


ORIGINAL ARTICLE

Complaints about police misconduct have adverse effects for Black civilians

Patrick W. Kraft¹  and Benjamin J. Newman²

¹Juan March Institute & Department of Social Sciences, Carlos III University of Madrid, Madrid, Spain and ²School of Public Policy & Department of Political Science, University of California, Riverside, CA, USA

Corresponding author: Benjamin J. Newman; Email: bnewman@ucr.edu

(Received 10 July 2022; revised 5 July 2023; accepted 10 July 2023)

Abstract

Existing literature examines the effectiveness of civilian oversight in reducing police misconduct. However, little-to-no quantitative research explores possible adverse consequences of this accountability mechanism. Utilizing time series analysis of administrative data on aggregate monthly civilian complaints and police behavior in the largest American city, this article offers evidence of racial inequality in police response to civilian complaints. For White civilians, complaint against the police abates subsequent police stops. For Black civilians, complaint is associated with subsequent intensification of police stops. This intensification only follows complaints against White officers, is conditional upon officer knowledge of the complaint, is confined to stops involving greater officer discretion to perform the stop, and is only observed in police precincts with large Black populations.

Keywords: police accountability; police reform; policing; racial bias; stop and frisk; systemic racism

1. Introduction

A persistent and major policy issue in the USA is racial bias in law enforcement (Terrill and Mastrofski, 2002; Engel and Calnon, 2004; Gelman *et al.*, 2007). Scholars, politicians, and reformers have proposed a variety of interventions to redress such bias, including more rigorous screening of recruits (Haas, 2013; Kluger, 2016), increasing the racial diversity of police forces (DOJ, 2016; Ba *et al.*, 2021), training about implicit bias (Baker, 2018), accountability-enhancing protocols (Warren and Tomaskovic-Devey, 2009; Mummolo, 2018), and requiring officers to wear body cameras (Voigt *et al.*, 2017; Yokum *et al.*, 2017). One intervention of growing prevalence and importance in the USA and abroad is civilian oversight of police and the inauguration of civilian oversight agencies (COAs) (Walker, 2001; Ferdik *et al.*, 2013; Ali and Pirog, 2019; Kim, 2022).

COAs are independent investigative and external oversight agencies comprised of civilian investigators who are not sworn officers. The primary functions of COAs are typically to establish formal civilian complaint processes, to perform independent investigations into civilian complaints against police officers, and to render a case disposition which may involve recommended disciplinary actions. Advocates of COAs contend that they have the capacity to deter officer misconduct and increase the social accountability of local police forces. Currently, the limited amount of empirical research on the effects of COAs suggests that they can encourage civilians to complain about perceived mistreatment by the police (Smith and Holmes, 2003), reduce police misconduct (Brereton, 2000), and diminish racial bias in arrests and police shootings (Ali and Pirog, 2019).

While scholarship has identified several potential problems with COAs as a means of redressing racial bias in law enforcement (Ferdik *et al.*, 2013; Faber and Kalbfeld, 2019), one potential problem that has not been subject to rigorous empirical analysis is the possibility of an adverse reaction by the police to receiving civilian complaints. This possibility leads us to the primary questions motivating this research: (1) How do the police respond to receiving civilian complaints, and (2) does civilian race factor into such response? Extant research demonstrates that racial minorities are more likely to be subjected to the use-of-force by police (Smith and Petrocelli, 2001; Terrill and Mastrofski, 2002; Engel and Calnon, 2004), that collective protest by Black citizens is more likely to be met with police presence and violence (Davenport *et al.*, 2011), and that violence against police officers (e.g., fatal shooting) is more likely to result in retaliatory policing when the perpetrator is Black (Legewie, 2016). However, what has yet to be empirically investigated is if civilian complaint about police misconduct by racial minorities results in retaliatory policing.

We address this gap in the literature by analyzing police behavior in New York City. Using a vector autoregression approach, we examine the dynamic relationship between monthly changes in complaints filed with the Civilian Complaint Review Board and monthly changes in New York Police Department's Stop, Question, Frisk incidents. Across various model specifications, the results reveal racial disparities in how the police respond to increases in civilian complaints. We find that increases in complaints filed by Black civilians are associated with subsequent increases in police stops of Black civilians. This adverse effect of complaint by Black civilians is concentrated in heavily Black precincts in Brooklyn and, to a lesser extent, the Bronx, which possessed the largest Black populations during our study period.

2. Race and police response to civilian complaints

Several bodies of literature provide a basis for the expectation that the police may react adversely to receiving civilian complaints. Most generally, research on protest and state repression finds that civilian protest can trigger state repression (Davenport, 1995; Carey, 2006; Ritter and Conrad, 2016) and that such repression is often enacted by police forces (Earl, 2003; Ahnen, 2007; Davenport, 2007). Research on police culture indicates that many police forces possess a deviant subculture characterized by a "code of silence," whereby officers refuse to report misconduct by fellow officers (Johnson, 2006; Lee *et al.*, 2013) and retaliate against those who do report such misconduct (Skolnick and Fyfe, 1993; Chin and Wells, 1997; Cancino and Enriquez, 2004). Adding to this, research on "police personality" indicates that authoritarian aggression and the desire to maintain or enhance social hierarchy are prevalent traits among police officers (Austin *et al.*, 1987; Haley and Sidanius, 2005) and correlated with an officer's use of force against civilians (Swencionis *et al.*, 2021). Finally, research on police behavior finds that officer behavior is highly responsive to civilian behavior, with officers more likely to be disrespectful and use force in response to perceived disrespect, antagonism, or challenges to their authority (Chevigny, 1969; Garner *et al.*, 1995; Worden, 1995; Terrill, 2003). On this point, recent research finds that, when their peers are injured by civilians in the line of duty, officers are more likely to use force against and injure the civilians they encounter in the following week (Holz *et al.*, 2023). Taken together, this literature provides a foundation for the expectation that police officers may have an adverse response to receiving civilian complaint.

Focusing on the role that race may play in shaping police response to civilian complaint, several bodies of literature provide a foundation for the expectation that officers may react adversely to receiving complaints from members of racial minority groups. First, literature on group conflict and racial threat (Blalock, 1967; Jackson, 1993) argues that dominant racial groups, such as White Americans, will leverage instruments of the state, such as police forces and the criminal justice system, to engage in social control of subordinate racial groups perceived to be threatening their power and group dominance. Paralleling this literature is research on systemic racism

(Weaver, 2007; Feagin, 2013), which begins with the observation that the criminal justice system in the United States is controlled by Whites. While there is some evidence that Black citizens play a role in policing their own communities through their support for punitive law enforcement policies (Fortner, 2015) and as agents of government (Forman, 2017), there is considerable evidence that police forces in the United States were historically (Williams, 2015) and are presently (Ashkenas and Park, 2004; Dewan, 2014) controlled by Whites. Building on this, the systemic racism framework argues that the criminal justice system is used to suppress collective efforts by African Americans to challenge racial inequality and White supremacy. According to this framework, police contact with Black civilians represents “hierarchical interaction” between an oppressor and an oppressed group. As such, this framework predicts that Whites, and the institutions they control, will react “swiftly and harshly” (Feagin, 2013) to behaviors by Blacks that resist or challenge White authority.

Complementing the group threat and systemic racism literature is research on bureaucratic accountability evaluating the efficacy of citizen-led action to improve the performance of government agencies, known as social accountability strategies (Joshi and Houtzager, 2012; Fox, 2015). This literature argues that social accountability strategies will be most effective when the interests of citizens are able to be aligned with those of state actors capable of constraining the behavior of a targeted agency. Given the importance of socioeconomic resources in accessing government officials (Kalla and Broockman, 2016), one implication of this argument is that lower status citizens lacking such resources (e.g., racial minorities) may be less able to access government officials to lobby for intercession. This implication is substantiated by field experiments demonstrating that a range of elected officials and bureaucratic agents are more responsive to Whites than racial minorities (Costa, 2017; Einstein and Glick, 2017). When applied to citizen oversight of law enforcement (Ali and Pirog, 2019), this literature suggests that higher status citizens, such as Whites, may be better positioned than marginalized citizens, such as racial minorities, to use civilian oversight to constrain police behavior.

Finally, intergroup emotions theory (Mackie *et al.*, 2000, 2008) argues that the relative social status of an outgroup shapes emotional reactions to threats by the outgroup. These emotional reactions, in turn, yield distinctive behavioral response tendencies, such as aggression or avoidance. One relevant prediction of this theory is that threats posed by members of lower status groups will trigger anger, which in turn can lead to aggressive action intended to harm the source of the threat. This research suggests that, when focusing on a racialized policing environment with hierarchical relations between White officers and non-White civilians, complaints filed by non-White civilians against White officers could trigger anger and the desire to act against the complainant.

Taken together, these bodies of literature provide a foundation for the expectation that police aggression in response to being the target of civilian complaint should be greatest when the complainants are racial minorities. Thus, in addition to over-policing communities of color and establishing a stronger presence at Black protest events, law enforcement may also react in a discriminatory manner when encountering formalized “pushback” (e.g., civilian complaints about police misconduct) from policed communities. According to the theoretical frameworks advanced in the bodies of literature discussed above, retaliatory increases in policing (e.g., police stops, searches, ticketing) may serve as a reassertion of authority over a presumed subordinate group that is perceived to be challenging the racial hierarchy and to suppress future “claims making” efforts by such subordinate groups (Davenport *et al.*, 2011; Feagin, 2013). In sum, the reviewed literature suggests that complaints filed by White civilians may function to constrain subsequent police behavior, while complaints filed by non-White civilians (e.g., Blacks and Latinos) could result in subsequent increases in police-initiated enforcement-oriented interactions.

3. Data and methods

We empirically explore these expectations using the case of civilian complaint and police behavior in New York City (NYC). NYC is the largest American city, the New York Police Department

(NYPD) is the largest police force in the country, and the Civilian Complaint Review Board (CCRB) of NYC is one of the oldest and largest COAs in operation. NYC represents “ground zero” in the development and implementation of the types of aggressive policing systems that have spread throughout the nation and come to stand at the center of public discourse over racially biased policing (Fagan and Davies, 2000; Weaver and Lerman, 2010). Prominent policing programs in NYC have been found to contain substantial racial bias, such as the “vertical patrols” of public housing under the trespass abatement program (Fagan *et al.*, 2012) and marijuana arrests under the zero-tolerance program (Golub *et al.*, 2007). In addition, NYC is a notable example of historic and present control by Whites over a police force, as reports by the NYPD’s records and management system document a persistent racial gap between the police and the policed in NYC (Foley *et al.*, 2008; Kelly, 2009; O’Neill, 2019). These reports demonstrate that non-Latino Whites are chronically more prevalent among rank-and-file officers, and drastically more prevalent among command staff, than in the city as a whole. In short, NYC is an important center of gravity for our research questions and a prime case for exploring the repercussions of civilian complaint about police misconduct. While these characteristics of NYC make it an important test case, they also may make it difficult to extrapolate observed findings to other cities lacking similar characteristics. This point of caution aside, our use of a single large city to test our hypothesis is consistent with leading policing research focusing on NYC (Legewie, 2016; Sullivan and O’Keeffe, 2017; Mummolo, 2018), Chicago (Ba *et al.*, 2021; Holz *et al.*, 2023), and other cities (Voigt *et al.*, 2017; Hoekstra and Sloan, 2022) as test cases.

The analysis that follows explores the relationship between civilian complaint and police behavior within the context of the NYPD’s Stop, Question and Frisk (SQF) program. Leading scholarship conceptualizes police stops under the SQF program as largely “disciplinary” in nature (Weaver and Lerman, 2010) in that the majority of stops were unproductive in uncovering contraband (White and Fradella, 2016) or reducing crime (Rosenfeld and Fornango, 2014; MacDonald *et al.*, 2016) and a substantial portion failed to meet “reasonable suspicion” standards (Fagan and Davies, 2000). Moreover, the SQF program was fraught with evidence of racial bias (Gelman *et al.*, 2007; Center For Constitutional Rights, 2012) and, as a result, was ruled unconstitutional in 2013 by a federal judge in *Floyd v. City of New York* and was significantly down-scaled afterward.

The main dependent variable in our analysis is the monthly number of SQF incidents between 2003 and 2014, which is publicly available from the New York City Police Department.¹ The original data set contains individual entries for each suspect involved in a police stop and includes information about the location, time, reasons for the stop, as well as basic characteristics of the suspect (gender, race, etc.). We aggregate the number of stops on a monthly basis since our main predictor—civilian complaints—is only available on a monthly basis (see discussion below). Notwithstanding these limitations, monthly aggregates provide a sufficient number of time points (and incidents per time unit) to model their dynamic relationship (see Appendix A for more information). Moreover, as we discuss further below, month is an appropriate time unit for observing police response to changes in civilian complaint for a variety of reasons (e.g., officer awareness of complaints). While our initial analysis will focus on aggregate stops and complaints across NYC as a whole, we will also report results for smaller geographical areas such as boroughs and police precincts. This approach to testing our research questions by focusing on a time series within a single city and policing jurisdiction is common in the literature (Legewie, 2016; Sullivan and O’Keeffe, 2017; Mummolo, 2018).

To construct our independent variable, we turn to data provided by the CCRB—an independent investigative agency of the NYC municipal government. The CCRB is empowered to receive and investigate complaints by civilians against NYPD officers and their staff is comprised entirely of civilians.² The CCRB releases monthly and annual reports to the public that provide

¹<https://www1.nyc.gov/site/nypd/stats/reports-analysis/stopfrisk.page>

²For more information about the CCRB, including its history and main functions, see the CCRB official website: <http://www1.nyc.gov/site/ccrb/index.page>

descriptive statistics on received complaints, the types of misconduct alleged in complaints, and the final disposition (i.e., decision) of CCRB investigations. The CCRB does not make their complete database publicly available; however, different pieces of the data can be found archived online through various databases. We obtained a redacted copy of the entire CCRB database—that is, data for every single complaint filed with the CCRB since they began electronic documentation in 1993—through a request submitted to the Deputy Executive Director for Policy and Strategic Initiatives of the CCRB, which was granted in early 2015. Following the granting of our initial data request, we submitted additional Freedom of Information Law (FOIL) requests to the CCRB for sensitive geographic information concerning each filed civilian complaint—namely, the NYPD precinct linked to the allegation and target NYPD officer. These additional data requests were granted in the summer of 2021.

The CCRB database consists of two separate parts. The first part contains information about every allegation filed with the CCRB, such as location (borough and police precinct), time, type of incident/complaint, the board disposition, and information about the officer involved (e.g., ethnicity, gender). The second part of the database contains further demographic information about the complainants, victims, and witnesses related to a given complaint. It is important to note here that any information that can be used to uniquely identify individual officers receiving complaints or individual complainants was redacted from the data prior to release by the CCRB. This includes the redaction of the exact day of the incident and complaint, which implies that we can only aggregate complaints on a monthly basis. Furthermore, both parts of the CCRB database can only be linked on the basis of unique complaint IDs. Any complaint ID, however, may involve multiple officers, multiple complainants and witnesses, and multiple allegations (e.g., racial slur, beat, gun pointed, etc.) associated with a instance of complaint. Additional information that would allow for a more fine-grained match of individual complainants to allegations about specific officers within a complaint ID were redacted and therefore unavailable.

In order to prepare these data for further analyses, it is necessary to link both parts of the CCRB database based on the unique complaint IDs. For each allegation in the officer data, we selected the first victim on record in the complainant data and matched their demographic characteristics to the respective complaint ID (i.e., ignoring demographic characteristics of witnesses or individuals who filed a complaint on behalf of someone else). Alternative matching procedures do not change the substantive results reported below (see Appendix B.1). Each observation in the merged data set therefore consists of a single allegation and includes information about the officer as well as the alleged victim of misconduct. Since the day of the incident and complaint has been redacted and only the respective month and year are included in the data, we aggregate the number of complaints on a monthly basis.

3.1 The appropriateness of aggregating by month

We complement the administrative complaint data with vital information about CCRB procedures obtained through interviews with CCRB senior administrators, policy and legal analysts, and investigators (see Appendix A.2 for full list). These interviews indicate that month is a plausible unit of time for our analysis since it reflects the approximate bureaucratic lag for officers to become aware of a complaint filed against them.

Once a complaint is received by the CCRB, it is assigned to an individual investigator, who begins by scheduling interviews with complainants and proceeds by scheduling interviews with targeted officers, which typically occurs within two to four weeks from the receipt of a complaint. Individual officers are not contacted directly by CCRB investigators; rather, CCRB investigators contact officers' supervisors and schedule an interview through a system called "appearance control." While there does exist variation in the time frame due to variation in the amount of time required to schedule and complete interviews with complainants and witnesses, interviews with officers are typically set up within two to four weeks of receipt of a complaint, and nearly all

interviews, according to various CCRB administrators, are set up within six weeks at the most. In short, within one month of receipt of a civilian complaint, officers are contacted to schedule an interview with a CCRB investigator, and thus, *are made aware* of being the target of a civilian complaint. When officers are contacted to schedule their interview with the CCRB, it is standard procedure for officers to be given information about the nature of the complaint filed against them, the identity of the complainant, the date and location of the complaint, and the presence and testimony of witnesses involved in the case. Importantly, CCRB staff indicate that the two-to-four week time frame for contacting officers to schedule interviews applies to all complaints regardless of method of filing the complaint (i.e., in-person, online, by telephone, or by fax).

A secondary means by which officers become aware of receipt of complaints is through data reporting routines. The CCRB files monthly reports containing descriptive statistics (e.g., the total number of received complaints, type of complaint, and complaints by police precinct) to the NYPD as part of the CompStat program. Additionally, CCRB staff attend regular CompStat meetings held by the NYPD, which are also attended by precinct level supervisors. Finally, Integrity Control Officers across NYPD precincts have access to the CCRB database, including individual level data on complaints concerning officers under their supervision. Thus, over the time period in our analysis, there was a fluid exchange of information between the CCRB and NYPD.

3.2 Granger causality and the vector autoregression framework

Figure 1 displays an overview of the total number of SQF incidents (in 1000s) and CCRB complaints (in 100s). The number of SQF incidents per month increased over time to a peak of about 60,000 in 2011 and dropped quickly thereafter. Immediately after the court ruling in August 2013, which declared the SQF program unconstitutional, the total number of SQF incidents leveled

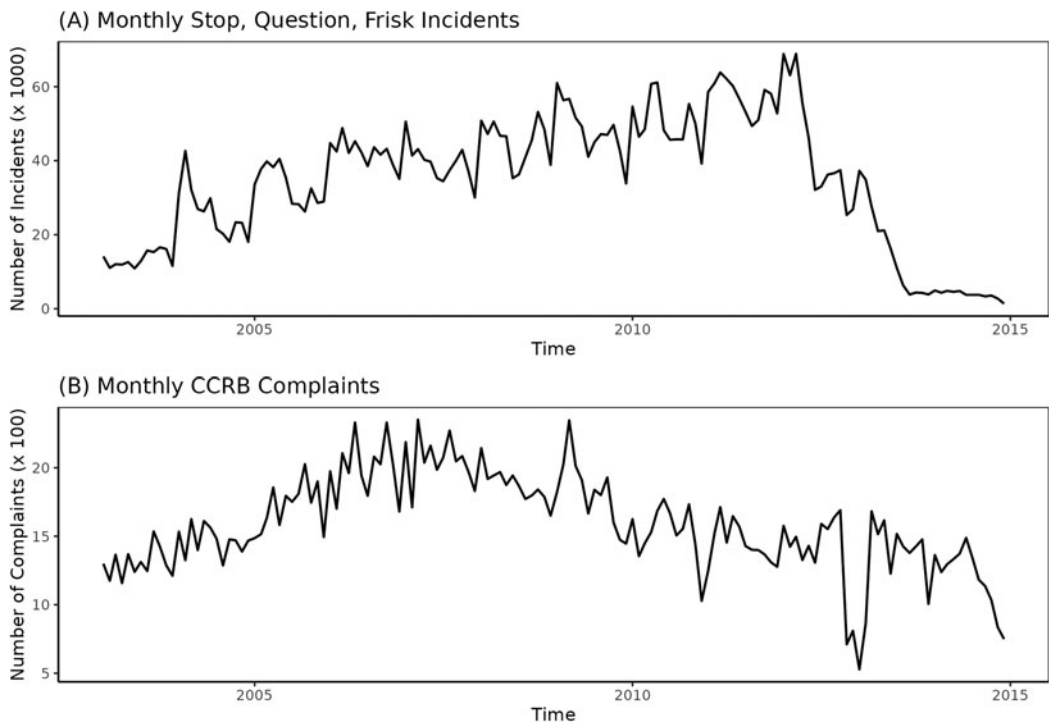


Figure 1. Monthly SQF incidents (A) and CCRB complaints (B) in NYC per month.

around only 4000 per month. The total number of CCRB allegations, on the other hand, increased initially to a maximum of about 2000 per month around 2006, but decreased thereafter. Using these data, we construct time series of monthly *changes* in the number of police stops and civilian complaints about police misconduct. We focus on the differenced series to avoid biases and spurious relationships often encountered when analyzing non-stationary variables (Dickey and Fuller, 1979; Enders, 1995) and to account for potential confounding factors that impact long-term dynamics of stops and complaints. The substantive focus of our article is therefore the reciprocal relationship between short-term changes in CCRB complaints and SQF incidents.

In many applications, such temporal dynamics are examined in terms of their Granger causality (Freeman, 1983; see also Hatemi *et al.*, 2019 for a recent example). In brief, Granger causality describes whether the prediction of a variable at time t based on its previous history prior to time t can be improved by including the previous history of another variable. Consider the following two models:

$$y_t = \alpha + \sum_{p=1}^P \beta_p y_{t-p} + \varepsilon_t \tag{1}$$

$$y_t = \alpha^* + \sum_{p=1}^P \beta_p^* y_{t-p} + \sum_{p=1}^P \delta_p x_{t-p} + \varepsilon_t^* \tag{2}$$

In Equation 1, we predict y_t based only on its own previous values going back up to P time points. In Equation 2, on the other hand, we add previous values of a covariate x to predict y_t . In this example, x is said to Granger cause y if the inclusion of $x_{t-p} \forall p \in \{1, 2, \dots, P\}$ in model 2 improves the prediction of y_t as compared to model 1. A simple way to check for Granger causality is to test against the joint null hypothesis $H_0: \delta_p = 0 \forall p \in \{1, 2, \dots, P\}$. As such, the Granger test shows whether the inclusion of any x_{t-p} improves our prediction of y_t beyond previous values of the dependent variable (i.e., $y_{t-p} \forall p \in \{1, 2, \dots, P\}$).³

We can further extend this model by simultaneously estimating the temporal relationship between y and x in both directions by adding an equation that models x as a function of its own history as well as previous values of y . This basic framework, however, does not account for potential confounding factors affecting both x and y . We can therefore add further exogenous covariates (i.e., not affected by the previous history of x or y) in each equation in order to control for potential confounding factors. Such a framework, which simultaneously captures the interdependencies between multiple time series, is called a vector autoregression model (Freeman *et al.*, 1989; Toda and Phillips, 1994) and can be written as:

$$\vec{y}_t = \vec{\alpha} + \sum_{p=1}^P B_p' \vec{y}_{t-p} + \Gamma' \vec{z}_t + \vec{\varepsilon}_t, \tag{3}$$

where \vec{y}_t is now a vector of outcomes for multiple endogenous time series at time t , $\vec{\alpha}$ denotes a vector of intercepts, B_p is a square matrix of β coefficients capturing the effect of each value of \vec{y}_{t-p} on each element in \vec{y}_t , Γ denotes a matrix of coefficients capturing the effect of additional exogenous covariates \vec{z}_t , and $\vec{\varepsilon}_t$ is a vector of error terms. We use this framework for our main analyses of changes in CCRB complaints and SQF incidents. Thus, elements in B_p denote the effects of previous changes in CCRB complaints at time $t-p$ on its own current values as well as on current changes in SQF incidents. Other elements in the matrix describe the effects of previous changes in SQF incidents at time $t-p$ on its own current changes as well as current

³See Appendix A.3 for an initial analysis using this framework. The results indicate that changes in CCRB complaints about potential police misconduct are predictive of SQF incidents in subsequent months, but not the reverse.

changes in CCRB allegations. Of course, this immediately raises the question of choosing the lag order P in the model. In other words, how far into the past do we have to go to estimate the effect on current values of each time series? For this analysis, we estimate each model with up to 5 lags and report the result that minimizes the model's Akaike information criterion. However, all results presented herein are robust for alternative specifications for P (see Appendixes B.2 and B.3). It is important to keep in mind that the vector autoregression described here models SQF incidents based on CCRB complaints in previous months (and vice versa) without considering potential contemporaneous effects. While non-zero contemporaneous effects between the endogenous variables could affect the estimates, it is reasonable to assume that complaints do not have instantaneous effects on SQF incidents due to the usual timeline to process complaints by the CCRB and the resulting delay in reporting (see discussion above).

Of course, there are potential confounding factors that may jeopardize our inferences about the dynamic relationship between complaints and police stops. For instance, changes in CCRB complaints and SQF incidents may both be correlated with recent changes in crime. Although monthly crime statistics are not available for the entire period of our analysis, we account for this potential confounder by controlling for the monthly change in fingerprintable arrests as a proxy.⁴ The vector \bar{y}_t therefore contains the outcomes of three time series: the monthly change in CCRB complaints, SQF incidents, as well as fingerprintable arrests.

We incorporate a set of additional control variables in \bar{z}_t to capture contemporaneous confounders. Keeping in mind in this context that we are modeling differenced time series, we only have to focus on *short-term* factors that may be associated with monthly *changes* in civilian complaint and police activity (rather than with overall levels). The first potential confounder that falls into this category is news coverage. In order to account for the potential impact of media attention on crime and the police, we control for monthly changes in the number of articles mentioning the NYPD in two local tabloids, the *New York Daily News* as well as the *New York Post*.⁵ Other potential confounders include changes in economic conditions (such as during the Great Recession) and the varying influx of tourists. We therefore include controls for monthly changes in the unemployment rate in NYC as well as the monthly change in the total number of overseas visits to the USA as a proxy for tourism. In addition, we account for the possibility that police patrols may be impacted by changing weather conditions by controlling for monthly mean temperatures and total precipitation in NYC. Lastly, there is a striking regularity in that the police engages in systematically fewer stops in December of every year (on average about 25 percent fewer SQF incidents). To correct for the effect of this pattern on monthly changes in policing, we include a dichotomous indicator for the month of January. We note, however, that the results reported hereafter are robust across a wide range of model specifications and the inclusion of alternative controls.⁶

It is worth emphasizing that the Granger causality framework is—notwithstanding its name—purely predictive and not a causal model in a strict counterfactual sense. In principle, any causal claim has to satisfy three conditions: empirical association, temporal precedence, and the absence

⁴See Appendix B.4 for additional analyses using available statistics on monthly felonies.

⁵We include media attention as an *exogenous* variable in order to account for its potential contemporaneous effect on complaints and because it results in a more parsimonious model specification. In other words, we considered it more likely that changes in media attention on the police may impact complaints in the same month instead of in subsequent months. However, it is also possible that police activity may impact subsequent news media attention. To address this potential concern, we have included an additional robustness check in Appendix B.5 that incorporates media attention as an endogenous variable in the vector autoregression instead. The results remain unchanged. We thank an anonymous reviewer for raising this point.

⁶Furthermore, any time series model such as the vector autoregression described here requires balanced equations to allow for valid statistical inferences (Pickup, 2022; Pickup and Kellstedt, 2023). Appendix A.4 presents additional analyses showing that (1) our focus on *changes* in CCRB complaints and SQF incidents results in a balanced vector autoregression with white noise residuals, and (2) alternative specifications focusing on monthly *totals* would not be balanced due to the fact that total CCRB complaints and total SQF incidents are not cointegrated.

of spuriousness. By itself, establishing Granger causality only provides evidence for the first two conditions of a causal effect (association and precedence) since it shows that lags of one time series are predictive of another series above and beyond its own history.⁷ Absent a stronger design-based identification, we minimize spuriousness in two ways: (1) by focusing on monthly *changes* instead of levels, rendering long-term confounders less of an issue, and (2) by controlling for changes in crime and other potential short-term confounders. Ultimately, however, it remains a theoretical question whether there may be additional *unobserved* confounders that jeopardize our interpretation of the results. While we cannot categorically refute this possibility, we can nevertheless use the vector autoregression framework to assess the plausibility of spuriousness as a potential threat to inference.

There are dangers involved in implying a causal relationship between civilian complaint and police behavior when such relationship does not exist. As such, we note before proceeding that our empirical analyses are observational and cannot directly identify the underlying causal mechanism driving observed relationships. It would be neither feasible nor ethical for researchers to randomly manipulate the level of complaints about police misconduct in order to ascertain their true causal effect on subsequent policing. While we try to rule out potential alternative explanations wherever possible, our analysis can only provide a first step in characterizing the dynamic relationship between aggregate civilian complaint about officer misconduct and subsequent aggregate police behavior under the SQF program. Such descriptive inference is nevertheless important as it provides the necessary groundwork to address potential racial biases in police response to civilian complaints. Rather than simply ending our analysis after presenting the association between civilian complaint and police stops, our strategy is to complement these main findings with a comprehensive set of ancillary placebo and mechanism tests that paint an overall picture of the relationship of civilian complaint to subsequent police behavior that is exceedingly difficult to explain by any parsimonious alternative mechanism and thereby mitigates concerns about spurious relationships.

4. Results

We begin by examining the association between overall changes in civilian complaint and subsequent changes in the total number of police stops. The top section of [Figure 2](#) displays the results. Panel (A) shows the parameter estimates of lagged changes in CCRB complaints predicting subsequent changes in SQF incidents. To facilitate substantive interpretation, panel (B) displays 1000 simulations of monthly SQF incidents after a shock in CCRB allegations in month 0 (holding all other variables at their respective means). Each shock is equivalent to the largest monthly increase observed in CCRB allegations from an average month of the respective series. While the model estimates in panel (A) directly compare the effect of 100 additional CCRB complaints across different groups on subsequent SQF incidents, the simulations in panel (B) provide additional intuition about the monthly changes in complaints that are actually observed in each series as well as the predicted change in stops relative to the average monthly totals. The simulations help contextualize the estimated raw coefficients since the monthly fluctuation in complaints differs across racial groups (e.g., we observe monthly increases in complaints by Blacks of up to 330 while complaints by Whites only increase by up to 100 from one month to the next). At the same time, the baseline number of total SQF incidents varies across racial groups as well (e.g., on average 3500 monthly stops involving White suspects versus 18,000 monthly stops involving Black suspects), so the predicted change should be interpreted relative to this baseline.

In the top right panel of the figure, for instance, we see that following a surge of 820 additional complaints in month 0, the number of predicted SQF incidents increases slightly in the following month (from 35,000 to 38,000 stops) but quickly reverts to its original level. While there appears

⁷See Hatemi *et al.* (2019) for a similar application of the Granger causality framework.

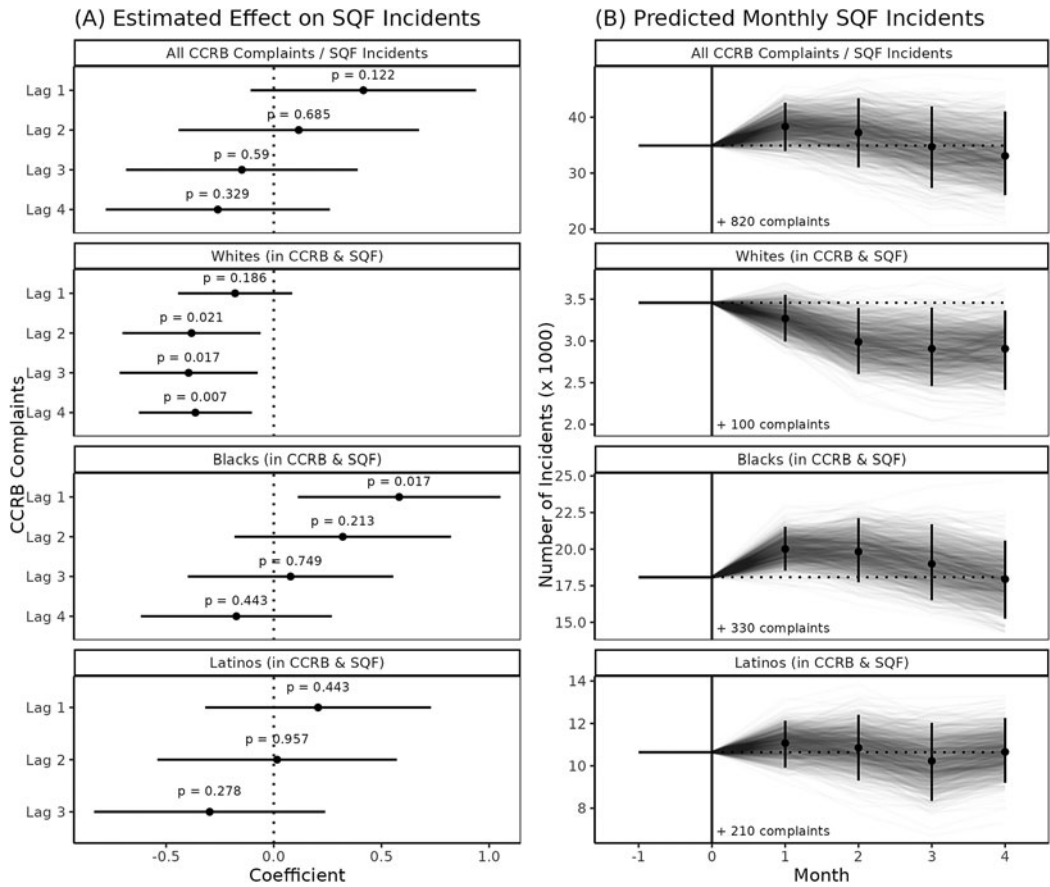


Figure 2. Association between lagged CCRB complaints (monthly change in 100s) and subsequent SQF incidents (monthly change in 1000s). (A) Vector autoregression coefficients with 95 percent confidence intervals. (B) Predicted SQF incidents after CCRB shock in month 0 (increasing average allegations by the maximum monthly growth observed in the original series). Each line represents a single iteration of the simulation scenario. The plot additionally displays expected average levels along with 95 percent confidence intervals. Total number of time points $T = 143$ for each series. Full model results are displayed in Appendix D.1.

to be a tendency of increased SQF activity in the first month after complaints rise, the coefficients do not reach conventional levels of statistical significance. In other words, there is too much uncertainty in our model to ascertain whether changes in *overall* complaints are associated with changes in *overall* police stops. This pattern changes dramatically, however, once we break down both series by racial groups to investigate if the police respond differently to complaints by minority groups—especially Black civilians.

The second row of panels in Figure 2 restricts each time series to incidents involving White complainants in the CCRB data and White suspects in the SQF data. Here, we observe that an increase in complaint is associated with a *decrease* in police activity in subsequent months: a simulated shock in complaints by Whites from an average month (+100 complaints) reduces the average number of SQF incidents involving White suspects from about 3500 to 3000 over the course of four months. What is particularly notable about this result is the duration: an increase in White complaint at time 0 is associated with statistically significant decreases in police stops of White civilians for up to four months (panel (B), second row). This finding comports with existing research suggesting that the privileged status of Whites in the United States should

engender greater ability to use social accountability mechanisms to constrain the behavior of “street-level” bureaucrats, such as the police. More specifically, this finding aligns with past research demonstrating that police officers allot greater respect to White civilians (Voigt *et al.*, 2017) and do not escalate conflict with White civilians following an attack made against a uniformed officer by a White assailant (Legewie, 2016).

The opposite is the case, however, when looking at Black complainants and Black suspects. The results are displayed in the third row of plots in Figure 2. In contrast to the results for Whites, CCRB complaints by Blacks is associated with a statistically significant *increase* in subsequent SQF incidents involving Black civilians. A simulated shock of 330 additional CCRB complaints by Blacks increases the predicted number of SQF incidents in the subsequent month from about 18,000 to 20,000. In other words, an additional complaint by a Black civilian is associated with a predicted increase of more than five additional stops involving Blacks in the following month. In contrast to the findings for Whites, this effect gradually dissipates over time, being most pronounced in the first month following a spike in Black complaint and essentially returning to zero by the fourth month (panel (B), third row). This finding aligns with research demonstrating that police officers speak with greater disrespect to Black civilians (Voigt *et al.*, 2017), are more likely to use force against Black civilians (Knox *et al.*, 2020), and greatly escalate conflict with Black civilians following an attack made against a uniformed officer by a Black assailant (Legewie, 2016).

Looking next at the relationship for Latinos in the last row of Figure 2, we observe no clear association between complaints and police stops. The absence of a significant effect of Latino complaint is notable given that extant research finds that Latinos are subject to racial bias in policing (Fagan and Davies, 2000; Gelman *et al.*, 2007). That said, while Latinos were subject to more stops than non-Latino Whites within the context of SQF in NYC, Black residents were by far the most pronounced target of over-policing (Fagan and Davies, 2000). In keeping with this fact, publicly available CCRB annual reports indicate that, during the heyday of the SQF program, the majority of misconduct allegations against NYPD officers were filed by Black residents. Much of this complaint activity had to do with the SQF program: for example, the 2010 Annual Report by the CCRB indicates that, of complaints about SQF, 66 percent were filed by Black residents compared to 24 percent by Latino residents. Moreover, various aggressive policing programs in NYC beyond SQF, such as vertical patrols of public housing and marijuana misdemeanor arrests, were disproportionately targeted at Black residents relative to Latinos and Whites (Golub *et al.*, 2007; Fagan *et al.*, 2012). In short, the insignificant effect of Latino complaint (compared to the significant effect of Black complaint) may reflect specific features of the racialized law enforcement context in NYC between 2003 and 2014.

In summary, the results reported in this section present troubling correlational evidence of a *differential response* of police behavior to civilian complaint as a function of the race of the civilians complaining about their treatment by the police. After complaints by Whites increase, we observe a noticeable and sustained *contraction* in police activity targeting Whites. However, after complaints by Blacks increase, our model predicts a swift and sizeable *expansion* in police activity targeting Black civilians. Of course, these results are based on observational data that limit any causal claims regarding the nature of this relationship. Nevertheless, such observational evidence of racial bias in police response to complaints about misconduct should be deeply concerning and—at a minimum—warrants further investigation. As a first step in this effort, we conduct several robustness checks in the Appendix to rule out potential alternative explanations of the patterns observed in the data.⁸

⁸Specifically, we demonstrate that the results in Figure 2 are robust to a variety of checks. First, we replicate the analysis while restricting our data to a reduced time series that only considers cases prior to the 2013 ruling rendering NYC’s SQF Program unconstitutional (Appendix B.6). Second, we find the same patterns when focusing only on SQF incidents involving

4.1 Assessing the plausibility of racially biased retaliatory policing

Robust causal claims about the racially biased nature of the relationship between civilian complaints and subsequent policing could be based on randomly manipulated levels of complaints by White and Black civilians, or some exogenous shock thereof that is unrelated to police behavior or any potential confounding variable. Unfortunately, neither is available in the context of our current study. The absence of such designs, however, does not imply that there is no way to further corroborate our findings. In this section, we proceed by deriving and testing falsifiable predictions that would be expected under the presumption of racially-motivated retaliatory policing but that are hard to account for by any alternative explanation or conceivable time-varying confounder. Here, we compare the association between complaints and subsequent stops involving Black civilians conditional on (a) the race of the officers involved in complaints, (b) the officers' awareness of filed complaints, (c) the veracity of the complaint, (d) the officers' level of discretion to perform stops, and (e) the productivity of police stops. As we will show, these additional results provide a veritable pattern of findings that comport with police retaliation as the data-generating process and are extremely difficult to explain by any conceivable time-varying confounder.

To the extent that the observed relationships derive from a racialized policing environment where Black complaints represent “pushback” that is perceived as a threat to White authority, the observed effect of Black complaints should be confined to complaints targeting *White* officers. In other words, presuming that the increase in stops of Black civilians following increases in Black complaint reflects a reassertion of White authority in response to Black claims-making, it implies that our findings should be confined to Black complaints involving White officers. In contrast, observing the same pattern for non-White police officers would lead us to reject retaliatory policing borne out of racial threat to White authority as the mechanism generating our findings and instead would suggest an alternative data-generating process. [Figure 3](#) shows that the relationship is only found for complaints by Black civilians against *White* officers. For complaints against *non-White* officers, we observe relationships that are smaller in magnitude and statistically insignificant. Additional “racial” placebo tests reported in the Appendix indicate that the spike in stops against Blacks is confined to Black complaint against White officers, as increases in White complaint against White officers *does not* lead to subsequent increases in stops against Black civilians (Appendix C.1).

Beyond officer race, a necessary precondition for police retaliation as an explanation for our findings is *officer awareness* of complaints filed against them. Officers receiving complaints are made aware of the complaint through being contacted by CCRB investigators through the NYPD's Appearance Control Unit to schedule an investigative interview. As noted above, standard CCRB procedure is for this contact with subject officers to occur within two to four weeks of receipt of a complaint (see Methods for details). Upon scheduling this interview, officers receive summary information on the misconduct allegation(s), date and time of occurrence, and the name of the complainant(s). However, prior to contacting and interviewing officers, CCRB investigators contact and interview complainants. Critically, a significant number of filed complaints never get to the point where officers are notified of receipt of a complaint due to complainants (1) withdrawing their complaint when contacted by CCRB investigators, (2) being unable to identify the perpetrating officer, or (3) being uncooperative or unavailable for interview. In these cases, the complaint is administratively closed and the precondition for retaliation is not present: officers are not contacted by the CCRB, have no knowledge of the complaint, and have no motive to retaliate. Thus, an important test of retaliation as an explanation for our findings involves demonstrating the *absence* of an effect of complaints when the officer is not contacted by the CCRB.

the use of force (Appendix B.7), indicating that the results are not only driven by minor interactions with police. Third, the results from a temporal placebo test (Appendix B.8) demonstrate that *leads* in Black complaints are not associated with changes in stops of Black civilians. This finding is critical, as it illustrates that police stops of Black residents only responds to preceding, and not future, changes in Black complaint.

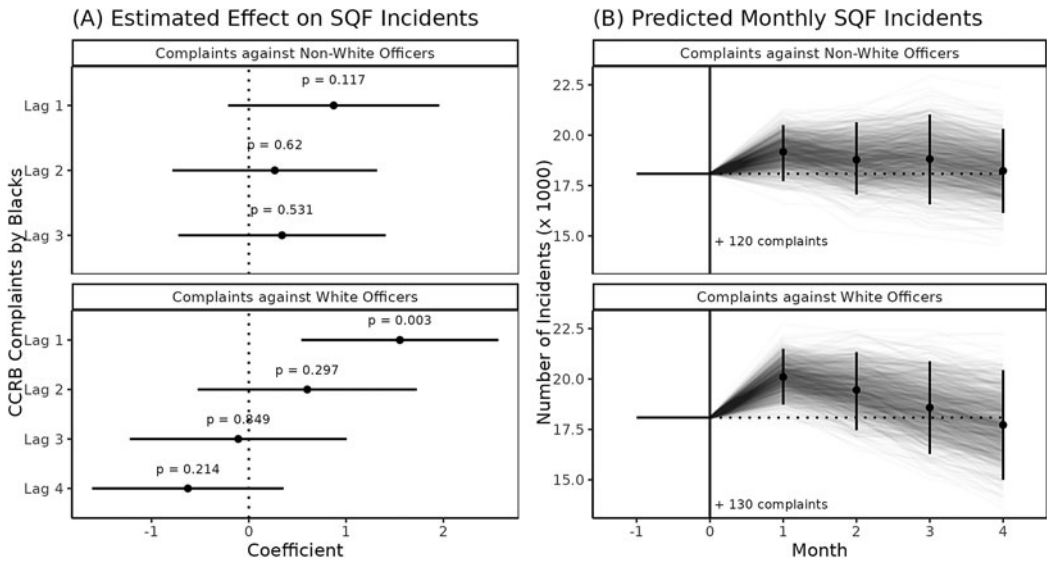


Figure 3. Association between lagged CCRB complaints by Black civilians (monthly change in 100s) and subsequent SQF incidents involving Black suspects (monthly change in 1000s) by officer race. (A) Vector autoregression coefficients with 95 percent confidence intervals. (B) Predicted SQF incidents after CCRB shock in month 0 (increasing average allegations by the maximum monthly growth observed in the original series). Each line represents a single iteration of the simulation scenario. The plot additionally displays expected average levels along with 95 percent confidence intervals. Total number of time points $T=143$ for each series. Full model results are displayed in Appendix D.2.

If complainants complete an interview with CCRB investigators and subject officers are subsequently interviewed, CCRB investigators render a “case disposition” where complaints lacking evidence are deemed unsubstantiated while those possessing sufficient credible evidence are deemed substantiated (Civil Complaint Review Board, 2013). With either outcome, the *precondition for retaliation is met*—officers are contacted and informed about the complaint. However, complaints that are substantiated involve a greater threat to officers because they (a) possess evidence of officer misconduct, and (b) can result in disciplinary action against the officer (e.g., loss of vacation time, charges or prosecution). If the results observed for Black complaints derive from the motivation of officers to punish a perceived lower status group for engaging in “pushback” that threatens their well-being, we should observe stronger relationships for complaints involving a more imminent element of threat. The veracity of complaints (as determined by the CCRB) provide the only available means within the data to gauge the culpability of subject officers, with substantiated complaints implying a scenario where officers engaging in misconduct against a subordinate racial group are “called out” for their behavior.

In summary, we can disaggregate the complaint data by CCRB case disposition to perform a critical “mechanism” placebo test that checks for a null effect of complaints lacking the precondition for retaliation (i.e., complaint withdrawn, officer unidentified, complainant uncooperative/unavailable). Further, we can test for (1) the existence of significant relationships for Black complaints when the precondition is met (i.e., officers are contacted and interviewed), and (2) larger effects for Black complaints accompanied by credible evidence of officer misconduct conveyed during officer contact with CCRB investigators.

The results displayed in the top panel of Figure 4 demonstrate that increases in CCRB allegations that did not involve contacting the officer exert *no effect* on subsequent SQF incidents involving Black civilians. This finding is critical, as it demonstrates that the effect of Black complaint reported in Figure 2 disappears when focusing on complaints filed by Black civilians that are not brought to target officers’ attention. Next, the results displayed in the middle and bottom

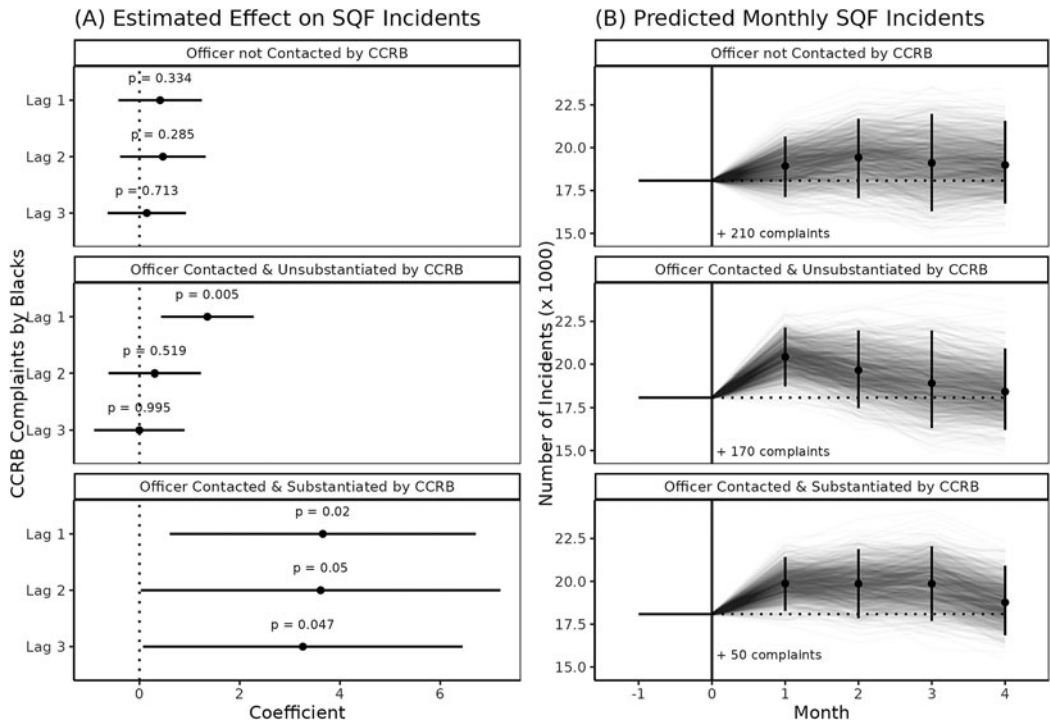


Figure 4. Association between lagged CCRB complaints by Black civilians (monthly change in 100s) and subsequent SQF incidents involving Black suspects (monthly change in 1000s) by CCRB disposition. (A) Vector autoregression coefficients with 95 percent confidence intervals. (B) Predicted SQF incidents after CCRB shock in month 0 (increasing average allegations by the maximum monthly growth observed in the original series). Each line represents a single iteration of the simulation scenario. The plot additionally displays expected average levels along with 95 percent confidence intervals. Total number of time points $T = 143$ for each series. Full model results are displayed in Appendix D.3.

panels of [Figure 4](#) demonstrate significant effects of Black complaints on subsequent stops of Black civilians when restricting the analysis to complaints involving officer contact with the CCRB. Interestingly, even when complaints lack sufficient evidence and are rendered unsubstantiated, we observe a statistically significant effect on subsequent stops of Black civilians; these effects, however, are comparatively modest, as a simulated shock of about 170 additional *unsubstantiated* CCRB complaints by Blacks increases the number of SQF incidents involving Blacks from about 18,000 to more than 20,000. Finally, the bottom of [Figure 4](#) reveals larger and longer lasting effects for Black complaints that are substantiated by the CCRB. Indeed, we observe a comparable increase in simulated SQF incidents involving Black suspects following a shock of only 50 additional *substantiated* CCRB complaints by Blacks. The effect of a single additional substantiated Black complaint is therefore approximately twice as large and more enduring than the effect of a single additional unsubstantiated Black complaint. These findings not only corroborate a required mechanism (i.e., officer awareness of complaints) for police retaliation as an explanation for the finding for Black complaint reported in [Figure 2](#), they also provide suggestive evidence that the *veracity* of complaints filed by Black civilians, and thus the putative culpability of subject officers, factors into police response. The results reported here suggest that complaints filed by Black civilians engender the most pronounced adverse effect when they disclose police misconduct and carry the potential for disciplinary consequences for officers.

One additional requirement in applying a “police retaliation” interpretation for our findings is *officer discretion* to perform a stop. If our findings are due to officers performing an SQF on Black civilians in reprisal for receiving complaints by Black civilians, this type of process would require

the presence of officer discretion to perform said stops. Importantly, many types of stops performed by officers in the SQF data involve the reported presence of factors compelling officer performance of an SQF or circumscribing the target of an SQF. For example, many SQF incidents occur in the context of an ongoing investigation, in response to a report filed by a victim or witness, within a defined vicinity of a crime scene, or in response to officers directly observing criminal activity. Under such conditions, there are *compulsory bases* present for performing an SQF that reduce the element of officer discretion, consequently making it less plausible that such stops could be driven by prior complaints. According to information provided by staff at the CCRB, these compulsory bases for performing a police stop eliminate elements of officer discretion in choosing whether or not to perform a stop, choosing whom to stop, or the feasibility of fabricating a pretext for performing a stop given the ability for the reason to be falsified. In short, stops with compulsory bases should not be responsive to preceding changes in Black complaint; instead, we should only observe increases in stops absent factors compelling officers to perform the stop.

Figure 5 shows the result of several analyses comparing cases where officers have more or less discretion over SQF activity. The left column displays the association between CCRB complaints by Black civilians on subsequent SQF incidents involving Black suspects where external circumstances compelling officers to engage in stops were absent (i.e., no ongoing investigation or report by a victim or witness). In the right column, on the other hand, the corresponding relationships are displayed where documented circumstances surrounding a stop make purely discretionary stops less plausible. For simplicity, we omit the simulated monthly SQF incidents in this figure. Across five different external circumstances that influence an officer's discretion to perform police stops, a strikingly consistent pattern emerges: we only observe significant associations between Black complaints and subsequent stops of Black civilians if officers had more discretion in choosing whether and whom to stop.

Finally, if the observed findings reported in Figure 2 are due to officers performing stops of Black residents in reprisal for receiving complaints by Blacks, there is little reason to expect this increased police activity to be productive in terms of law enforcement outcomes (e.g., discovery of contraband or arrest made). Rather, prior research on racial threat and systemic racism (Lerman and Weaver, 2014) points toward these stops being “disciplinary” in nature, where the purpose may instead be the reassertion of authority over members of a subordinate group. Indeed, research in this vein contends that being stopped in public space represents a public discounting of worth, where the stop conveys to the suspect that “the state has no problem displaying its power and control over the citizen on a public stage” (Fagan *et al.*, 2016: 9). Thus, one final test we can perform to corroborate police retaliation as an explanation for our findings is to estimate the effect of Black complaint on subsequent stops of Black civilians conditional upon the *productivity* of the stops. Here, if the additional stops of Black civilians observed to follow spikes in Black complaint are due to racially motivated police retaliation, we should expect this additional police activity to be unrelated to criminal behavior and unproductive with respect to law enforcement outcomes.

The results presented in Appendix C.2 reveal that increases in Black complaint only exert a significant effect on subsequent stops of Black suspects when focusing on stops that were unproductive in nature (i.e., no contraband uncovered, no arrest made). In contrast, when focusing on productive stops of Black suspects, we fail to observe a significant effect of Black complaint. These results further corroborate a police retaliation interpretation of our main findings by suggesting that prior changes in Black civilian complaint only systematically factor into the data-generating process for subsequent police behavior when focusing on a subset of this behavior that prior research conceptualizes as “disciplinary” in nature. Indeed, monthly changes in *productive* policing of Black residents under the SQF program appears entirely unaffected by Black civilian complaint behavior.

4.2 Geographic disaggregation of complaints and police stops

One remaining concern with our analysis is the city-wide geographic aggregation of our time series. After all, analyzing the association between civilian complaints and police stops on the city

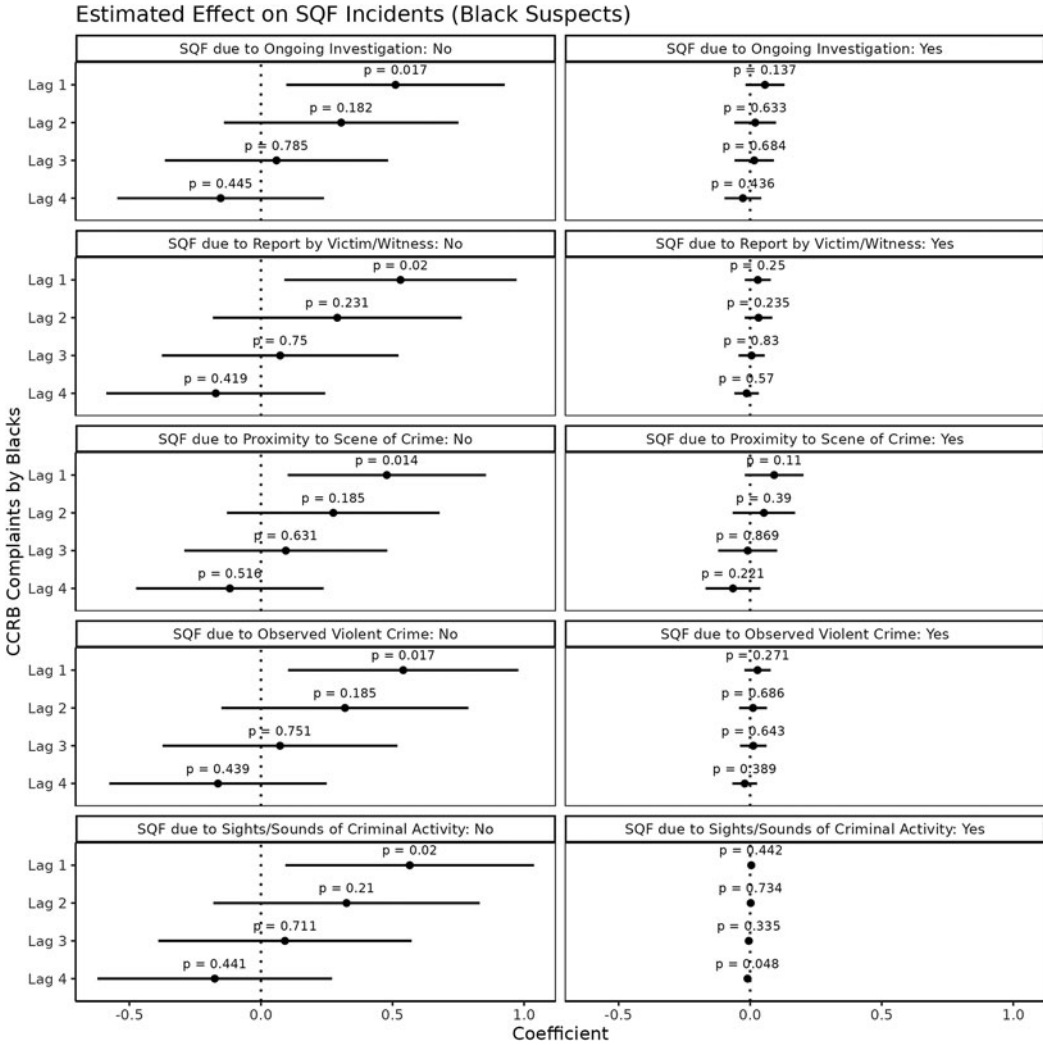


Figure 5. Association between lagged CCRB complaints by Black civilians (monthly change in 100s) and subsequent SQF incidents involving Black suspects (monthly change in 1000s) by documented reason for performing SQF. Vector autoregression coefficients with 95 percent confidence intervals. Total number of time points $T = 143$ for each series. Full model results are displayed in Appendix D.4.

level may obscure important racial disparities that manifest themselves in smaller geographical units. To address this issue, we leverage information on the borough and NYPD precinct housing complaints and police stops in both data sets. Re-estimating the relationship between complaints and stops involving Blacks by borough and policing precinct provides the opportunity to assess whether or not our findings are concentrated in geographic contexts with larger populations of Black civilians. This is a relevant test given that prior research indicates that aggressive policing and officer misconduct are more likely to occur in communities of color (Kane, 2002; Terrill and Reisig, 2003; Fagan *et al.*, 2012) and that dominant racial groups are more likely to experience racial threat in contexts housing larger Black populations (Blalock, 1967). Additionally, demonstrating that our city-level findings hold when disaggregating to police precinct is critical given that the processes we theorize between civilian complaint and subsequent police behavior unfold within smaller policing jurisdictions.

Figure 6 presents the estimated relationship between complaint and stops in each of the five NYC boroughs. Consistent with previous findings on aggressive policing and officer misconduct (Fagan and Davies, 2000; Kane, 2002), we find that the spike in police stops of Black suspects following increases in Black complaint is spatially concentrated in Brooklyn, and to a lesser extent, the Bronx. These findings align with expectations derived from theories of racial threat, as Brooklyn and the Bronx contained the largest Black populations of the five boroughs during the time period under analysis. In addition to being geographically concentrated in certain boroughs, the relationship between complaints and policing should also be *spatially bounded* in the sense that it only emerge *within* a given borough and not spill over between boroughs. This logic provides the opportunity for “spatial” placebo tests since interactions between Black civilians, officers, and CCRB investigators occur within police precincts contained within boroughs. As such,

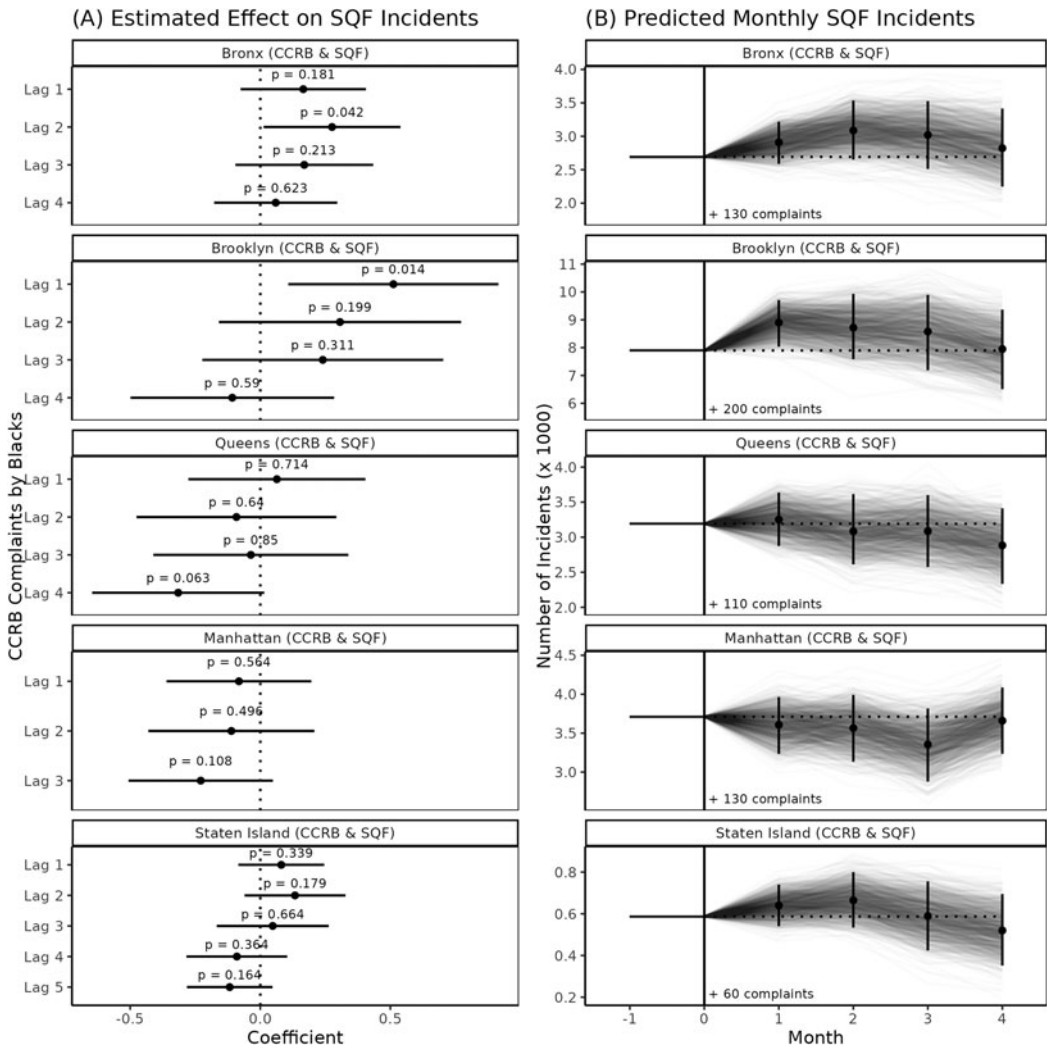


Figure 6. Association between lagged CCRB complaints by Black civilians (monthly change in 100s) and subsequent SQF incidents involving Black suspects (monthly change in 1000s) by borough. (A) Vector autoregression coefficients with 95 percent confidence intervals. (B) Predicted SQF incidents after CCRB shock in month 0 (increasing average allegations by the maximum monthly growth observed in the original series) with 95 percent confidence intervals. Total number of time points $T = 143$ for each series.

complaints filed by Black civilians in one borough should *not affect* police stops in a neighboring borough. As we show in Appendix C.3, these spatial placebo tests indicate that the spike in police stops of Black civilians in Brooklyn following increases in Black complaints in Brooklyn is *not* observed when substituting Black complaints in neighboring boroughs as the independent variable. It is only increases in complaint by Blacks in the *same borough* that appears to lead to a surge in SQF incidents; complaints in neighboring boroughs, on the other hand, have insignificant effects.

Last, we further disaggregate our analysis down to NYPD precinct. In theory, the processes generating our observed results should be driven by precinct-level processes occurring between Black civilians, police officers, and CCRB investigators. Moreover, according to theories of racial threat, these processes should be most pronounced in precincts housing the largest Black populations. Thus, a final geographic check on our results is to assess whether or not our results hold when analyzing the relationship between Black complaint and police stops of Black civilians at the precinct level, with a focus on precincts housing large Black populations. Given the results in Figure 6 indicate that the association between Black complaint and stops of Black civilians is most pronounced in Brooklyn and, to a lesser extent, the Bronx, we focus this analysis on police precincts in Brooklyn and the Bronx. Using census block group data from the 2010 Decennial Census and block group and NYPD precinct shape files, we performed a weighted spatial join to estimate the percent Black in each NYPD precinct. We then split the precincts by median percent Black to isolate the precincts housing the largest Black populations.

Figure 7 presents the relationship between Black complaint and subsequent stops of Black civilians broken down by above and below median precinct Black populations for precincts within the Bronx and Brooklyn. As can be seen, the spike in police stops of Black civilians following an increase in Black complaint is most pronounced in heavily Black police precincts in Brooklyn. In the Bronx, the first lag for Black complaint in above-median percent Black precincts is positive but only marginally significant ($p = 0.08$). These results build on the results by NYC

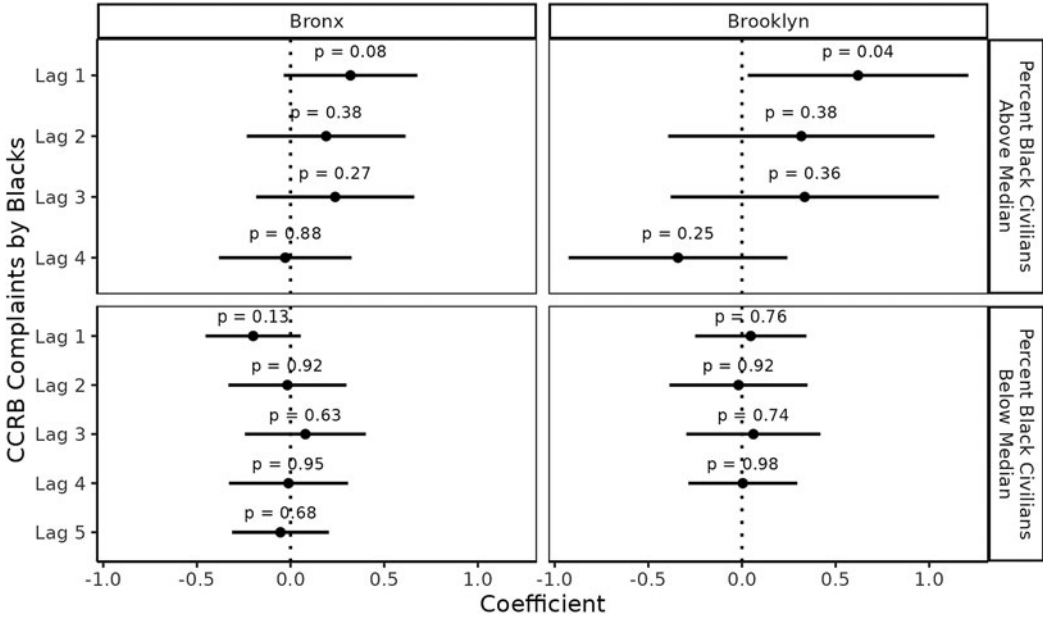


Figure 7. Association between lagged CCRB complaints by Black civilians (monthly change in 100s) and subsequent SQF incidents involving Black suspects (monthly change in 1000s) by police precinct Black population size. Vector autoregression coefficients with 95 percent confidence intervals. Total number of time points $T = 143$ for each series.

borough by demonstrating that the pattern of findings hold when breaking the analysis down to police precinct. Finally, the bottom row of graphs in [Figure 7](#) demonstrate that the observed relationship between Black complaint and subsequent stops of Black civilians in heavily Black police precincts fails to emerge when focusing on precincts with below median Black populations (i.e., below 29 percent in Brooklyn and 33 percent in the Bronx). This finding suggests that increases in precinct-level Black complaint only generate an adverse reaction among precinct police officers in heavily Black communities, where extant theories stipulate the perceived threat posed by Blacks to be the most acute. Additional analyses reported in [Appendix C.4](#) show that these precinct-level results by racial composition are not driven by economic differences between the respective precincts.

In sum, the results from our geographic disaggregation of the data and spatial placebo checks align with expectations derived from racial threat theory and comport with a police retaliation interpretation of our main finding: they reveal that our findings are confined to racialized policing environments where Black “pushback” against the police is theorized to represent the strongest threat to White authority and the prevailing racial hierarchy. Crucially, these results cast serious doubt on alternative explanations that are based on unobserved confounding variables operating on the city-level as a whole—as these should lead to spurious relationships that are neither geographically concentrated nor spatially bounded.

5. Discussion

We begin this section with a discussion of the limitations of our research. First, of primary importance is the need to reiterate that the findings presented in this article are based on observational data, which requires that we refrain from making the claim that the patterns we uncover provide *causal evidence* of police retaliation. Second, limitations in the type of data released to the public do not allow for a more fine-grained analysis. For instance, the data set we have received from the CCRB redacted the exact day of each complaint, which prevents us from replicating our analyses using alternative levels of temporal aggregation (see also [Freeman et al., 1989](#)). That said, the reporting timelines and procedures implemented by the CCRB suggest that aggregation by month is most appropriate. Third, the temporal relationships between SQF incidents and complaints are based on aggregate changes across various geographical units (city-wide, by borough, by precinct), which raises the possibility of ecological fallacies. Unfortunately, the data limitations described above rule out the use of a research design with individual officer as the unit of analysis. In addition to improving the precision of our analysis, possession of individual officer-level data would enable researchers to determine whether the findings reported in this article are due to individual officers receiving complaints by Black civilians subsequently stopping more Black suspects, peer-officers in their precinct stopping more Black suspects on their behalf, or both. In the end, the theories we draw on suggest both processes possible. Moreover, the import of our findings remains strong *regardless* of whether officers receiving complaints, their precinct peers, or both, are responsible for enacting retaliation. Each possibility is alarming and suggests the need for corrective measures to safeguard the civilian complaint process against the possibility of police retaliation.

To address the limitations inherent in our analysis of aggregate observational data, we performed an exhaustive series of tests to determine if the patterns of relationships we uncover support the application of a given interpretation of the findings (i.e., police retaliation). Our findings across a range of robustness, validity, mechanism, and placebo tests are consistent with expectations of retaliatory policing derived from theories of racial threat and systemic racism and support the application of a police retaliation interpretation of the findings. The findings we report mitigate concerns over omitted variable bias, as we are unable to identify any omitted time-varying factor that would explain the countervailing effects of White and Black complaint on police stops of White and Black suspects, as well as our many other divergent findings, such as those by officer race, officer knowledge of complaint, veracity of complaint, discretion to perform a stop,

productivity of stops, and borough and precinct. At a minimum, our results provide alarming evidence of divergence in police behavior following complaints by White and Black civilians. Aside from multiple tests providing converging suggestive evidence of retaliation, the feasibility of our findings were corroborated by staff at the CCRB, who indicated that the findings from our analyses comport with their personal observations and experiences. Indeed, communications with CCRB staff drew our attention to the presence of hundreds of cases of complaints about police retaliation in the CCRB data and published CCRB reports documenting fear of police retaliation as a reason why alleged victims often do not file complaints against NYPD officers or withdraw complaints after filing them.

The results from this analysis suggest that during the height of the SQF program in NYC—a program yielding extensive evidence of racial bias in its execution—Black residents collectively experienced a seemingly retaliatory spike in policing in response to filing civilian complaints about their treatment by the police. In contrast, our results depict the opposite process for White residents of NYC, where usage of the civilian complaint process appears to have meaningfully constrained police behavior toward White residents. While ample scholarship documents racial bias in policing, the findings reported in this article make a vital contribution to the literature by documenting the repercussions faced by Black Americans for exercising their “voice” (Hirschman, 1970) in opposition to the police. In addition to being subject to over-policing under the SQF program, our findings suggest that when Black residents in NYC attempted to register pushback against the police through an institutionalized process of conveying grievances, the result is that they were subjected to *even more* policing.

A primary implication of our findings is that policy reforms aimed at reducing racial bias in policing should give attention to the possibility of “cobra effects,” whereby interventions designed to solve a problem end up exacerbating the problem. COAs are an institutional innovation whose inauguration holds the promise to increase the social accountability of the police, deter officer misconduct, and reduce racial bias in policing outcomes (Ali and Pirog, 2019). However, the results from our analyses suggest that, in order to fully realize the potential benefits of their establishment, the creation of a COA should be accompanied by safeguards to protect against retaliatory policing in response to civilian complaint—especially complaint originating in communities of color. Such safeguard could range from required training designed to deter retaliation, required follow-up behavioral reviews for officers receiving complaints, and the implementation of sanctions for retaliatory behavior.

A secondary—and related—implication concerns the data access policies of police agencies and their associated COAs. Throughout this article, we highlighted various limitations of our analyses that could be (partially) addressed with more fine-grained data about individual complaints (e.g., exact timing and location). We submitted several Freedom of Information requests for access to more detailed records, none of which were granted at the time. While the data we have access to therefore limits our ability to further corroborate the evidence for racial bias in response to civilian complaint, we hope that our results illustrate the importance of expanding transparency in policing and thereby bolsters future requests for data access.

The findings presented in this article yield myriad directions for future research. First, future research could work to further corroborate a “police retaliation” interpretation of our findings through legal requests for further disaggregated data. While the CCRB and NYPD do not release officer-identified data enabling analysis of the behavior of individual officers receiving complaints, the findings presented here may motivate stakeholders to submit requests for such data. Second, future research could endeavor to identify exogenous shocks in civilian complaint akin to work analyzing random changes in police behavior (Sullivan and O’Keeffe, 2017), as such events create the opportunity to observe the causal effect of changes in civilian complaint on police behavior. One challenge here, however, is that police behavior is such an integral feature of the data-generating process for civilian complaint that it is exceedingly difficult to locate an event that caused a large change in civilian complaint but did not also affect police behavior.

While extant research has identified natural experiments involving exogenous shocks to police behavior (Legewie, 2016; Sullivan and O’Keeffe, 2017), we were unable to locate any published research identifying equivalent random events causing changes in civilian complaint behavior.

Our own investigation into random events causing changes in civilian complaint flows led us to identify Hurricane Sandy in late 2012 as a possible shock. According to the CCRB 2013 Annual Report, the hurricane flooded the building housing the CCRB, which resulted in the agency’s loss of electricity, phone service, computer servers, and access to its physical space. The 2013 Annual Report shows that civilian complaints experienced a precipitous drop in the four months following the hurricane until March 2013 when all agency operations were completely restored. The effects of the hurricane, unfortunately, were not confined to the operations of the CCRB or civilian complaint and instead had disruptive effects throughout the city, with the most relevant being its effects on crime and policing. Several reports exist, in addition to data provided by the NYPD, that specific types of crime increased in the wake of the storm, such as looting and burglary (Parascandola and Jacobs, 2012). In response, there was an immediate intensification of police presence in areas affected by the storm, which included the addition of street patrols and light towers intended to deter crime (Gibbs and Holloway, 2013). In short, Hurricane Sandy seemed promising as an exogenous shock to civilian complaint but further investigation into the case renders it ineligible due to its concomitant effects on crime and policing. While past research, as well as our own, were unable to locate exogenous shocks to civilian complaint flows, we see value in future research pursuing such shocks as a means of testing the causal effect of civilian complaint on police behavior.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/psrm.2023.49>. To obtain replication material for this article, <https://doi.org/10.7910/DVN/O1G1KO>.

Competing interest. None.

References

- Ahnen RE (2007) The politics of police violence in democratic Brazil. *Latin American Politics and Society* 49, 141–164.
- Ali MU and Pirog M (2019) Social accountability and institutional change: the case of citizen oversight of police. *Public Administration Review* 79, 411–426.
- Ashkenas J and Park Ha (2004) *The Race Gap in America’s Police Departments*. The New York Times. <http://www.nytimes.com/interactive/2014/09/03/us/the-race-gap-in-americas-police-departments.html?r=2>.
- Austin TL, Hale DC and Ramsey LJ (1987) The effect of layoff on police authoritarianism. *Criminal Justice and Behavior* 14, 194–210.
- Ba BA, Mummolo J and Rivera R (2021) The role of officer race and gender in police–civilian interactions in Chicago. *Science* 371, 696–702.
- Baker A (2018) Confronting implicit bias in the New York Police Department. The New York Times. Last accessed 13 January 2020. <https://www.nytimes.com/2018/07/15/nyregion/bias-training-police.html>.
- Blalock HM (1967) *Toward a Theory of Minority-Group Relations*. New York: Wiley.
- Brereton D. (2000) Evaluating the performance of external oversight bodies. In Goldsmith A and Lewis C (eds). *Civilian Oversight of Policing: Governance, Democracy and Human Rights*. Hart Publishing, pp. 105–124.
- Cancino JM and Enriquez R (2004) A qualitative analysis of officer peer retaliation: preserving the police culture. *Policing: An International Journal of Police Strategies & Management* 27, 320–340.
- Carey SC (2006) The dynamic relationship between protest and repression. *Political Research Quarterly* 59, 1–11.
- Center For Constitutional Rights (2012) Stop and Frisk: The Human Impact. <https://ccrjustice.org/sites/default/files/attach/2015/08/the-human-impact-report.pdf>.
- Chevigny P (1969) *Police Power: Police Abuses in New York City*. New York: Pantheon Books.
- Chin GJ and Wells SC (1997) The blue wall of silence as evidence of bias and motive to lie: a new approach to police perjury. *University of Pittsburgh Law Review* 59, 233.
- Civil Complaint Review Board (2013) Annual Report. <https://www1.nyc.gov/site/ccrb/policy/annual-bi-annual-reports.page>.
- Costa M (2017) How responsive are political elites? A meta-analysis of experiments on public officials. *Journal of Experimental Political Science* 4, 241–254.
- Davenport C (1995) Multi-dimensional threat perception and state repression: an inquiry into why states apply negative sanctions. *American Journal of Political Science* 39, 683–713.

- Davenport C** (2007) State repression and political order. *Annual Review of Political Science* **10**, 1–23.
- Davenport C, Soule SA and Armstrong DA** (2011) Protesting while Black? The differential policing of American activism, 1960 to 1990. *American Sociological Review* **76**, 152–178.
- Dewan S** (2014) *Mostly White Forces in Mostly Black Towns: Police Struggle for Racial Diversity*. The New York Times. <https://www.nytimes.com/2014/09/10/us/for-small-police-departments-increasing-diversity-is-a-struggle.html>. Last accessed 13 January 2020.
- Dickey DA and Fuller WA** (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**, 427–431.
- DOJ** (2016) Advancing Diversity in Law Enforcement. Report by Equal Employment Opportunity Commission. Last accessed 13 January 2020. <https://www.justice.gov/crt/policediversity>.
- Earl J** (2003) Tanks, tear gas, and taxes: toward a theory of movement repression. *Sociological Theory* **21**, 44–68.
- Einstein KL and Glick DM** (2017) Does race affect access to government services? An experiment exploring street-level bureaucrats and access to public housing. *American Journal of Political Science* **61**, 100–116.
- Enders W** (1995) *Applied Econometric Time Series*. 1st edn., New York: John Wiley & Sons.
- Engel RS and Calnon JM** (2004) Examining the influence of drivers' characteristics during traffic stops with police: results from a national survey. *Justice Quarterly* **21**, 49–90.
- Faber JW and Kalbfeld JR** (2019) Complaining while Black: racial disparities in the adjudication of complaints against the police. *City & Community* **18**, 1028–1067.
- Fagan J and Davies G** (2000) Street stops and broken windows: terry, race, and disorder in New York City. *Fordham Urban Law Journal* **28**, 457.
- Fagan J, Davies G and Carlis A** (2012) Race and selective enforcement in public housing. *Journal of Empirical Legal Studies* **9**, 697–728.
- Fagan J, Tyler TR and Meares TL** (2016) Street stops and police legitimacy in New York. In Thierry Delpuech and Jacqueline E. Ross (eds), *Comparing the Democratic Governance of Police Intelligence: New Models of Participation and Expertise in the United States and Europe*. Edward Elgar Publishing, pp. 203–231.
- Fagin J** (2013) *Systemic Racism: A Theory of Oppression*. New York: Routledge.
- Ferdik FV, Rojek J and Alpert GP** (2013) Citizen oversight in the United States and Canada: an overview. *Police Practice and Research* **14**, 104–116.
- Foley PF, Guarneri C and Kelly ME** (2008) Reasons for choosing a police career: changes over two decades. *International Journal of Police Science & Management* **10**, 2–8.
- Forman J Jr** (2017) *Locking Up Our Own: Crime and Punishment in Black America*. Farrar, Straus and Giroux.
- Fortner MJ** (2015) *Black Silent Majority: The Rockefeller Drug Laws and the Politics of Punishment*. Cambridge, MA: Harvard University Press.
- Fox JA** (2015) Social accountability: what does the evidence really say?. *World Development* **72**, 346–361.
- Freeman JR** (1983) Granger causality and the times series analysis of political relationships. *American Journal of Political Science* **27**, 327–358.
- Freeman JR, Williams JT and Lin T** (1989) Vector autoregression and the study of politics. *American Journal of Political Science* **33**, 842–877.
- Garner JH, Schade T, Hepburn J and Buchanan J** (1995) Measuring the continuum of force used by and against the police. *Criminal Justice Review* **20**, 146–168.
- Gelman A, Fagan J and Kiss A** (2007) An analysis of the New York City police department's "stop-and-frisk" policy in the context of claims of racial bias. *Journal of the American Statistical Association* **102**, 813–823.
- Gibbs LI and Holloway CF** (2013) Hurricane Sandy After Action: Report and Recommendations to Mayor Michael R. Bloomberg. <https://tinyurl.com/qxscxz>. https://www1.nyc.gov/assets/housingrecovery/downloads/pdf/2017/sandy_aar_5-2-13.pdf.
- Golub A, Johnson BD and Dunlap E** (2007) The race/ethnicity disparity in misdemeanor marijuana arrests in New York City. *Criminology & Public Policy* **6**, 131–164.
- Haas B** (2013) Tenn. Police chief screening for racists. USA Today, <https://tinyurl.com/rfhy7gr>. Last accessed 13 January 2020. <https://eu.usatoday.com/story/news/nation/2013/01/24/police-department-racism/1862179/>.
- Haley H and Sidanius J** (2005) Person–organization congruence and the maintenance of group-based social hierarchy: a social dominance perspective. *Group Processes & Intergroup Relations* **8**, 187–203.
- Hatemi PK, Crabtree C and Smith KB** (2019) Ideology justifies morality: political beliefs predict moral foundations. *American Journal of Political Science* **63**, 788–806.
- Hirschman AO** (1970) *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States*. Vol. **25**. Cambridge, MA: Harvard university press.
- Hoekstra M and Sloan CW** (2022) Does race matter for police use of force? Evidence from 911 calls. *American Economic Review* **112**, 827–860.
- Holz JE, Rivera RG and Ba BA** (2023) Peer effects in police use of force. *American Economic Journal: Economic Policy* **15**, 256–291.
- Jackson JW** (1993) Realistic group conflict theory: a review and evaluation of the theoretical and empirical literature. *The Psychological Record* **43**, 395.

- Johnson RA** (2006) Whistleblowing and the police. *Rutgers Journal of Law & Public Policy* 3, 74.
- Joshi A and Houtzager PP** (2012) Widgets or watchdogs? Conceptual explorations in social accountability. *Public Management Review* 14, 145–162.
- Kalla JL and Brookman DE** (2016) Campaign contributions facilitate access to congressional officials: a randomized field experiment. *American Journal of Political Science* 60, 545–558.
- Kane RJ** (2002) The social ecology of police misconduct. *Criminology; an Interdisciplinary Journal* 40, 867–896.
- Kelly RW** (2009) Crime and Enforcement Activity in New York City (1 January–31 December 2008). New York City Police Department Report. (<https://tinyurl.com/vwvjg2b>). https://www1.nyc.gov/assets/nypd/downloads/pdf/analysis_and_planning/YearEnd2008EnforcementReport.pdf.
- Kim J** (2022) The effect of civilian oversight on police organizational performance: a quasi-experimental study. *The American Review of Public Administration* 52, 256–291.
- Kluger J** (2016) There's a Test That May Reveal Racial Bias in Police—and in All of Us. Time <https://time.com/4398505/implicit-association-racism-test/>. Last accessed 13 January 2020. <https://time.com/4398505/implicit-association-racism-test/>.
- Knox D, Lowe W and Mummolo J** (2020) Administrative records mask racially biased policing. *American Political Science Review* 114, 619–637.
- Lee H, Lim H, Moore DD and Kim J** (2013) How police organizational structure correlates with frontline officers' attitudes toward corruption: a multilevel model. *Police Practice and Research* 14, 386–401.
- Legewie J** (2016) Racial profiling and use of force in police stops: how local events trigger periods of increased discrimination. *American Journal of Sociology* 122, 379–424.
- Lerman AE and Weaver VM** (2014) *Arresting Citizenship: The Democratic Consequences of American Crime Control*. Chicago, IL: University of Chicago Press.
- MacDonald J, Fagan J and Geller A** (2016) The effects of local police surges on crime and arrests in New York City. *PLoS ONE* 11, e0157223.
- Mackie DM, Devos T and Smith ER** (2000) Intergroup emotions: explaining offensive action tendencies in an intergroup context. *Journal of Personality and Social Psychology* 79, 602.
- Mackie DM, Smith ER and Ray DG** (2008) Intergroup emotions and intergroup relations. *Social and Personality Psychology Compass* 2, 1866–1880.
- Mummolo J** (2018) Modern police tactics, police–citizen interactions, and the prospects for reform. *The Journal of Politics* 80, 1–15.
- O'Neill JP** (2019) Crime and Enforcement Activity in New York City (1 January–31 December 2018). New York City Police Department Report. (<https://tinyurl.com/y5dt5hpb>). https://www1.nyc.gov/assets/nypd/downloads/pdf/analysis_and_planning/year-end-2018-enforcement-report.pdf.
- Parascandola R and Jacobs S** (2012) Hurricane Sandy drives down major crimes in New York City—but burglaries surge. *New York Daily News* <https://www.nydailynews.com/new-york/burglaries-surge-hurricane-sandy-article-1.1196245>.
- Pickup M** (2022) Equation balance in time series analysis: lessons learned and lessons needed. *Political Science Research and Methods* 10, 1–11.
- Pickup M and Kellstedt PM** (2023) Balance as a pre-estimation test for time series analysis. *Political Analysis* 31, 295–304.
- Ritter EH and Conrad CR** (2016) Preventing and responding to dissent: the observational challenges of explaining strategic repression. *American Political Science Review* 110, 85–99.
- Rosenfeld R and Fornango R** (2014) The impact of police stops on precinct robbery and burglary rates in New York City, 2003–2010. *Justice Quarterly* 31, 96–122.
- Skolnick JH and Fyfe JJ** (1993) *Above the Law: Police and the Excessive Use of Force*. New York, NY: Free Press.
- Smith BW and Holmes MD** (2003) Community accountability, minority threat, and police brutality: an examination of civil rights criminal complaints. *Criminology; an Interdisciplinary Journal* 41, 1035–1064.
- Smith MR and Petrocelli M** (2001) Racial profiling? A multivariate analysis of police traffic stop data. *Police Quarterly* 4, 4–27.
- Sullivan CM and O'Keeffe ZP** (2017) Evidence that curtailing proactive policing can reduce major crime. *Nature Human Behaviour* 1, 730–737.
- Swencionis JK, Pouget ER and Atiba Goff P** (2021) Supporting social hierarchy is associated with White police officers' use of force. *Proceedings of the National Academy of Sciences* 118, 1–9.
- Terrill W** (2003) Police use of force and suspect resistance: the micro process of the police-suspect encounter. *Police Quarterly* 6, 51–83.
- Terrill W and Mastrofski SD** (2002) Situational and officer-based determinants of police coercion. *Justice Quarterly* 19, 215–248.
- Terrill W and Reisig MD** (2003) Neighborhood context and police use of force. *Journal of Research in Crime and Delinquency* 40, 291–321.
- Toda HY and Phillips PCB** (1994) Vector autoregression and causality: a theoretical overview and simulation study. *Econometric Reviews* 13, 259–285.

- Voigt R, Camp NP, Prabhakaran V, Hamilton WL, Hetey RC, Griffiths CM, Jurgens D, Jurafsky D and Eberhardt JL** (2017) Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences* **114**, 6521–6526.
- Walker S** (2001) *Police Accountability: The Role of Citizen Oversight*. Belmont, CA: Wadsworth.
- Warren PY and Tomaskovic-Devey D** (2009) Racial profiling and searches: did the politics of racial profiling change police behavior?. *Criminology & Public Policy* **8**, 343–369.
- Weaver VM** (2007) Frontlash: race and the development of punitive crime policy. *Studies in American Political Development* **21**, 230–265.
- Weaver VM and Lerman AE** (2010) Political consequences of the carceral state. *American Political Science Review* **104**, 817–833.
- White MD and Fradella HF** (2016) *Stop and Frisk: The Use and Abuse of a Controversial Policing Tactic*. New York, NY: NYU Press.
- Williams K** (2015) *Our Enemies in Blue: Police and Power in America*. Chico, CA: AK Press.
- Worden RE** (1995) Police officers' belief systems: a framework for analysis. *American Journal of Police* **14**, 49.
- Yokum D, Ravishankar A and Coppock A** (2017) *Evaluating the Effects of Police Body-Worn Cameras*. Washington, DC: The Lab@ DC 20.