry

Mean Diff. = 0.014***
$p<0.01$

District Mean Allegations

Years Relative to Entry of High Misconduct Officers

B. High Misconduct Officer Departure

Mean Diff. = 0.024***
$p<0.01$

District Mean Allegations

Years Relative to Departure of High Misconduct Officers

C. Low Misconduct Officer Entry

Mean Diff. = -0.005***
$p<0.01$

District Mean Allegations

Years Relative to Entry of Low Misconduct Officers

D. Low Misconduct Officer Departure

Mean Diff. = -0.006***
$p<0.01$

District Mean Allegations

Years Relative to Departure of Low Misconduct Officers

⊙——— Fixed Non-Switching Cohort
△ – – – Cohort with Switchers in the Worst or Best 5% Differential

Notes: The figures plot event studies of average civilian allegations for the cohort of non-switching officers (solid line) and the cohort of all officers (dashed line) in a district as officers enter or leave a district-time period cell in time t = 0. Non-switching officers are in the district for 2 periods before officers switch into the district and remain in a district for at least 2 periods after. A shrunken civilian allegation is estimated for each officer time-period based only on the allegations that occur outside the event window. For each cohort of officers entering or departing a district, we calculate the difference in the mean shrunken allegation of the switching cohort and the non-switching cohort and define a switching high (low) misconduct officer cohort as one that is in the top (bottom) 5 percent of this differential.

Figure 7: The Relationship between Shrunken Civilian Allegations and Future Non-Civilian Allegations

A. Future Internal Allegations
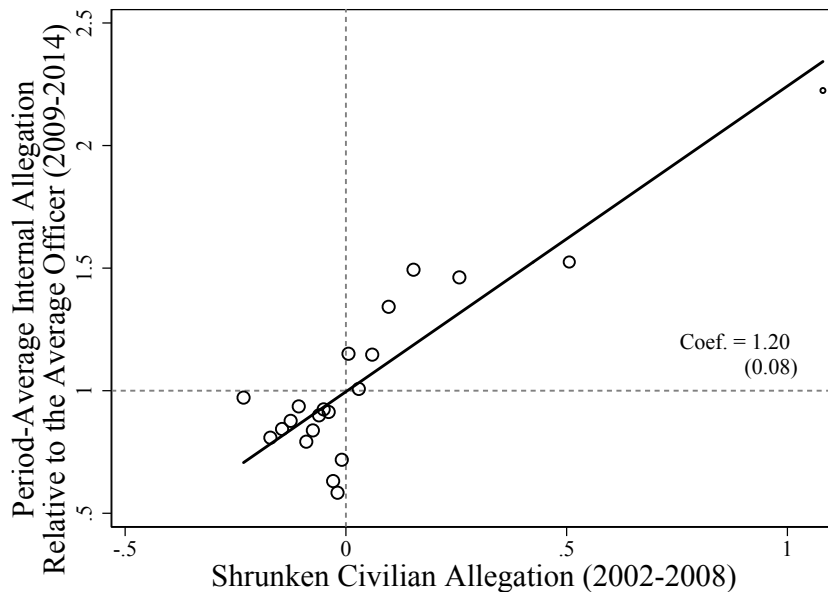
B. Future Off-Duty Allegations
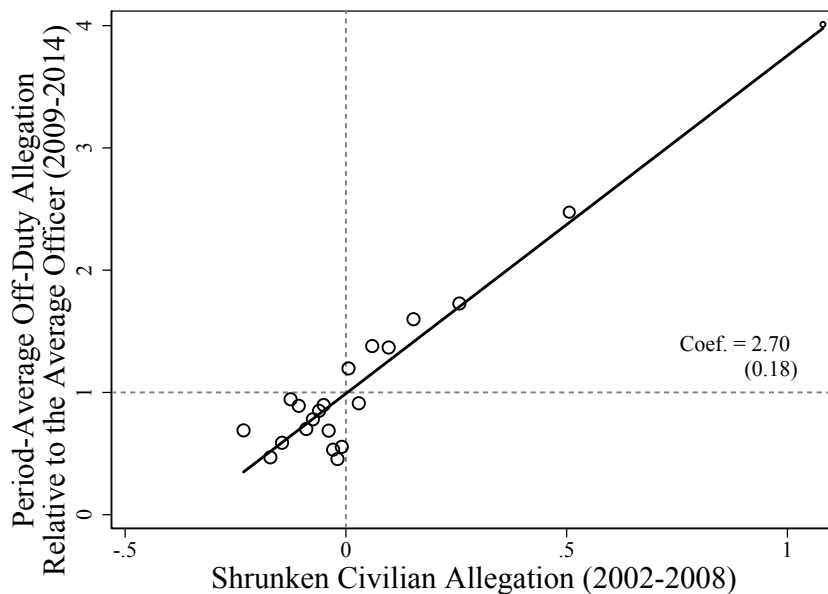
Notes: The figures report binned scatterplots of the relationship between shrunken civilian allegations from 2002 to June 2008 and internal allegations (Panel A) and off-duty allegations (Panel B) from July 2008 to 2014. Each figure breaks out the top 1 percent of officers into a separate bin. The y-axis is scaled by dividing the non-civilian allegations by their sample averages, where a "1" on the y-axis means that the outcome is at the sample average of non-civilian allegations and a "2" means that the outcome is at twice the average.

Electronic copy available at: https://ssrn.com/abstract=2866696

Figure 8: The Relationship between Shrunken Civilian Allegations and Future Civil Rights Litigation

A. Number of Lawsuits Named In

B. Payouts in Lawsuits

C. Payouts in Lawsuits Greater than $100k

D. OLS Damages in Levels

E. OLS Damages in Natural Log

F. Damages: Quantile Regression (95th)

Notes: The figures report point estimates from regressing litigation outcomes from 2009 to 2014 on the exposure-adjusted shrunken civilian allegations ("misconduct propensity") from 2002 to June 2008. The estimates are on indicators for groups of officers according to the decile of their misconduct propensity, where we break the top decile into 90th to 95th percentiles, 95th to 99th percentiles, and the top 1 percent of officers. The x-axis indicates the mean misconduct propensity for a given group of officers. Officers in the lowest decile are the reference group and each panel reports the sample mean for comparison purposes. Standard error bars represent 95 percent confidence intervals.

Tables

Table 1: Police Officer Misconduct Allegations

Type of Misconduct	Number of Allegations	Percent of Total (%)
Civilian Allegations		
First Amendment and Illegal Arrest	23840	46.6
Arrest/Lock-up Procedures	19944	39.0
Search-Related	4624	9.0
Verbal Abuse	2793	5.5
Total	51201	100
Internal Allegations		
Failure to Provide Service	12033	42.4
Neglect of Duty	4112	14.5
Failure to Make a Report	1655	5.8
Other Internal	10553	37.2
Total	28353	100
Off-Duty Allegations		
Traffic	1636	31.5
Conduct Unbecoming (Off-duty)	1490	28.7
Criminal Misconduct	1132	21.8
Substance Abuse	343	6.6
Total	4601	100

Notes: The table reports allegations made between 2002 and 2014. Civilian allegations relate to interactions between civilians and on-duty police officers. Internal allegations arise from reports from supervisors or fellow officers. Off-duty allegations are mostly generated by off-duty behavior including criminal behavior.

Table 2: Descriptive Statistics

Independent Variables from Years 2002 to 2008	
Mean Civilian Allegations	0.16
Total Variance of Civilian Allegations ($\text{Var}(\nu_{ijt})$)	0.21
Variance of Predicted Allegations ($\text{Var}(\mu_j)$)	0.05
Outcomes: Years After 2008	
Mean Per-Period Civilian Allegations	0.11
Mean Per-Period Internal Allegations	0.06
Mean Per-Period Other Allegations	0.01
Number of Lawsuits Per Officer	0.23
Number of Payouts Per Officer	0.14
Number of Payouts Per Officer in Lawsuit >$100k	0.03
Mean Lawsuit Damages	$40,519

Notes: The table reports descriptive statistics of the sample and of the shrunken allegations produced from Bayes estimation. The first panel reports the shrunken allegations from 2002 to June 2008. The second panel reports descriptive statistics of the outcome variables from July 2008 to 2014.

Table 3: Effect of Leave-Out Shrunken Civilian Allegations on Actual Civilian Allegations when Officers Switch Districts

	Levels		Changes	
	(1)	(2)	(3)	(4)
A. Officer Entry				
Measure of Shrunken	0.98	1.01	1.17	1.24
Civilian Allegation	(0.17)	(0.14)	(0.25)	(0.37)
Events	780	780	780	780
B. Officer Departure				
Measure of Shrunken	1.05	1.06	3.21	2.78
Civilian Allegation	(0.03)	(0.03)	(0.88)	(1.21)
Events	1461	1461	775	775
Covariates				
Time Period FE	No	Yes	No	Yes
District FE	No	Yes	No	Yes

Notes: See notes to Figure 5 for a description.

Table 4: Impacts of Outlier Officer Entry and Departure on Civilian Allegations

	Civilian Allegations		
	(1)	(2)	(3)
A. Officer Entry			
Entering Officer	0.013	0.018***	0.025***
	(0.009)	(0.005)	(0.005)
Entering Officer	0.851***	0.809***	0.803***
× High Misconduct Officer	(0.078)	(0.074)	(0.072)
Entering Officer	-0.138***	-0.215***	-0.210***
× Low Misconduct Officer	(0.014)	(0.015)	(0.016)
R-squared	0.012	0.071	0.088
Dep Var Mean	0.178	0.178	0.178
B. Officer Departure			
Departing Officer	0.003	-0.001	0.026***
	(0.009)	(0.010)	(0.009)
Departing Officer	0.818***	0.768***	0.767***
× High Misconduct Officer	(0.074)	(0.071)	(0.070)
Departing Officer	-0.166***	-0.195***	-0.220***
× Low Misconduct Officer	(0.012)	(0.036)	(0.032)
R-squared	0.007	0.055	0.074
Dep Var Mean	0.200	0.200	0.200
Covariates			
Time Period FE	No	Yes	No
District FE	No	Yes	No
District-Time FE	No	No	Yes

Notes: The table reports regressions of civilian allegations after officers enter or depart a district on leave-out shrunken allegations from outside the event. Panel A reports the results of outlier officers entering a district, and Panel B reports the results of outlier officers departing a district. Outlier officers are defined as in the top or bottom 5 percent of the differential between their shrunken allegation and the mean shrunken allegation of the non-switching cohort. Column 1 has no controls, Column 2 adds district and time fixed effects, and Column 3 adds district-time fixed effects. Standard error clustered by event in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 5: The Relationship between Exposure-Adjusted Shrunken Civilian Allegations from 2002 to 2008 and Civil Litigation from 2009 to 2014

| | Number of Lawsuits | Payouts in Lawsuit | | Amount of Lawsuit Damages | | |
		All	>$100k	Levels	Logs	95th Quantile
	(1)	(2)	(3)	(4)	(5)	(6)
A. Linear Specification						
Misconduct Propensity	0.048***	0.041***	0.008***	9.08***	0.31***	16.7***
	(0.002)	(0.002)	(0.001)	(1.88)	(0.01)	(2.0)
B. Linear Specification with Controls						
Misconduct Propensity	0.044***	0.040***	0.007***	8.03***	0.30***	16.4***
	(0.002)	(0.002)	(0.001)	(2.00)	(0.02)	(2.1)
C. Linear Specification with Spline						
Misconduct Propensity (≤ 0)	0.014**	0.016***	0.000	4.89	0.14***	-2.4
	(0.007)	(0.005)	(0.002)	(5.54)	(0.04)	(7.5)
Misconduct Propensity (> 0)	0.057***	0.048***	0.010***	10.24***	0.36***	62.3***
	(0.003)	(0.002)	(0.001)	(2.37)	(0.02)	(3.2)
N	12,075	12,075	12,075	12,075	12,075	12,075
Dep Var Mean	0.232	0.144	0.030	41.32	1.32	41.3

The table reports regressions of litigation outcomes on exposure-adjusted shrunken civilian allegations from 2002 to 2008. We adjust for exposure in the post-period by multiplying the number of periods observed in the 2009 to 2014 time frame by the shrunken civilian allegations. Standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Appendix

Figure A1: Civil Rights Lawsuits against Chicago Police Officers over Time
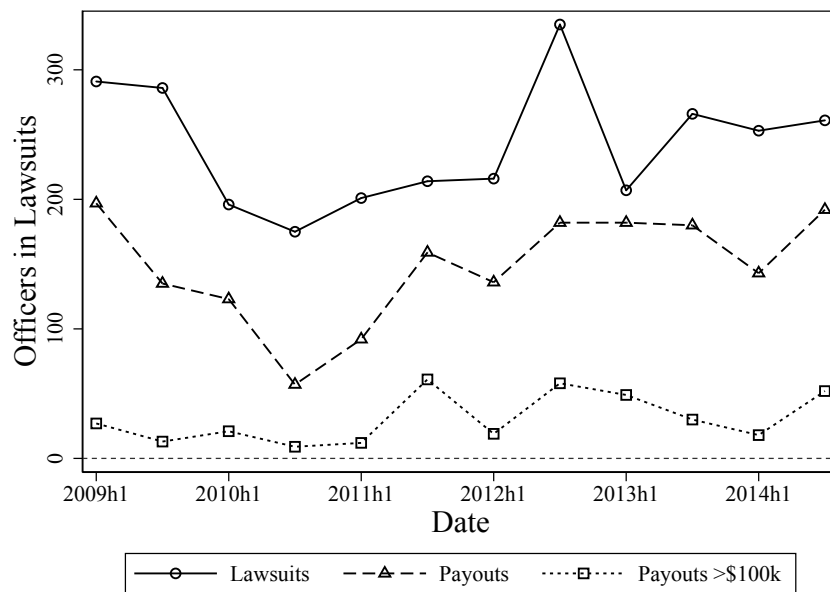
Notes: The figure reports the number of lawsuits, payouts, and payouts over $100,000 between 2009 and 2014 for each six-month period.

Figure A2: Quantile Regressions
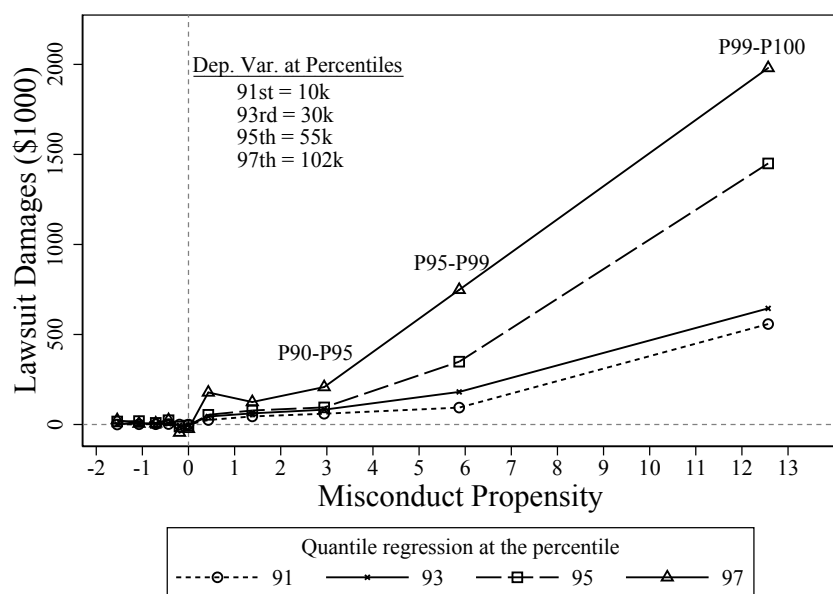
Notes: The figure is analogous to Panel F of Figure 9, but reports 91st, 93rd, 95th, and 97th quantile regression results.

Figure A3: Relationship between Shrunken Civilian Allegations and Future Civil Litigation using Different Controls

A. Lawsuits Named In
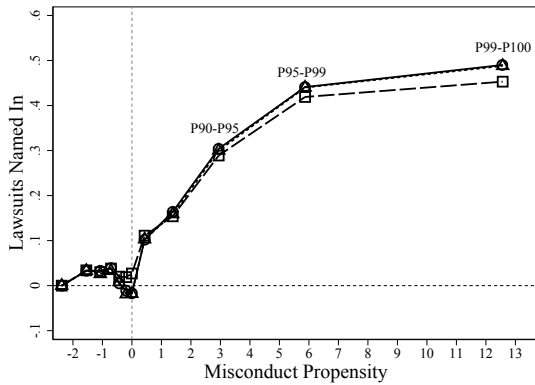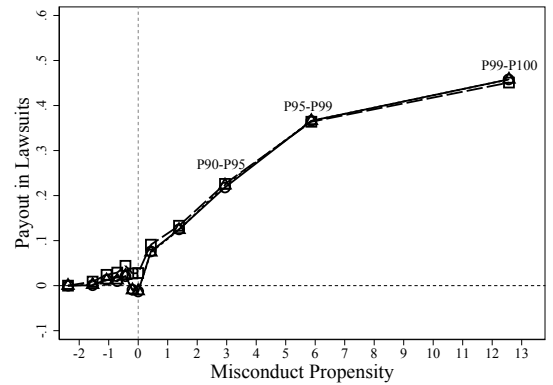
B. Payouts in Lawsuits

C. Payouts in Lawsuits Greater than $100k
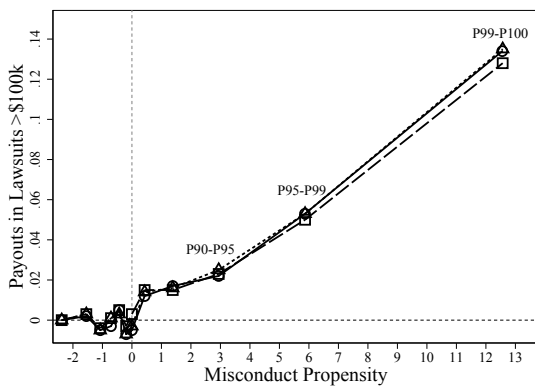
D. OLS Damages in Levels
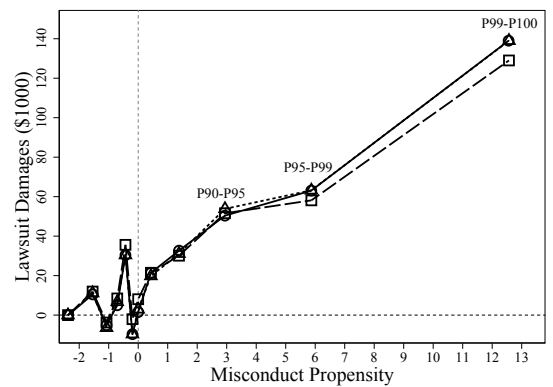
E. OLS Damages in Natural Log
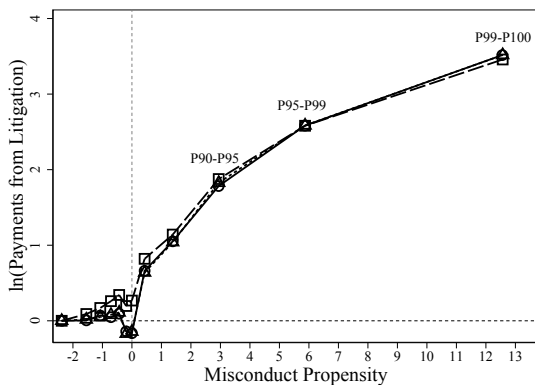
F. Damages: Quantile Regression (95th)
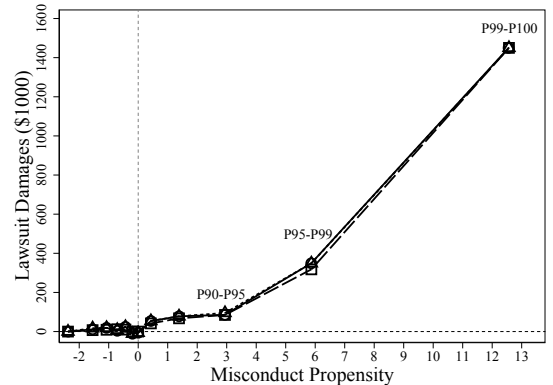
Controls
— ○ — District FE + Time FE + Lagged District Mean & Variance
····△···· District-Time FE + Officer Characteristics
— ☐ — District-Time FE + Officer Characteristics + Other Misconduct

Notes: The figure is analogous to Figure 10 in the main body and reports point estimates from regressing litigation outcomes from 2009 to 2014 on the exposure-adjusted shrunken civilian allegations ("misconduct propensity") from 2002 to June 2008. The estimates are on indicators for groups of officers according to the decile of their misconduct propensity, where we break the top decile into 90th to 95th percentiles, 95th to 99th percentiles, and the top 1 percent of officers. The x-axis indicates the mean misconduct propensity for a given group of officers. Officers in the lowest decile are the reference group and each panel reports the sample mean for comparison purposes. Each line represents a different set of controls as defined in the accompanying legend.

Figure A4: Effect of Leave-Out Shrunken Civilian Allegation on Allegations when Officers Depart a District

A. Differences in Allegations from Before to After the Arrival
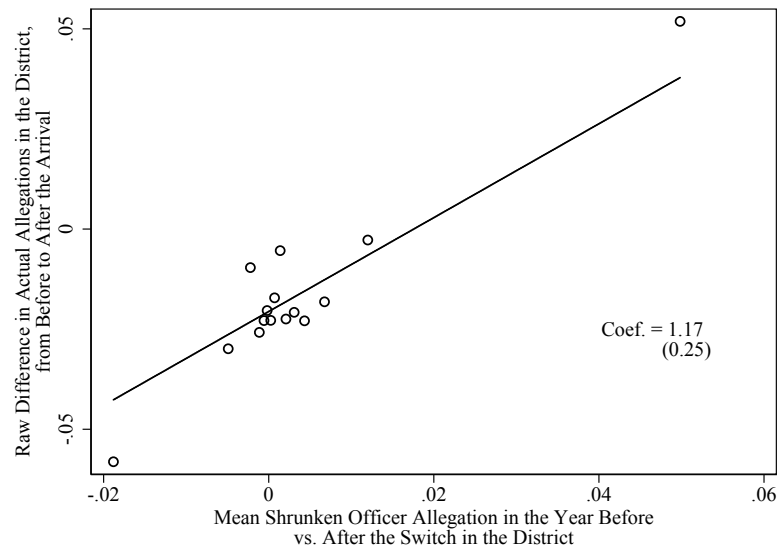
B. Differences in Allegations from Before to After the Departure
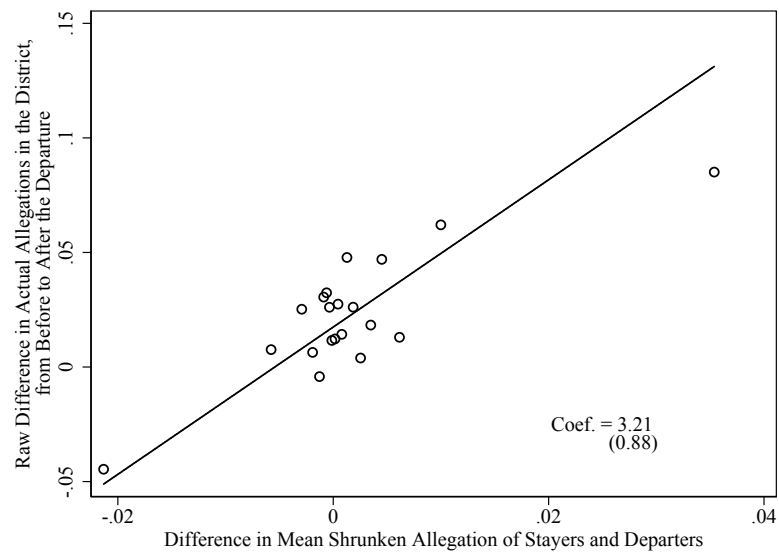
Notes: The figures are binned scatterplots to test the adequacy of our controls relying on police officers entering or departing a district. Switching officers are those who are in a district for at least two periods and then depart the district. Non-switching officers are in the district for 2 periods before officers switch out of the district and remain in the district for at least 2 periods after. A shrunken "leave-out" measure of the civilian allegations is estimated for each officer time-period based only on the allegations that occur outside the event. We estimate the leave-out shrunken allegation separately for each event, using all others that are not in the event as controls. In Panel A, the x-axis is the difference in average shrunken allegations of officers in the district from before the entry event, which does not include the switching officers, to after the event, which includes the switching officers; the y-axis is the difference in average allegations of officers in the district in the event window from before the entry event, which does not include the switching officers, to after the event, which includes the switching officers. There are 780 events in which officers switch into a district. In Panel B, the x-axis is the difference in average shrunken allegations of officers in the district from before the departure event, which includes the switching officers, to after the event, which does not include the switching officers; the y-axis is the difference in average allegations of officers in the district in the event window from before the entry event, which includes the switching officers, to after the event, which does not include the switching officers. There are 775 events in which officers switch out of a district. See text for more details.
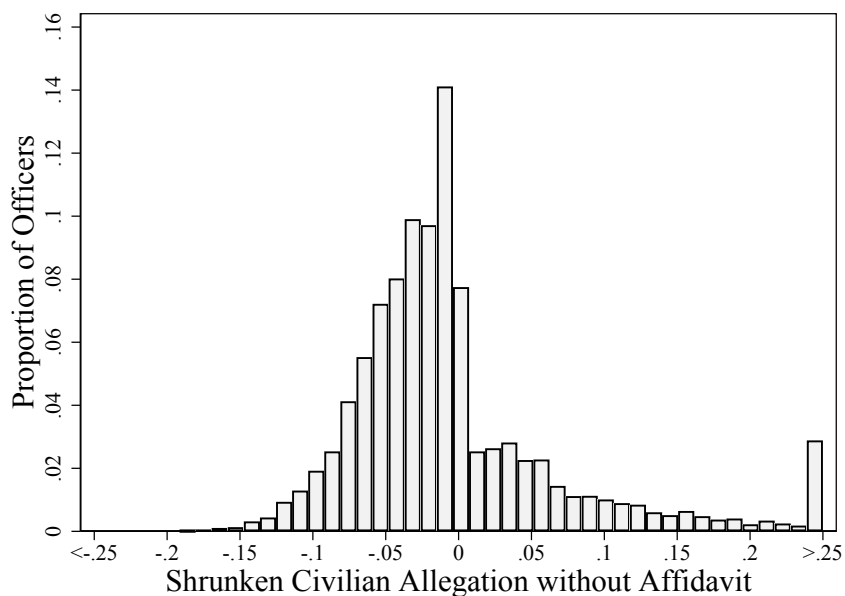
Figure A5: Distribution of Shrunken Allegations by Affidavit-Based Allegations and Non-Affidavit Based Allegations

A. Shrunken Civilian Allegation with Affidavit Distribution
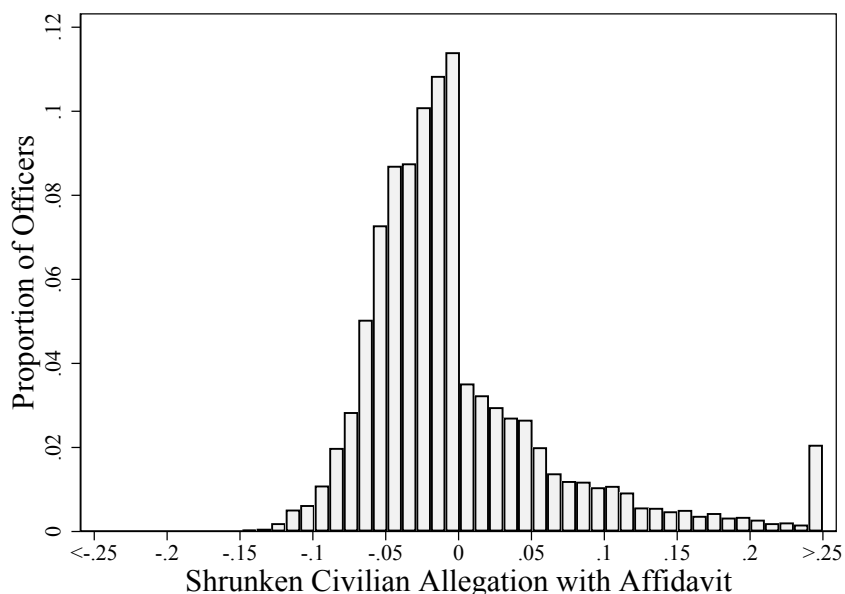
B. Shrunken Civilian Allegation without Affidavit Distribution

Notes: The figures plot distributions of the Bayes estimates of officer civilian allegations that shrink noisy estimates toward zero. Panel A reports the distribution of the officer-level "shrunken" civilian allegations from 2007 to 2011 that controls for officer assignment-time fixed effects and officer characteristics for allegations supported by an affidavit. Panel B reports the distribution for allegations not supported by an affidavit.

Table A1: Probability Officers Remain on the Force

	Still On the Force	
	in 2009	in 2015
	(1)	(2)
Shrunken Civilian Allegation	0.018	0.011
	(0.014)	(0.014)
Appointment Year FE	Yes	Yes
Birth Year FE	Yes	Yes
Obs	13,468	11,281
Dep Var Mean	0.830	0.790

Notes: The table reports linear probability models taking on the value 1 if the officer is observed in the time period in the column heading. In the first column, the dependent variable takes on the value 1 if the officer from the period 2002 through June 2008 remains on the force in 2009. This tests for whether an officer is observed in the first period but never observed in the second. In the second column, the dependent variable takes on the value one if an officer observed in the second period remains on the force through 2015. This tests for selection during the post-period.

Figure A6: The Relationship between First-Period Shrunken Civilian Allegations and Second-Period Civilian Allegations

A. Shrunken Allegation With Affidavit
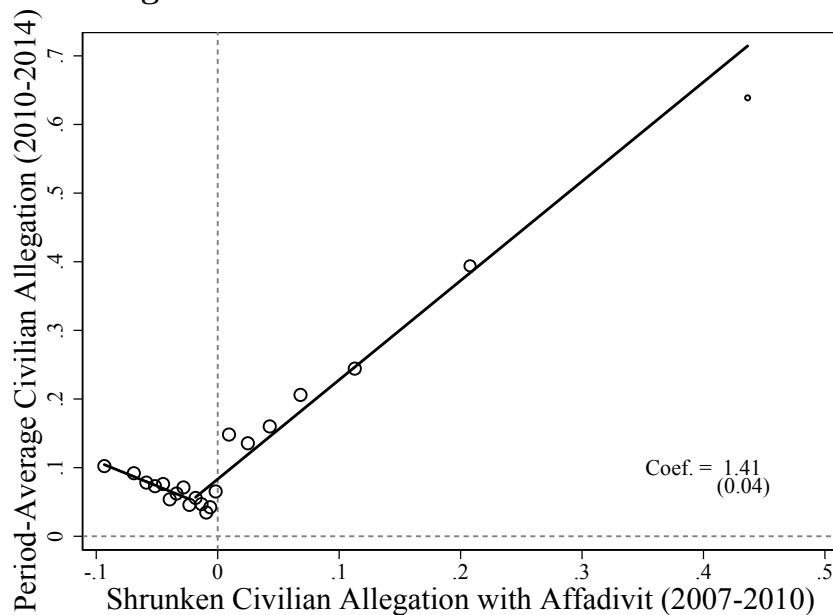
B. Shrunken Allegation Without Affidavit
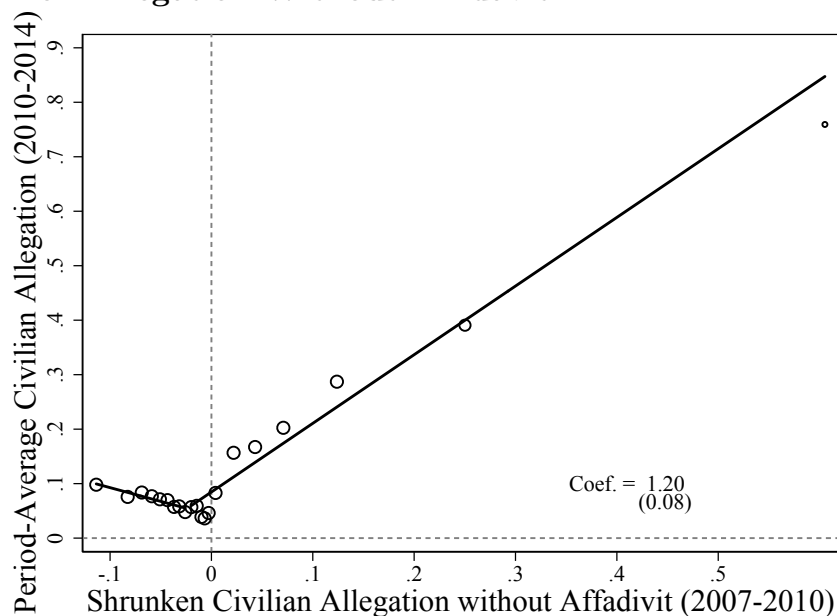
Notes: The figure is a binned scatterplot of the relationship between shrunken civilian allegations estimated relying on data from 2007 to June 2010 and civilian allegations from July 2010 to 2014. Panel A is for shrunken allegations with an affidavit and Panel is for shrunken allegations without an affidavit. The shrunken civilian allegations are from Bayes estimates that control for officer assignment and officer characteristics and account for the reliability of the allegations by shrinking noisy estimates toward zero.

Figure A7: The Relationship between Shrunken Civilian Allegations with and without Affidavits and Future Civil Rights Litigation
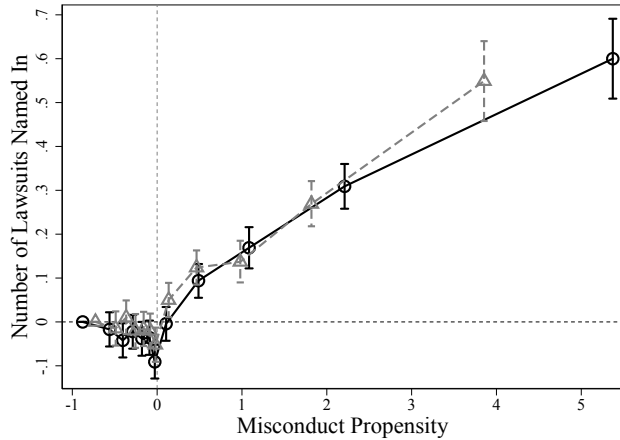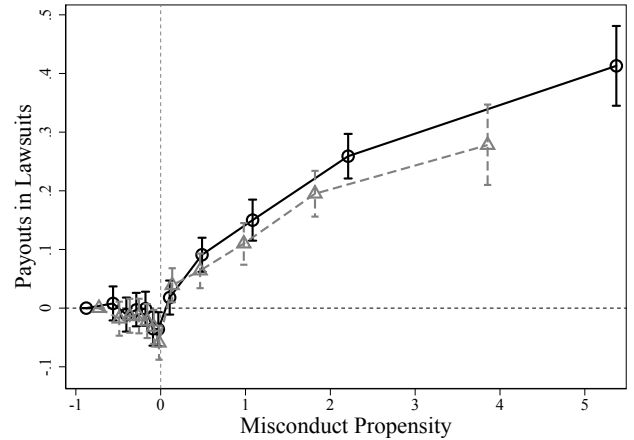
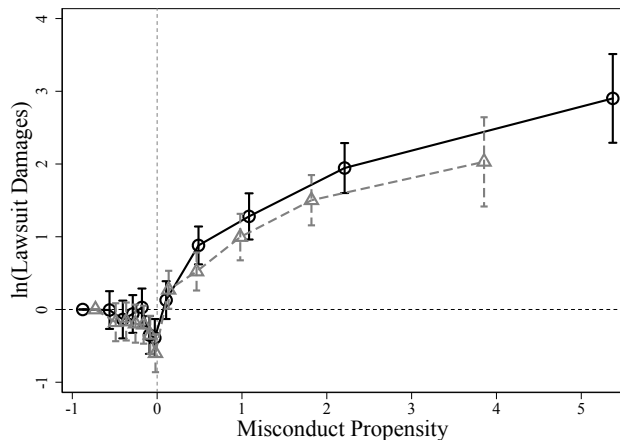### A. Number of Lawsuits Named In

B. Payouts in Lawsuits
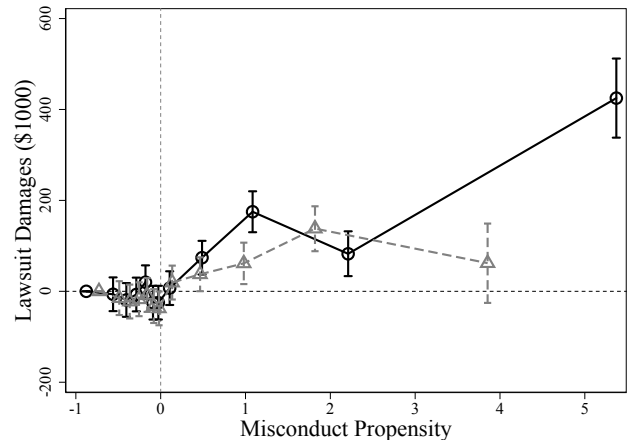

C. Payouts in Lawsuits Greater than $100k
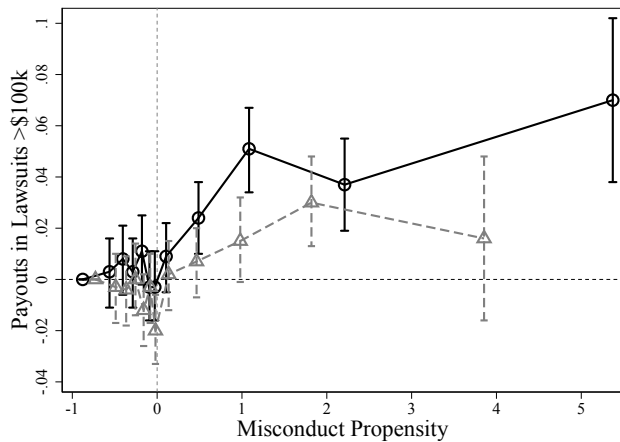
D. OLS Damages in Levels
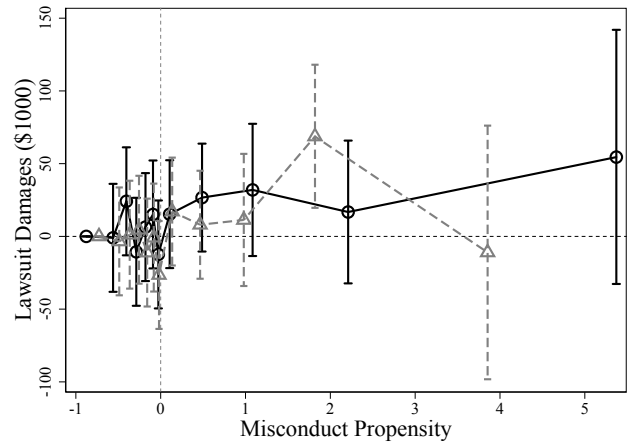
E. OLS Damages in Natural Log

F. Damages: Quantile Regression (95th)

Without Affidavit — With Affidavit

Notes: The figures report point estimates from regressing litigation outcomes from July 2010 to 2014 on the exposure-adjusted shrunken civilian allegations with and without affidavits ("misconduct propensity") from 2007 to June 2010. The estimates are on indicators for groups of officers according to the decile of their misconduct propensity with and without affidavits, where we break the top decile into 90th to 95th percentiles, 95th to 99th percentiles, and the top 1 percent of officers. The x-axis indicates the mean misconduct propensity for a given group of officers. The solid line reports the results for allegations without an affidavit. The dashed line reports the results for allegations with an affidavit. Officers in the lowest decile are the reference group and each panel reports the sample mean for comparison purposes. Standard error bars represent 95 percent confidence intervals.

Table A2: The Relationship between Exposure-Adjusted Shrunken Civilian Allegations with and without Affidavits from 2007 to June 2010 and Civil Litigation from July 2010 to 2014

	Named in Lawsuit	Payouts in Lawsuit		Amount of Lawsuit Damages		
		All	>$100k	Levels	Logs	95th Quantile
	(1)	(2)	(3)	(4)	(5)	(6)
Misconduct Propensity	0.127***	0.078***	0.011***	13.38**	0.62***	34.5***
(Allegations with Affidavit)	(0.006)	(0.005)	(0.002)	(6.10)	(0.04)	(4.8)
Misconduct Propensity	0.114***	0.081***	0.013***	9.13*	0.61***	32.4***
(Allegations without Affidavit)	(0.005)	(0.004)	(0.002)	(4.85)	(0.03)	(4.0)
N	11,651	11,651	11,651	11,651	11,651	11,651
Dep Var Mean	0.178	0.112	0.026	37.75	1.06	37.7

Notes: The table reports regressions of litigation outcomes from July 2010 to 2014 on exposure-adjusted shrunken civilian allegations from 2007 to June 2010. We adjust for exposure in the post-period by multiplying the number of periods. Panel A reports the results for civilian allegations with an affidavit and Panel B reports the results for civilian allegations without an affidavit.

To investigate the extent that affidavits are relevant to the quality of the signal contained in allegations, we estimate a modified version of Equation (5) that places unequal weight on allegations with and without an affidavit. In particular, we estimate a set of misconduct propensities according to Equation (9), and then estimate the relationship between misconduct propensity and litigation in Equation (6) for each iteration of the misconduct propensity.

$$M_j = \left(p \times \hat{\mu}_j^{\text{Affidavit}} + (1 - p) \times \hat{\mu}_j^{\text{No Affidavit}} \right) \times E_j \tag{9}$$

where $\hat{\mu}_j^{\text{Affidavit}}$ is the shrunken civilian allegation for allegations with an affidavit, $\hat{\mu}_j^{\text{No Affidavit}}$ is the shrunken civilian allegation for allegations with an affidavit, $p \in [0, 1]$ is the weight placed on the shrunken allegation for allegations with an affidavit, and E_j is the exposure to litigation. We vary p from 0 to 1 in 0.001 increments, and estimate Equation (6) for each of the litigation outcomes. Using the distribution of point estimates on the coefficients in each iteration of p, we then plot the median and 95 percent confidence region for the point estimates, where the value of misconduct propensity on the x-axis for each iteration is that for equal weighting.

The results are reported in Figure A6. The estimated size of the relationship between misconduct propensity and the outcomes partly depends on the weighting of allegations with and without affidavits, but the overall relationship remains strong in all weighting iterations. Figure A7 reports the Akaike's Information Criterion, which is a standard criterion for comparing the quality of the model to explain data, from each of the 1,000 regressions with different values of p. The model that best fits the data minimizes the AIC. The results suggest that close to equal weighting is favored, and reject weighting allegations with an affidavit more than allegations without an affidavit. We find similar results using other test criterion, including Bayesian Information Criterion (BIC) and Likelihood Ratio Chi-Square (G2), and well tests for goodness of fit including the R squared.

Figure A8: Sensitivity of Results to Weighting of Allegations with and without Affidavit
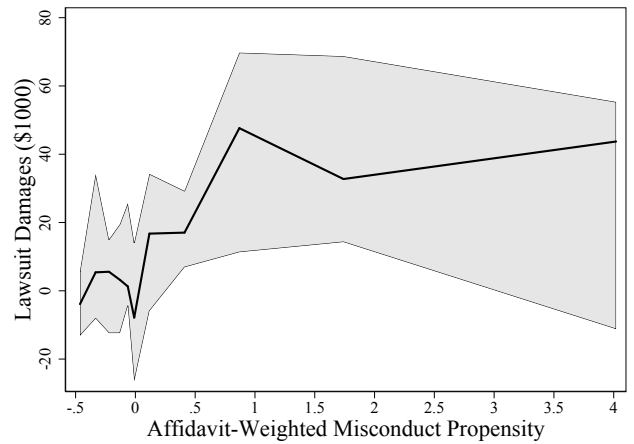
A. Lawsuits Named In
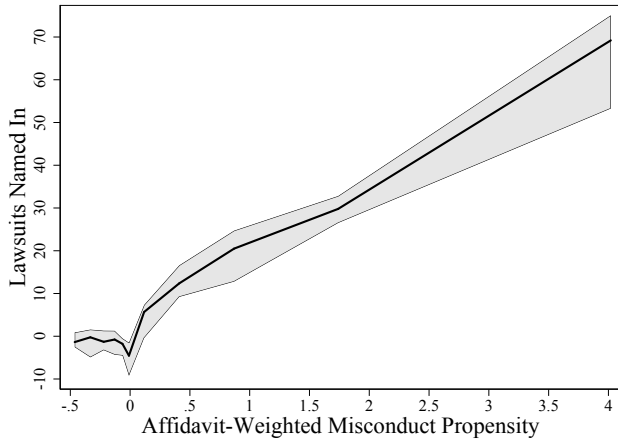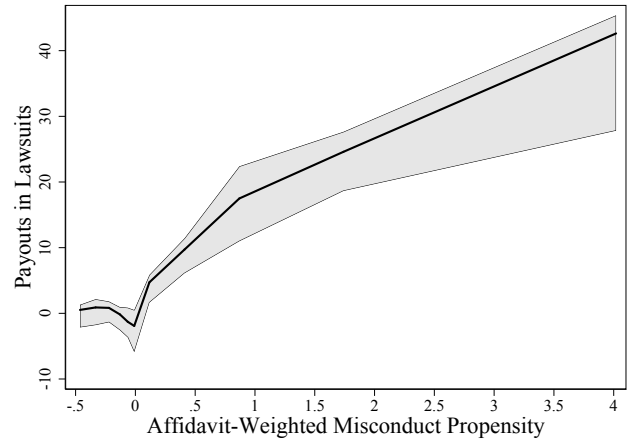
B. Payouts in Lawsuits
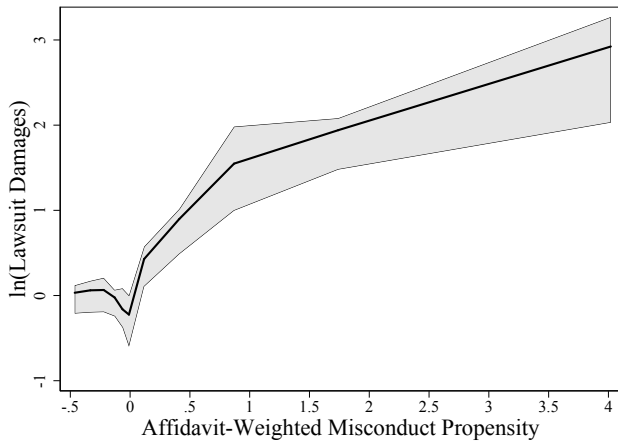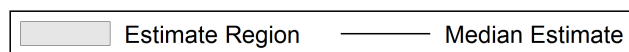

C. Payouts in Lawsuits Greater than $100k

## D. OLS Damages in Levels

E. OLS Damages in Natural Log

F. Damages: Quantile Regression (95th)

Estimate Region ──────── Median Estimate

Notes: Results reported by deciles of shrunken allegation. The excluded category is the first decile, and the last decile is broken into 90th to 95th percentile, 95th to 99th percentile, and the top percentile. Each panel presents results from a series of regressions placing different weights on allegations with and without affidavits. Using the distribution of point estimates on the coefficients in each iteration, the figure plots the median and 95 percent confidence region for the point estimates, where the value of misconduct propensity on the x-axis for each iteration is that for equal weighting.

Figure A9: Optimal Weighting Between Shrunken Allegation with and without Affidavit using Akaike's Information Criterion (AIC)
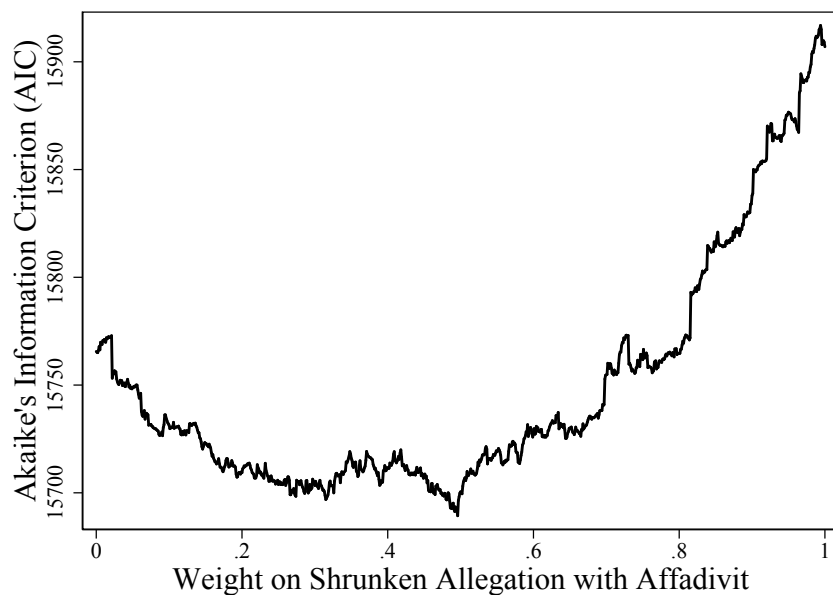
Notes: The figure reports the Akaike's Information Criterion, which is a standard criterion for comparing the quality of the model to explain data, from 1,000 regressions with different weighting of allegations with and without affidavits. The model that best fits the data minimizes the AIC. See text in the Appendix for more details.