

This is the pre-peer reviewed version of the following article: Ali, M.U. and Pirog, M. (2019), Social Accountability and Institutional Change: The Case of Citizen Oversight of Police. Public Admin Rev, 79: 411-426. <https://doi.org/10.1111/puar.13055>, which has been published in final form at <https://onlinelibrary.wiley.com/doi/10.1111/puar.13055>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Social Accountability and Institutional Change: The Case of Citizen Oversight of Police

Authors

Mir Usman Ali Ph.D.

Maureen Pirog Ph.D.

Abstract: *We examine the ability of social accountability to spur gradual institutional change at the municipal level, using the case of citizen oversight agencies (COAs) for police agencies. Using the gradual change framework (GCF) and social accountability framework to guide our empirical strategy, and data collected through an original survey of COAs, we test the impact of COAs on institutional outcomes in policing. We find, in accordance with the GCF, that the degree to which a COA reduces racial disparity in policing outcomes depends on its scope of authority, as well as the degree of discretion afforded by existing institutions to police officers. In general, the wider the scope of authority, and the broader the discretion afforded by existing institutions, the greater the likelihood of change in institutional outcomes.*

Evidence for Practice

- Investigative COAs, which conduct independent investigations into citizen complaints and have the authority to recommend discipline to police officers found guilty of misconduct, were found to be associated with a reduction in racial disparity in disorderly conduct arrests rates, as well as a reduction in racial disparity in police homicides of citizens.

- Monitoring COAs, which focus on analyzing trends in police misconduct, and recommending changes in police policies, procedures, and training, were found to be associated with a reduction in racial disparity in disorderly conduct arrest rates.
- COAs that were led by a board of citizens who were appointed by municipal district were found to be associated with a reduction in racial disparity in disorderly conduct arrest rates.
- Since most impacts of COAs become more evident over time, such agencies likely require an ongoing commitment from local governments, especially in terms of financial and human resources.

American policing is said to be suffering a crisis of legitimacy. Since the August 9, 2014, shooting of Michael Brown in Ferguson, Missouri, the public's attention has been captured by police brutality incidents, including numerous instances involving police homicides of citizens. The overrepresentation of black victims in such incidents has led to calls for institutional reforms, including demands such as ending broken windows policing (e.g., see American Civil Liberties Union 2016), appointing independent prosecutors for cases involving police violence, ending over-policing of minority communities, racially diversifying law enforcement agencies, revising use-of-force policies (see President's Task Force on 21st Century Policing 2015), as well as fundamentally rethinking the goals of policing itself (see Vitale 2017).

The above calls have also included demands for creating citizen oversight agencies (COAs), which are a type of social accountability mechanism (President's Task Force on 21st Century Policing 2015). Social accountability refers to "actions by citizens and civil society to push officeholders to report on and answer for their actions" (Brinkerhoff and Wetterberg 2016). COAs are institutional arrangements by which citizen complaints against police are reviewed at some point by persons who are not sworn officers (Walker and Bumphus 1992). In this article,

we examine their effectiveness as a means for changing institutional outcomes in policing, i.e. racial disparities in arrest rates and homicides of citizens by law enforcement officers, in the United States (US). We do this by combining the Gradual Change Framework (GCF) (Rocco and Thurston 2014), with the Social Accountability Framework (Fox 2015) to guide our empirical strategy and model building¹. We take a historical-institutionalist perspective, viewing institutions as political legacies of historical struggles which persist over time because they help maintain power imbalances favoring certain actors in society. According to the GCF, any set of rules or expectations, formal or informal, that structure action will privilege certain actors over others in terms of distributional consequences. We believe that this is plausible for certain policing practices even if they are defined in ostensibly race-neutral terms (e.g., “zero tolerance” policing and use of stop, question, and frisk tactics (see Weisburd and Majmundar 2018; Eterno, Barrow, and Silverman 2017; Gelman, Fagan, and Kiss 2007; Fagan et al. 2010)

As an antidote to the patterns of discrimination in policing, COAs are agencies that intended to enhance accountability and transparency in policing and build community trust through citizen oversight. In most police agencies across the US, when a citizen makes a formal complaint about a police officer, the complaint is investigated and adjudicated by other sworn officers. This internal investigation process is problematic because it creates a conflict of interest that tilts, or can be perceived to tilt, the accountability process in favor of the police. COAs attempt to address this accountability deficit by opening insular, internal police investigations to the scrutiny of citizens and/or professionals who serve on the COA. In theory, COAs deter police misconduct by performing a range of functions such as reviewing findings made by the police agency’s internal affairs division, assessing whether such investigations were conducted in a thorough and fair manner, and even conducting independent investigations and recommending

discipline and changes in police policies, if that authority is granted. As citizen-oriented accountability bodies, COAs can potentially create incentives for police to take pre-emptive steps that limit potential misconduct from occurring e.g., by increasing supervision of rookie officers, voluntarily changing policies pertaining to the use of force, reducing the level of engagement with minorities for minor offenses, among other steps.

Notwithstanding these implications, research on the impact of COAs, and social accountability mechanisms in general, is relatively thin in the context of developed countries (see Brinkerhoff and Wetterberg (2016), Ma (2012), Schatz (2013) and Altman (2002) for empirical social accountability studies in the context of developing countries). Little is known about the difference made by stronger COAs relative to those with limited authority in terms of impacting police behavior, especially in the US. Our study attempts to fill that gap in the literature. Given substantial evidence of the racially disparate impact of policing institutions (see Epp, Maynard-Moody, and Haider-Markel (2017), Ward and Rivera (2014), Vitale (2017), Alexander (2012), Kochel et al. (2011), Nix et al. (2017), and Menifield, Shin, and Strother (2018) among others), we believe it is incumbent upon scholars to evaluate strategies that are intended to make policing more responsive to social equity concerns. The President’s Task Force on 21st Century Policing echoed the necessity of such research, urging “evidence-based practices to implement successful civilian oversight mechanisms” (President’s Task Force on 21st Century Policing 2015). Our study highlights the institutional and organizational characteristics of COAs that are likely to be most effective in reducing the racial disparities in the aforementioned policing outcomes. Our findings should be of value to public administrators such as city managers, mayors, and police chiefs, who are interested in strategies and organizational interventions intended to make policing in their jurisdiction more racially equitable.

We believe that this is the first study that evaluates the impact of COAs on racial disparities in policing outcomes. While we use the GCF and social accountability framework to guide our model building and empirical strategy, our goal is to test the impact of different types of COAs on policing outcomes, not on institutional change, although we believe that changes in policing outcomes reflect changes in the institutions and rules governing police behavior. In short, we are concerned with the *impact* of social accountability mechanisms, such as COAs, on institutional *outcomes*.

The paper proceeds in seven sections: first, we review the findings from COA research. In the second section, we describe how the GCF, combined with propositions from the social accountability framework can serve as a blueprint to predict the impact of social accountability mechanisms. In the third section, we propose hypotheses about the conditions under which COAs are likely to lead to socially equitable outcomes in police enforcement activities that involve different degrees of discretion. In the fourth section, we introduce our empirical strategy and data set. In the fifth section, we present the results from the estimated models. In the sixth section we discuss the results and the limitations of this study, and in the seventh section, we draw conclusions about the effectiveness of COAs and the efficacy of GCF in explaining changes in institutional outcomes.

Literature Review

As stated earlier, COAs are government agencies, predominantly at the local level², that are intended to serve as a source of external oversight over police agencies. Since 1969, when the first currently existing COA was created in Kansas City, Missouri, the number of COAs in the US has gradually grown with around 145 COAs nationwide as of 2017³ (see Walker 2006, 2001) for a discussion of the history of citizen oversight in the US). COAs in the US operate in a

variety of political and socioeconomic milieus and exhibit considerable variation in terms of their formal authority, level of professionalization, staffing, budgetary authority, as well as the style of oversight. They are often created through a local government ordinance or through an amendment in the local government charter (De Angelis, Rosenthal, and Buchner 2016). The simplest COAs consist of a board of citizens which can review the findings of investigations conducted by the police agency's internal affairs division. Such COAs often have little or no budgetary authority, with the board of citizens serving on a volunteer basis. More organizationally complex COAs may include a paid, full-time staff of lawyers, investigators, and policy analysts, which in turn reports to the citizen board. Such COAs often have substantial budgetary authority, as well as the ability to conduct independent investigations into citizen complaints, and access to police evidence records and electronic databases. Based on our survey of COAs (more details provided below), while board members in most COAs are appointed by the mayor or city council, the methodology for the appointment of board members (i.e., by municipal district or at large) often varies across jurisdictions (De Angelis, Rosenthal, and Buchner 2016). Finally, prior research has postulated that COAs are likely to be created after an officer-involved shooting or incidents involving racially disparate policing (De Angelis, Rosenthal, and Buchner 2016).

Recent studies in public administration have emphasized the impact of institutions on racial disparities in policing outcomes, as opposed to individual-level behaviors (Kim and Kiesel 2018; Menifield, Shin, and Strother 2018; Epp, Maynard-Moody, and Haider-Markel 2017; Rivera and Ward 2017; Jennings and Rubado 2017; Eterno, Barrow, and Silverman 2017; Hong 2017; Maynard-Moody and Musheno 2012). While these studies effectively demonstrate the relevance and importance of widely prevalent norms, rules, and values in driving racially

disparate criminal justice outcomes, the role of accountability institutions has not received sufficient attention. On the other hand, while criminal justice and legal studies scholars have conducted a substantial degree of theorizing about the organizational characteristics of citizen oversight and the conditions under which it is likely to have an impact on policing and criminal justice outcomes, few studies have directly and empirically examined the impact of COAs on racial disparities in policing outcomes, a gap we attempt to fill. Studies in the latter category include Brereton (2000), Walker and Luna (2000), Terrill and Ingram (2016), Worden (2008), and Hickman (2006), among others. These studies examine the impact of COAs on overall police misconduct, general efficacy of citizen oversight, or trends in complaints against the police. We look in somewhat more detail at this literature to extrapolate insights for the current study.

Brereton (2000) highlighted the case of the Queensland Criminal Justice Commission (CJC), a well-resourced and well-staffed oversight body, with two-thirds of its Australian dollars 23 million budget devoted to conducting oversight operations. Introduced in the mid-1990s as part of a suite of measures to reform the QPS, the CJC played an active role in fostering robust investigations of complaints, made procedural and policy recommendations directly to the police service, and employed various proactive surveillance strategies to detect and deter serious police misconduct. Based on surveys of senior officers, several cohorts of new officers, as well as analyses of trends in complaints, Brereton found that subsequent to the establishment of the CJC, there had been (a) a reduction in the incidence of serious assault by police and other types of misconduct and (b) improvements in the police culture, including an increased tendency among officers to report misconduct by other officers. While this report only focused on a single police agency, it provides evidence that a well-resourced agency, with a substantive investigative as well as policy role, can reduce officer misconduct and improve the agency culture.

Walker and Luna (2000) conducted an evaluation of the citizen oversight system in Albuquerque, NM. Albuquerque had two oversight bodies from the late-1980s to the mid-1990s: the office of the Independent Counsel, and the Police Safety Advisory Board, in addition to the internal affairs division of the Albuquerque Police Department (APD). Despite having two oversight agencies, a racially diverse police agency, a use-of-deadly-force policy that conformed to professional standards, and a general decline in police-involved homicides of citizens in most other major US cities, the number of police-involved shootings in Albuquerque continued to be high relative to other major US cities. To understand the gaps in the accountability processes, the researchers reviewed the official documents of the above agencies, conducted interviews and surveys of a range of stakeholders. They found that the police safety advisory board and the independent counsel, while having substantial statutory authority to conduct oversight, generally failed to exercise the powers that were expressly granted to them by law. For instance, although the independent counsel had broad statutory authority to ‘direct the overall manner’ of the complaint investigation process in the APD, recommend discipline and policy changes wherever it deemed appropriate, the contract between them and the city restricted the role of the independent counsel to merely reviewing the findings of internal affairs investigations. Furthermore, the contract defined the independent counsel’s role in terms of a lawyer-client relationship, a constraint that, while *not* statutorily required, stipulated that the information gathered by the independent counsel could not be disclosed publicly. This insulated the independent counsel from the public, thus undermining public responsiveness and accountability. This study highlights that it is possible to blunt the potential impact of a COA by restricting its authority and public role, even if it originally had broad statutory authority.

Terrill and Ingram (2016) assessed the extent to which different oversight models (i.e., internal affairs, command level, or citizen oversight) were associated with whether citizen complaints against police were sustained. They found that when police agencies used COAs as part of their complaint process – whereby the COA could review internal affairs findings – the odds of a sustained disposition increased by 78%, relative to if they were referred to and investigated by internal affairs alone. Second, in cities where the COA which served only as an intake or complaint-receiving entity, the odds of a sustained disposition were 80% less than if complaints were referred the internal affairs division alone. Oversight models where complaints were referred to and investigated at the command level were 39% less likely to sustain complaints relative to when complaints were solely referred to the internal affairs division. Thus, this study suggests that review of police complaint investigations by COAs results in a higher proportion of complaints being sustained while having COAs with no ‘oversight’ authority tends to reduce the proportion of sustained complaints.

Finally, Worden (2008) indirectly evaluated the impact of a COA in deterring police misconduct⁴ in an anonymous jurisdiction as part of a larger study that examined public satisfaction with the complaint review system. This study relied on ex-post surveys and interviews of citizens who came into direct contact with police, complainants, as well as police officers. Worden argued that the likelihood of this particular COA deterring misconduct was limited because only 9 percent of respondents who perceived police misconduct actually filed a complaint with the COA or the internal affairs division (the rest complained to the police chief, another police official, or their lawyer). Even if up to 20% of these complaints were sustained (a high estimate), then only 0.5 of the perceived misconduct would be sanctioned as a result of complaint review. However, a caveat to this argument is that most of the people who

experienced police misconduct did not know that a COA existed. This study does not investigate whether a higher proportion of perceived misconduct would have been reported to the COA, and subsequently sustained, if the public knew that a COA existed? It also provides few details about the scope of authority or the powers of the COA, which can have a critical impact on the COA's potential for deterring misconduct. Per our reading, this COA seems to be a relatively weak agency with no budgetary or appointment powers (Worden 2008, 26). This study suggests that COAs which are not well-publicized, have limited powers and/or no budget, are not very likely to deter police misconduct.

The above literature, while useful, does not attempt to examine the overall impact of a large number of COAs or the impact of variation in their scope of authority on policing outcomes. Our contribution addresses this gap in the literature by examining the impact of COAs on policing outcomes, using survey data on a sample of municipal-level COAs in the US⁵. Most studies focus on either a few COAs (or just one) and hardly any studies use quasi-experimental designs. Our empirical strategy uses a two-way fixed effects methodology, with year and municipal-specific fixed effects which control for time-invariant factors and secular time trends that may impact policing outcomes, while also controlling for additional time-varying factors. Additionally, irrespective of geographic locale, no studies directly look at the impact of COAs or other oversight agencies on differential police enforcement by race as we attempt to do here. Furthermore, we bring a rigorous, empirical approach to this question, departing from literature reviews and less rigorous case study approaches. We focus on the disparity between blacks and whites because the prior literature has found that blacks are more likely to bear the brunt of policing relative to whites (Epp, Maynard-Moody, and Haider-Markel 2017; Kochel, Wilson,

and Mastrofski 2011) and because COAs are often created after an incident of police brutality against a citizen of color (De Angelis, Rosenthal, and Buchner 2016).

Gradual Institutional Change as a Blueprint for Social Accountability Impact

Per the GCF, two basic factors combine to facilitate change in institutional outcomes: (1) the conduciveness of the political context, and (2) whether the targeted institution affords opportunities for exercising discretion in interpretation or enforcement of rules (Rocco and Thurston 2014). Conduciveness of the political context often depends on the presence of powerful veto players or veto points. Veto possibilities are high, rendering the political context not conducive when actors have access to institutional or extra-institutional means to block change. These actors may have veto powers that keep change proposals off the table, or they may exercise substantial influence in how rules (formal or informal) are interpreted at the street-level.

The second factor, i.e., opportunities to exercise discretion, makes institutional change possible by allowing actors to interpret or enforce existing institutions in different ways. Thus, if existing institutions are ambiguous, they can potentially be reinterpreted in different ways or can be subject to incremental changes. Consequently, institutional change is more likely to occur, and outcomes are more likely to change. Dichotomizing the presence of institutional veto points, and the degree of discretion afforded by existing institutions results in a 2x2 table whereby four different kinds of change processes are possible. Figure-1 shows these four possibilities.

[Figure-1 about here]

The types of change referenced in each quadrant of figure 1 are briefly described in appendix I. For a more detailed explication of the different types of change see Rocco and Thurston (2014).

In this study, we use the scope of a COA's authority as an indirect measure of the strength of veto points (examples of veto points include police unions or presence of local politicians opposed to COAs). We do this for three reasons: First, we believe it is reasonable to suggest that when veto points are weak, a COA with a broad scope of authority would be more likely to be established compared to a COA with a narrow scope of authority. Conversely, when veto points are strong, a COA with a narrow scope of authority would be more likely to be established compared to one with a broad scope of authority. More broadly, using the scope of a COA's authority as a measure of the strength of veto points allows us to use the GCF to guide our empirical analysis.

Second, using the scope of COA authority as a measure of strength of veto points and therefore as a variable in model building allows us to focus on its impact on racial disparities in policing outcomes (i.e., one of the goals of this study), along with addressing a long-standing gap in the literature on the impact of COAs. Third, using the scope of authority as a variable in model building also adds to the prior literature on COAs which has highlighted but has not empirically teased out, the potential impact of the scope of a COA's authority on policing outcomes (see De Angelis, Rosenthal, and Buchner 2016; Walker and Archbold 2014; Lewis 1999). In summary, we have used the scope of COA authority as a measure of the strength of veto points because it allows us to align our empirical strategy with the GCF, while also enabling us to focus on the causal impact of scope of authority on racial disparities in policing outcomes, which is an important goal in and of itself.

Regarding the degree of discretion afforded by existing institutions, we consider two types of enforcement actions: disorderly conduct arrests (DCAs)⁶ and police homicides of citizens (PHCs). For assessing the impact of COAs on high-discretion police work we focus on DCAs, whereas for assessing the impact of COAs on low-discretion police work, we focus on the PHCs that involve the use of a firearm. We believe that DCAs can be viewed as an example of an enforcement action that affords officers high discretion because of their routine nature, which in turn makes them less likely to attract scrutiny from the public, police supervisors, or a COA, if one exists. On the other hand, PHCs involving the use of a firearm are highly salient events that are almost guaranteed to attract scrutiny from internal accountability mechanisms (Cordner and Scott 2014) and/or a COA (Fallik and Novak 2014). Thus, police officers' discretionary authority in using deadly force against citizens is checked by internal and external accountability mechanisms. Furthermore, the exercise of discretion in using deadly force is restricted by specific court rulings such as the US Supreme Court's *Tennessee v. Garner* (471 U.S. 1 [1985]) decision, which prohibits police officers from using deadly force against a fleeing felon unless the officer has probable cause to believe that the suspect poses a significant threat of death or physical injury to the officer or to others. By contrast, police officers' discretionary authority in making DCAs is significantly less likely to be checked by internal and external accountability mechanisms, which in turn increases the discretion they have when deciding to make a DCA.

In summary, we consider two levels of the scope of authority (broad and narrow), and two levels of discretion (high and low) in this study. Per the GCF, while institutional outcomes can change under any combination of veto points and discretion, the type of change that is likely to occur under a given set of conditions will likely be different. With a broad scope of authority

and low discretion, the disparity in PHCs is likely to diminish, owing to the supplanting of institutions (this type of change is referred to as ‘displacement’ in the GCF literature (see appendix I or Rocco and Thurston 2014)). With a broad scope of authority and high discretion, the disparity in DCAs is likely to diminish, owing to the enhanced opportunities for reinterpretation of institutions pertaining to DCAs (referred to as ‘conversion’ (Rocco and Thurston 2014)). With a narrow scope of authority and high discretion, the disparity in DCAs is likely to reduce incrementally (i.e., ‘layering’ (Rocco and Thurston 2014)), due to amendments to existing institutions pertaining to DCAs. Finally, with a narrow scope of authority and low discretion, the disparity in PHCs may become more pronounced, due to a failure to update institutions pertaining to the use of deadly force (i.e. drift (see appendix I)).

In the next section, we discuss propositions from the social accountability framework. By combining the propositions from the GCF and social accountability framework in our empirical strategy, we hope to contribute to a unique and nuanced understanding of social accountability impact, relative to what has been offered in the literature thus far.

Propositions from the Social Accountability Framework

Widespread replication of donor-funded social accountability initiatives in developing countries has led to an extensive literature on social accountability elements and the circumstances under which they are likely to have an impact. Meta-analyses have tended to converge on a number of propositions (e.g., see Fox 2015). Two propositions that are relevant to our study, include assertions that social accountability initiatives are likely to have an impact when they (1) combine citizen action and mobilization with the state’s capacity to respond to citizen actions, and (2) the need for citizen voice to be aggregated horizontally (across the

population) and scaled up vertically (that is, represented at the policy- or decision-making table) (Fox 2015).

Proposition 1 suggests that social accountability mechanisms are more likely to have an impact when citizen voice and the state's capacity to respond to citizens' voice mutually reinforce each other, i.e. when they have a positive interaction effect on institutional outcomes (Fox 2015). This claim is premised on the realization that a lack of accountability for bureaucratic misconduct is facilitated by *vertically integrated* relationships between elected leaders and bureaucrats. To the extent that these powerful actors shield each other's interests, they create '*low-accountability traps*,' whereby demands for citizen oversight are likely to be deflected by anti-accountability actors who may be averse to promoting pro-accountability reforms. This, in turn, suggests that oversight mechanisms can only be effective when they are *also* based in vertically integrated relationships between citizens and state actors i.e., pro-change citizens and reformist state actors should be able to empower each other. That is, *citizen voice and teeth* are needed *simultaneously* to bring about pro-accountability power shifts, rather than either voice or teeth.

Proposition (2) refers to whether the voices of marginalized communities, who have the most to gain from greater accountability of public officials, are heard or not. We propose that whether the COA is led by a board, as opposed to a single executive, and whether the board is appointed by municipal district, rather than at-large is important in this regard. Having a board that is appointed by municipal district would make it more likely that a cross-section of community voices, including minorities, would be projected up to political elites.

In summary, the GCF and social accountability framework identify four broad factors that drive the process of institutional change, and hence institutional outcomes. These include (1)

the conduciveness of the political context, which we operationalize as a COA's scope of authority, (2) the degree of discretion afforded by institutions relating to disparity in DCAs and PHCs, (3) the extent to which citizen voice and teeth reinforce each other, and (4) the governance of the COA. These factors serve as our study variables of interest, and in the next section, we present hypotheses that relate them to the outcome variables.

Hypotheses

COA Scope of Authority

Criminal Justice and Legal Studies scholars have long theorized the key elements of an effective system of citizen oversight of police (Lewis 1999; Walker and Archbold 2014). These elements tend to focus on the scope of authority of COAs, which in turn is codified at the time of their founding.

The existing literature classifies COAs into several categories based on their scope of authority (Police Assessment Resource Center 2005; Ferdik, Rojek, and Alpert 2013; De Angelis, Rosenthal, and Buchner 2016). Based on a factor analysis of thirteen survey questions, we adopt a slightly modified version of the oversight classification scheme proposed by Angelis, Rosenthal, and Buchner (2016), categorizing COAs as either investigative, monitoring, or review/audit COAs (see appendix II for a description of the factor analysis procedure used to categorize our sample COAs). We note that these categories are not mutually exclusive, rather, each type of COA is an amalgam of similar authorities that vary in emphasis.

Investigative COAs emphasize classifying citizen-initiated complaints, independent investigation, recommending findings to the police, and recommending discipline. Such agencies are also more likely to have full-time paid staff relative to other types of COAs. Monitoring COAs emphasize active monitoring of police complaint investigations and are likely to have

access to the internal affairs division's electronic databases, internal affairs files (including closed casefiles), which they use to analyze trends and patterns in police misconduct. Based on their analyses, these COAs are concerned with recommending changes to existing police policies to prevent future misconduct. They are also less likely than investigative COAs to have paid full-time staff. Finally, review/audit COAs emphasize reviewing and auditing of completed complaint investigations conducted by police. While these agencies may have access to closed internal affairs files, they are the least likely to have a full-time staff or recommend discipline or policy change. Since investigative COAs may recommend discipline against individual officers found guilty of misconduct and are the most likely to have paid full-time staff, we consider them to have the broadest scope of authority, followed by monitoring, and then review/audit COAs.

Thus, in agreement with GCF and social accountability frameworks that the conduciveness of political context matters, we hypothesize that COAs with a relatively broader scope of authority (i.e., investigative and monitoring COAs) are more likely to reduce racial disparity in DCAs and PHCs as compared to review/audit COAs. Thus, our hypotheses are:

Hypothesis 1: Investigative COAs will be associated with a reduction in racial disparity in DCAs.

Hypothesis 2: Monitoring COAs will be associated with a reduction in racial disparity in DCAs.

Hypothesis 3: Investigative COAs will be associated with a reduction in racial disparity in PHCs.

Hypothesis 4: Monitoring COAs will be associated with a reduction in racial disparity in PHCs.

Mutually Reinforcing Effect of Voice and Teeth

We operationalize teeth – a COA's ability to apply negative sanctions, as well as recommend policy change (Fox 2015) – along a three-point scale. At level-0, a COA does not

have the authority to recommend discipline to police officers it believes are guilty of some misconduct, nor does it have any authority to recommend changes in police policies. At level-1, a COA has one of the two authorities: it may either recommend discipline or policy changes – it does not have the authority to do both. At level-2 teeth, an agency can recommend both discipline and policy changes.

Voice refers to the views of citizens who would be otherwise excluded from the accountability process in the absence of a COA due to ethnic, racial or class bias (Fox 2015). As argued above, it is the synergistic combination of voice and teeth that are most likely to lead pro-accountability power shifts, i.e., institutional change. We use the number of civil rights organizations in each municipality, a time-varying variable, as a measure of a community's voice, with the assumption that an increase in the number of civil rights organizations would be associated with greater intensity in calls for police reform. Thus, hypotheses 5 and 6 are:

Hypothesis 5: COAs will reduce racial disparity in DCAs when they have the teeth to respond to citizens' voice, i.e., when there is an interaction between voice and teeth.

Hypothesis 6: COAs will reduce racial disparity in PHCs, when they have the teeth to respond to citizens' voice, i.e., when there is an interaction between voice and teeth.

Governance

COAs are often, but not always, led by a board of appointed citizens. As alluded to above, if the COA is led by a board of citizens, it may be more likely to represent minority communities and act as a bridge between citizens and the government, relative to individual efforts. Thus, we hypothesize that:

Hypothesis 7: COAs which report to a board appointed by municipal district will reduce racial disparity in arrests relative to COAs whose leaders are chosen at large, or those which are led by a single executive.

Hypothesis 8: COAs which report to a board appointed by municipal district will reduce racial disparity in the police homicide rate relative to COAs whose leaders are chosen at large, or those which are led by a single executive.

Discretion Afforded by Existing Institutions

Since DCAs are assumed to afford police officers more discretion than situations which involving PHCs, we expect COAs to be more likely to reduce disparity in DCAs as compared to the disparity in PHCs.

This leads to the following hypothesis:

Hypothesis 9: COAs are more likely to reduce racial disparity in DCAs as compared to PHCs.

The above hypotheses can be represented in diagrammatic form as shown in figure 2.

[Figure-2 about here]

Empirical Strategy

Our dependent variables are (1) disparity in DCAs between blacks and whites per 100,000 persons, and (2) disparity in PHCs between blacks and whites per 100,000 citizens.

Racial disparity in DCAs is calculated as follows:

Disparity in Arrest Rate_{mt}

$$\begin{aligned} &= \text{Arrest Rates for Disorderly Conduct for Blacks}_{mt} \\ &\quad - \text{Arrest Rates for Disorderly Conduct for Whites}_{mt} \end{aligned}$$

Where m and t index municipality and year respectively, and the arrest rate for disorderly conduct for race r_{mt} is calculated as follows:

Arrest Rate for Disorderly Conduct for race r_{mt}

$$= \left(\frac{\text{Number of Adult Arrests of race } r_{mt}}{\text{Adult population of race } r_{mt}} \right) \times 100,000$$

Racial disparity in PHCs is calculated as follows:

Disparity in Police Homicides of Citizens_{mt}

$$= \max \left\{ \begin{array}{l} (\text{Police Homicide Rate for Blacks}_{mt} - \\ \text{Police Homicide Rate for Whites}_{mt}) \\ \text{or } 0 \end{array} \right.$$

Where the police homicide rate for race r_{mt} is calculated as:

Police Homicide Rate for race r_{mt}

$$= \left(\frac{\text{Number of police homicides of citizens of race } r_{mt}}{\text{Adult population of citizens of race } r_{mt}} \right) \times 100,000$$

The measures of racial disparity in DCAs and PHCs that we use are thus population-based benchmark measures that use the race-specific adult population (i.e. blacks and whites 18 years and older) as the relevant risk set for each race. Such measures have frequently been used for measuring racial disparities in policing outcomes in a variety of studies (Wilkins and Williams 2008; Shjarback et al. 2017; Neil and Winship 2019). While scholars have pointed out the limitations of using population-based benchmarks for assessing racial disparities, it has also

been acknowledged in the literature that identification of appropriate benchmarks require detailed, localized information about the geography and nature of policing strategies – information that is generally not available in publicly available datasets (Weisburd and Majmundar 2018), or in studies that examine racial disparities in policing outcomes in more than a handful of jurisdictions. Ideally, we would have the resources to conduct surveys similar to those completed by Epp, Maynard-Moody, and Haider-Markel (2017) to develop a benchmark measure that not only accounts for the number of black and white adults, but also a measure of how much these groups tend to violate specific laws, and the extent of contact they are likely to have with the police. Given the lack of resources necessary to collect these additional data for the large number of municipalities in our sample, however, we offer the best available alternative often offered by scholars in similar situations (e.g., see Wilkins and Williams 2008; Shjarback et al. 2017). As such, the measures of racial disparity that we use represent the only plausible benchmark given the unit of analysis in our study and the large number of municipalities in our sample.

The period chosen for DCAs is from 1980 through 2014⁷, while the period chosen for PHCs is from 2000 through 2014. The sample consists of municipal-level COAs which responded to our survey. For the sake of consistency in the analysis and generalizability of conclusions, we excluded non-municipal jurisdictions from regression analyses. The method chosen is a two-way fixed effects model with year and municipality-specific fixed effects, that takes the following form:

$$\begin{aligned}
Y_{mt} = & X_{mt}\beta + \delta \textit{Investigation}_{mt} + \tau \textit{Monitoring}_{mt} + \varphi \textit{Review/Audit}_{mt} \\
& + \theta \sum_{i=1}^2 (\textit{Teeth} \times \textit{Voice}) + \vartheta \textit{Voice}_{jt} + \rho \textit{Governance}_{mt} + \pi \textit{Age}_{mt} + w_m \\
& + v_t + \varepsilon_{mt}
\end{aligned}$$

Y_{mt} represents the two dependent variables, X_{mt} contains a vector of time-varying covariates; *Investigation*, *Monitoring* and *Review/Audit* are the predicted factor scores for the three types of COAs in the sample (see appendix II for a description of the factor analysis procedure used to categorize our sample COAs). These scores are essentially indices, with higher values of an indicator denoting a broader scope of authority for a particular type of COA. For instance, a COA with a large, positive value for *Investigation*, is likely to be an investigative COA.

Teeth is a set of two indicator variables that represent level-1 and level-2 teeth. *Voice* represents the number of civil rights organizations in the city. *Teeth* \times *Voice* represents the interaction terms between levels of teeth and number of civil rights organizations. *Governance* is an indicator that equals one when the COA is led by a board whose members are appointed by district and zero otherwise, and *Age* denotes the age of the COA. w_m and v_t are municipality and year-specific fixed effects respectively, and ε_{mt} is a mean-zero random error. Our empirical strategy thus employs a two-way fixed-effects model in which the municipal-specific fixed effects control for unobservable, municipality-specific, time-invariant characteristics of individual municipalities which may be correlated with the treatment while year-specific fixed effects control for secular trends in the dependent variables. Secular time trends in the dependent variables can result from a variety of factors, such as shifts in general policing patterns, Supreme Court decisions governing police-citizen interactions in the U.S. as a whole (e.g., *Tennessee v.*

Garner (471 U.S. 1 [1985]) and *Graham v. Connor* (490 U.S. 386 [1989])). The above empirical strategy, in which the coefficients can be interpreted as within-unit effects, is often used to evaluate the impact of a policy that is adopted at different points in time by individual jurisdictions (e.g., Albalade 2008). Models for disparity in DCAs were estimated using fixed effects ordinary least squares (OLS), while models for disparity in PHCs were estimated using fixed effects Poisson regression. We used the Poisson fixed effects estimator as it is a true fixed effects estimator (as opposed to the negative binomial) – and has been found to reliably estimate the effects of regressors as well as establish statistical significance, even in the presence of over-dispersion (Wooldridge 1999). All models account for intragroup correlation using robust standard errors clustered at the municipality level.

Data

Table 1 summarizes the data sources utilized in this study. Data on COAs' scope of authority, teeth, and governance were obtained from an online survey administered in mid-2017. We contacted 114 COAs located in 111 municipal jurisdictions as part of our survey^{8,9}. A few municipalities have multiple COAs with varied roles and responsibilities. We focused on the COA within each municipality that was most directly involved with conducting oversight (i.e., reviewing complaint investigations, auditing completed complaint investigations, conducting independent investigations, recommending discipline and policy), using it as the representative form of oversight in that municipality.

Of the COAs which were contacted, 91 COAs located in 88 municipal jurisdictions responded to the survey, resulting in a gross municipal-level response rate of 79.28% ($=88/111$)¹⁰. Among the COAs responding from these 88 municipalities, one had been created in

2016 as a result of which it was not used in our analysis (which spanned 1980 through 2014). Another COA returned a survey that was only partially complete, and 6 additional COAs (which although returned completed surveys) belonged to municipalities for which arrest data were not available through the UCR program. We also made a freedom of information request for arrest data for these 6 municipalities which, unfortunately, was declined. After removing these 8 municipalities from our sample, we were ultimately left with 80 municipalities (corresponding to 83 COAs) that could be used in our study. The net (or effective) municipal-level response rate for our survey was thus 72.07% ($=80/111$). Our analytic sample thus represents a snapshot of the institutional structure of citizen oversight in 80 distinct municipalities.

[Table 1 about here]

Figure 3 shows the geographical dispersion of COAs in our sampling frame, highlighting that COAs tend to be concentrated in cities with large populations. Table 2 presents summary statistics from the survey. The median 2014 population in municipalities that responded to our survey is around 254,000. The median percentage of black citizens (in 2014) in a given municipality in our sample is 16.1%, which was larger than the country overall (12.6% as of 2014 (U.S. Census Bureau 2014)). Considering that the median year of COA establishment for our sample is 2000 suggests that these agencies have mushroomed in the past couple decades. Overall, the above facts suggest that COAs tend to be established in large cities with a relatively

high proportion of blacks, and thus where interactions between blacks and police officers are likely to be more frequent.

[Figure 3 about here]

[Table 2 about here]

Regarding endogeneity, a concern is whether municipalities self-select into creating COAs. We tested for this possibility by testing for the significance of time distance dummies from the year in which a COA was created, as well as testing whether the parallel trend assumption is met for cities that established COAs versus those that did not (for untreated cities, we used all US cities that had never established a COA, whose 2010 population exceeded 100,000, and for which data on DCAs and PHCs were available). Overall, these tests indicated that there is little or no self-selection into the creation of COAs. Appendices IV and V explain these tests in more detail and present regression estimates from them.

Data on PHCs were obtained from the crowd-sourced database, the Fatal Encounters (FE) Project, which is maintained and fact-checked by a nonprofit. This database catalogs every

police-involved death since 2000. While the Federal Bureau of Investigation (FBI) does collect data on police-involved civilian deaths, these data are subject to substantial underreporting, the extent of which varies across jurisdictions (Barber et al. 2016). Several recent public administration studies studying PHCs have either directly used the FE database, or have used it to cross-check data gathered from other crowdsourced databases (see Jennings and Rubado 2017; Menifield, Shin, and Strother 2018; Nicholson-Crotty, Nicholson-Crotty, and Fernandez 2017).

Following previous research utilizing the FE database (see Jennings and Rubado 2017), we only retained firearm-related deaths in our data, while ignoring other types of deaths, e.g., those involving tasers, vehicles or asphyxiation, etc., because they are less likely to be intentional. We also excluded deaths where more than one police agency was involved, due to the difficulty in attributing such deaths to specific police agencies. Furthermore, as we were interested in municipalities, deaths involving state or federal agencies were also eliminated. Since we are evaluating the impact of COAs only within the municipalities where they have been established, we were left with 1,269 police-involved deaths across 80 police agencies from 2000 through 2014.

All demographic variables for census years were obtained from US census reports or the American Community Survey (ACS)¹¹, while values for intercensal years were obtained through linear interpolation. We also used municipality-level violent crime rates for each jurisdiction from the UCR as a covariate. These time-varying covariates were used because, in prior research, socioeconomic variables have been found in prior research to be correlated with policing tactics as well as social disorder and crime (Zhao, He, and Lovrich 2006; Zhao, Ren, and Lovrich 2012). We included an indicator variable for any municipality under federal investigation or bound by an agreement to reform policing practices vis-à-vis use of force or

racial profiling (e.g., through a consent decree, settlement agreement, or memorandum of agreement), which could potentially reduce racial disparity by itself. Finally, in the model for racial disparity in PHCs, we included racial disparity in DCAs as a baseline measure of bias in policing.

Results

Model estimates for racial disparity in DCAs and PHCs are shown in table 3 and table 4, respectively. In each case, we started with a basic model containing a creation indicator which switches from zero to one from the year the COA came into existence. In subsequent models we substituted the creation indicator, sequentially adding factor scores for the scope of authority, followed by indicators for teeth-voice interaction and governance. In the final model, which is the preferred specification, we added covariates to determine whether the observed effects changed after controlling for background characteristics.

Racial Disparity in Disorderly Conduct Arrests

Table-3 presents various model estimates for the impact of COAs on racial disparity in disorderly conduct arrest rates.

[Table-3 about here]

Starting with model 1, where we estimate the overall impact of COA creation on the dependent variable, we find that the created coefficient was is negative but insignificant. The age

coefficient shows a significant, negative impact on racial disparity in DCAs over time. The latter effect persists generally for all types of COA vis-à-vis DCAs in models 1 through 6 and suggests that all types of COAs, regardless of their individual characteristics, decrease racial disparity in DCAs *over time*, and implies the time-intensive nature of social accountability mechanisms.

In addition, the scope of authority coefficients for investigative COAs are consistently negative across models 1 through 6, and their impact attains statistical significance in model 5 ($p=0.049$). In model 6, the investigation score coefficient has a sizeable magnitude of -121.80 (relative to the average pre-treatment disparity in DCAs of 487.5 per 100,000 adults), however, it is not significant at the 5% level ($p=0.082$). The interaction coefficient for investigation score and age is positive ($\beta_{Investigation\ score \times Age} = 7.07, p=0.039$), which seems to partially offset the main effect of age ($\beta_{Age} = -20.67, p=0.024$)¹². Nevertheless, we believe these results suggest that investigative COAs lead to a reduction in racial disparity in DCAs over time, and potentially in the year of creation as well.

The coefficient for monitoring COAs fluctuates between negative and positive values and does not approach statistical significance. However, its interaction with the age coefficient is negative and significant ($p=0.049$), which suggests that monitoring COAs reduce disparity in DCAs over time. Finally, neither the review/audit score coefficient nor its interaction with the age coefficient is significant in the preferred specification. Thus, review/audit COAs do not appear to have a differential effect on the disparity in DCAs beyond the main effect of age. This result makes sense as review/audit COAs have the least authority among the three types of COAs.

Reviewing the teeth-voice interaction terms, as well as their interactions with age, we can see that while they are all negative in the preferred specification, none of them approach

statistical significance, which shows that there is little evidence from our study that mutually reinforcing teeth and voice, in and of themselves, reduce disparity in DCAs¹³. The main effect of voice in models 1 through 6 also does not attain statistical significance, however, it is consistently negative per our expectation.

The governance indicator is highly negative and significant in the preferred specification¹⁴.

Racial Disparity in Police Homicides of Citizens

Table-4 presents various model estimates for the impact of COAs on racial disparity in PHCs.

[Table 4 about here]

Model 1B estimates the impact of COA creation on racial disparity in PHCs. Neither the creation coefficient ($p=0.472$) nor the age coefficient ($p=0.582$) approach significance. However, like the model for racial disparity in DCAs, the creation coefficient is negative.

Regarding the scope of authority coefficients in models 1B through 6B, we note that the investigation and monitoring score coefficients are consistently negative. The interaction of these coefficients with the age coefficient is also consistently negative, with the coefficient for the interaction between investigation score and age negative and significant ($p=0.037$), which suggests that investigative COAs reduced disparity in PHCs by about 6% per year. On the other hand, neither the review/audit coefficient nor its interaction with age is significant. Thus, it

appears that investigative COAs are more likely to reduce disparity in PHCs, relative to the two other types of COAs.

None of the teeth-voice interaction terms, the main effect of voice, or the governance indicator, are significant in the preferred specification. However, it bears mentioning that the coefficient for voice was negative and approached significance in models 4B ($p=0.054$) and 5B ($p=0.082$). In supplementary analyses, we also estimated separate models for the impact of COAs on police homicide rate of blacks, police homicides rate of whites, and the combined police homicide rate of blacks and whites. Results from the latter model show that investigative COAs are the only COAs associated with some reduction (3%, $p=0.093$) in the combined police homicide rate of blacks and whites¹⁵.

Discussion

We found that all COAs, regardless of type, reduce racial disparity in DCAs by around 20.7 arrests per 100,000 adults per year (showing support for hypotheses 1 and 2). On the other hand, the impact of COAs on PHCs varies by the type of COA. Investigative COAs reduce racial disparity in PHCs by around 6% per year, while neither monitoring nor review COAs reduce racial disparity in PHCs, indicating a lack of support for hypothesis 4. Thus, it seems that less aggressive forms of oversight are sufficient to reduce the racial disparity in high discretion enforcement actions such DCAs. However, more aggressive forms of oversight are needed to reduce racial disparities in enforcement actions such as PHCs, where institutions afford officer low discretion and the citizen-police encounters pose danger.

Regarding the teeth-voice interaction and their interaction terms with age, neither has an impact on the disparity in DCAs or PHCs, which leads us to find lack of support for hypotheses 5 and 6. However, we do not believe that the above results necessarily suggest that teeth are

useless with regard to deterring racial disparity in DCAs or PHCs. Rather, we believe that teeth are likely to be effective when part of a larger repertoire of authorities. Indeed, while conducting the factor analysis procedure (see appendix II), we found that the authority to recommend discipline heavily loads onto the factor for investigative agencies, while the authority to evaluate and recommend policy changes heavily loads onto the factor for monitoring agencies. However, we intentionally did not include these authorities in predicting the scope of authority factors because doing so would have led to collinearity with the teeth-voice interaction, whose net impact we wanted to test separately. Thus, while the above results suggest that while agency teeth, in and of themselves, may not be enough to reduce disparity in DCAs or PHCs, we believe they are likely to be important components in the overall authority of COAs.

Hypotheses 7 and 8 posited that COAs that report to a board whose members are appointed by municipal district would reduce disparity in DCAs and PHCs, respectively, relative to COAs that either (a) do not have a board, or (b) have a board whose members are not appointed by municipal district. We reasoned that COAs whose members are appointed by district would be more likely to project the concerns of minority citizens. By district representation on the COA board was found to reduce racial disparity in DCAs as well as PHCs, however, only the reduction in the disparity in DCAs was statistically significant ($p=0.047$). Specifically, we found that by district COA governance reduces the racial disparity in disorderly conduct arrests by around 41% $=(200.28/487.5)$ relative to the average pretreatment disparity in such arrests. It is worth noting that the governance indicator is negative in every model for *both* dependent variables, even though each dependent variable was obtained from different data sources. This suggests that the governance of COAs likely has major implications in terms of

reducing racial disparity in policing outcomes, even if we did not obtain statistically significant results.

Finally, hypothesis 9 posited that COAs would be more likely to reduce disparity in DCAs relative to PHCs, as DCAs afford officers relatively broader discretion. While we did not test this hypothesis directly, we believe that two findings lend support to it. First, a reduction in the racial disparity of DCAs was observed for each additional year of a COA's existence, regardless of the type of COA that was in place (i.e. 20.7 fewer DCA arrests per 100,000 adults per year). On the other hand, only investigative COAs reduced racial disparity in PHCs (by 6% per year) for each additional year of existence of the COA. This finding shows that the impact of COAs on the racial disparity in DCAs are much more broad-based than the impact on racial disparity in PHCs.

Second, the impact of COAs on racial disparity in DCAs was also more sizeable than the impact of COAs on the racial disparity in PHCs. Specifically, racial disparity in DCAs decreased by 41% when the COA was led by a board with members appointed by municipal district. By contrast, COA governance by municipal district did not have any impact on the racial disparity in PHCs. In fact, only the investigative form of COAs reduced racial disparity in PHCs, although it was by a notable 6% per year. In sum, the broad-based impact of COAs on racial disparity in DCAs, as well as the difference in the scale of impact of COAs on DCAs vis-à-vis PHCs lends credence to the hypothesis that institutional outcomes are more likely to change when institutions afford bureaucrats broad discretion.

Another issue deserves elaboration. Since investigative COAs were found to reduce disparity in DCAs as well as PHCs, it is important to highlight the characteristics of the agencies in this category. A factor analysis conducted to determine the various types of COAs in the

sample is helpful in this regard (see appendix II). Per this analysis, authorities that loaded heavily onto the investigation score factor included the authority to (a) classify the nature of a citizen-initiated complaint, (b) conduct investigations of citizen complaints independent of the police agency, (c) recommend/issue investigation findings to the police, (d) recommend discipline to officers it found guilty of misconduct, and (e) have paid, full-time staff.

There were sixteen COAs in our sample which had all the above-mentioned authorities. These agencies had a median 2016 budget of \$999,420, a median full-time staff of 9, compared to a median budget of \$112,500 and a median full-time staff of 4 for the sample overall. Thus, it appears that the most effective COAs not only have a broad scope of authority, but they were also likely to have more resources to perform their role relative to the overall sample of COAs.

The following table summarizes the impacts that COAs were observed to have.

[Table-5 about here]

Our study raises a potential question about the **desirability of discretion**. On the one hand, scholars have contended that the exercise of discretion is not only an unavoidable element of street-level decision-making (Lipsky 1980), but that it enable bureaucrats to treat *people as people*, allowing them to tailor their decisions to clients' circumstances, motives, intent, and promises (Cordner and Scott 2014). However, several studies (including this one) have found that the greater the discretionary authority that street-level bureaucrats have, the greater the likelihood for bias to permeate their decisions (Maynard-Moody and Musheno 2012; Lipsky

1980; Cárdenas and Ramírez de la Cruz 2017). Does this finding suggest that discretionary authority ultimately undermines social equity and that public managers ought to design jobs to reduce discretion as much as possible?

We do not believe that to be the case. As Skolnick pointed out in his 1966 classic *Justice without trial*, increased discretion does not necessarily lead to more bias. Rather, bias is likely to be evinced when officers perceive an encounter with a citizen to be dangerous, when some degree of coercion is likely to be involved, and when the officer does not believe that the suspect is likely to ultimately face some measure of ‘justice’ without the officer exercising discretion in this particular case. In light of Skolnick’s insight, it is possible that our finding that investigative COAs reduce the racial disparity in PHCs merely reflects that COAs induce police officers to use their discretionary authority in a more restrained way, rather than reducing their discretion. This study’s findings should thus not be interpreted as suggesting that reducing police officers’ discretion is necessarily a desirable end goal.

The unique contribution of this study is three-fold. The issues of whether (a) COAs impact policing outcomes, and whether (b) such impacts vary by type of COA, are long-standing and of great concern in the literature on citizen oversight of police (see De Angelis, Rosenthal, and Buchner 2016; Walker and Archbold 2014; Walker 2006). Our study not only responds in the affirmative to *both* the above questions from the prior literature but also, (c) theoretically extends the state-of-the-art of social accountability impact evaluation research by combining the social accountability framework (Fox 2015) from the Development Studies literature and the gradual change framework from Political Science (Rocco and Thurston 2014; Mahoney and Thelen 2010). Thus, our approach not only bridges two literatures that have at times been disconnected from each other, but also offers a considerable advantage in terms of parsimony in

terms of the variables. Moreover, combining the social accountability framework with the GCF allows us to test a broader range of hypotheses than would be possible using just one of these frameworks. We encourage scholars of social accountability and social change to explore other literatures which could be integrated with the social accountability and GCF frameworks (e.g., social movement theory (see Schneiberg and Lounsbury 2017)).

That COAs with a broad scope of authority lead to reductions in racial disparity of DCAs and PHCs, supports contentions in the prior literature that external triggers for social equity can enhance the ability of public organizations to treat clients in an equitable manner (Gooden 2014), and that street-level bureaucrats are less likely to treat clients differentially if it is likely that they will incur major costs for doing so (Lipsky 1980). Our study builds upon the foregoing scholarship by finding that all external triggers for social equity are not created equally – COAs with a relatively broader scope of authority are likely to be more effective in reducing racial disparity in PHCs relative to those with a narrow scope of authority. Furthermore, we found that by district COA governance reduces racial disparity in DCAs which in turn shows that institutional rules for stakeholder engagement in participatory processes affect stakeholders' influence on policy *outcomes*. This result adds to the prior finding in the literature that stakeholder engagement in participatory processes enhances stakeholder influence on policy decisions (Baldwin 2018).

Our study also challenges the narrative of the 'punctuated equilibrium' model of change that has traditionally been embraced in rational choice, sociological, and historical-institutionalist accounts of institutional change. These paradigms typically stress the importance of the stability of institutions and situate the sources of institutional change in exogenous forces (DiMaggio and Powell 1983; Hannan and Freeman 1989; Thelen and Conran 2016). Per the above accounts,

when substantive organizational change does occur, it is likely to be “episodic and dramatic, responding to institutional change at the macrolevel” (DiMaggio and Powell 1991).

However, recent scholarship in the historical-institutionalist tradition has increasingly focused on institutions as distributional instruments laden with power imbalances, which lead to contestation or uncertainty regarding the implementation and meaning of institutional rules, and hence compliance with them (Thelen and Conran 2016). Under this view, institutions and hence institutional outcomes may change gradually as institutions undergo piecemeal modification or reinterpretation. Moreover, due to variations in the sociopolitical context and the discretion afforded by existing institutions, organizational change may not necessarily be ‘isomorphic’ as contended in sociological institutional accounts (DiMaggio and Powell 1983; Dobbin 1994; Scott 1995). While we did not trace the historical evolution of specific institutions in individual cases in this paper, the pattern of findings we uncover lend credence to the historical-institutionalist perspective on gradual institutional change: First, we found that in both low discretion enforcement work (i.e., PHCs) as well as high-discretion enforcement (i.e., DCAs), at least part of the change in outcomes occurs *gradually and over time rather than abruptly*. This was evidenced by the gradual reduction in the racial disparity in DCAs for COAs of all types, as well as the gradual reduction in the racial disparity of PHCs for investigative COAs. Second, consistent with the gradual change framework (GCF), changes in institutional outcomes appear to be a function of *both* the sociopolitical context (operationalized here as the scope of COA authority), as well as the degree of discretion afforded by existing institutions. Specifically, racial disparities in DCA narrow regardless of whether the scope of COA authority is broad or narrow i.e., change through either conversion or layering is plausible (see Thelen and Conran 2016; Rocco and Thurston 2014). On the other hand, racial disparities in PHCs narrowed when the

scope of COA authority was broad, i.e., outcomes plausibly changed through displacement of institutions. We did not observe an increase in racial disparities in PHCs when the COAs' scope of authority was narrow, suggesting overall a relatively low likelihood of institutional change through drift. The implication here is that the type of institutional change plausibly *varies* depending on the scope of COA authority, as well as the discretion afforded by the institutions themselves. Nevertheless, it bears emphasizing that the foregoing types of institutional changes are based on our broad-brush impressions, and as noted below, we encourage scholars to investigate the specific patterns of institutional change occurring in response to COAs through single-subject case studies.

Finally, this study broadens the typical focus of social accountability impact evaluations from developing countries to developed countries. We believe that social accountability mechanisms may also be a viable mechanism for institutional change in developed countries, where accountability deficits rooted in historic political struggles are also likely to exist, and whose symptoms are increasingly evident (e.g., the emergence of the Black Lives Matter movement) in an era of increasing political and economic inequality. We hope that broadening the scope of social accountability studies to developed countries prompts deeper reflection in the public administration field on the historical-institutional reasons as to why accountability deficits exist in the governance of essential public services, the long-term impacts of such deficits on social equity, as well as the development of conceptual frameworks and theories about how such deficits can be narrowed.

There are certain limitations of this study that ought to be acknowledged. First, we assumed that COAs' authorities remain constant over time. This assumption is needed in a fixed effects model to rule out the presence of time-varying factors which may be confounded with the

treatment. We verified this assumption by determining for a random sample of COAs, whether their authorities had changed since their inception by searching newspaper archives on Lexis-Nexis. While we did not find any newspaper reports of changes in the powers of the COAs in our sample, it is possible that some changes were not reported in the press which might have biased the treatment effect.

Second, as alluded to above, while the results of this study suggest institutional outcomes changed as a result of the creation of COAs, we cannot pinpoint the exact type of institutional change that is likely to have occurred. We believe that a deeper, qualitative study at the individual-COA level would be needed to determine the specific change process that occurred in a city. Future research should attempt to open the black box of the institutional change process through which policy interventions such as COAs can enhance the social equity of police services.

Conclusion

We started this study with the question of whether COAs have an impact on racial disparity in policing outcomes, and whether their outcomes vary by the kind of oversight that the COA conducts. A major weakness of prior studies that have considered these questions is that they have been largely atheoretical and have thus not yielded clear predictions of (a) whether COAs are likely to be effective, and (b) the circumstances under which they are likely to be effective. Other weaknesses of prior evaluations include the fact that they have almost never considered racial disparities in policing outcomes as dependent variables and have almost exclusively been restricted to case studies of individual jurisdictions. Such case studies, while illuminating, have not provided generalizable insights about when COAs are likely to be effective.

Our analyses show that COAs reduce racial disparities in policing outcomes over time when the enforcement action affords officers broad discretion. Moreover, this effect seems to be a function of a COA being present, rather than a specific kind of COA. However, COAs with a broad scope of authority are more likely (compared to COAs with a narrow scope of authority) to reduce the racial disparity in policing outcomes when the enforcement action affords narrow discretion and likely arises from dangerous citizen-police encounters. The results also suggest that the institutional rules by which oversight agencies are governed (i.e., by municipal district governance or at large governance) can also have an impact on racial disparity, at least for enforcement actions that afford officers broad discretion, and which do not involve dangerous citizen-police encounters.

Our results thus indicate that public administrators should consider whether certain police enforcement actions in their municipality are associated with racial disparities that disproportionately affect blacks. If blacks are only overrepresented in arrests for ‘quality-of-life’ offenses such as disorderly conduct, and if the local government is unable to commit to the resource outlays that investigative COAs entail, then public administrators may consider establishing a review-focused COA. On the other hand, if blacks are overrepresented in PHCs (or PHCs as well as DCAs) then the public administrators should consider establishing an investigative COA – one with a substantial scope of authority, budgetary authority, as well as the appropriate quality and quantity of human resources. Underpowered COAs that are essentially intended to serve as window-dressing may not necessarily have any impact on more serious forms of police violence that disproportionately affects blacks.

Taken together, the above findings suggest that robust citizen oversight is likely an effective approach to enhance equity in service provision contexts that have a history of

institutional racism and that also have the potential for inflicting direct physical harm. We believe this is true for policing – indeed, no other public service has the authority to use deadly force against citizens on behalf of the state.

COAs seek to deter police misconduct through a variety of means such as controlling police discretion through recommending policy changes, evaluating individual complaint investigations as well as broader patterns in misconduct, and by recommending discipline for officers found guilty of misconduct. As such, they aim to alter the incentives of police by generating extrinsic and intrinsic motivators that encourage equitable treatment of citizens. We urge scholars in public administration and cognate fields to conduct further research into how social accountability mechanisms promote social equity, obstacles that hinder them, and as to how they can be made more effective.

Notes

1. GCF has earlier been referred to as the theory of gradual institutional change (Mahoney and Thelen 2010).
2. The vast majority of oversight agencies in the US are at the municipal level, followed by COAs at the county level, and COAs in universities and transit systems.
3. This is an estimated count based on corroboration with other researchers, the National Association for Citizen Oversight of Law Enforcement (NACOLE), as well as internet searches conducted by us. Nonetheless, it is possible that the actual number of COAs is higher as there might be certain COAs that are not known to the sources we consulted, or which had no presence on the internet.
4. Misconduct was defined as “any alleged improper or illegal act, omission or decision” by a police officer that directly affects the person or property of an individual by reason of a violation of the police agency’s policy, or any local, state or federal law (Worden 2008).
5. A copy of the survey instrument is available from the authors upon request.
6. Disorderly conduct is defined on the Federal Bureau of Investigation’s (FBI) Uniform Crime Report website as “Any behavior that tends to disturb the public peace or decorum, scandalize the community, or shock the public sense of morality.”
7. We decided to use data from 1980 onwards because the Uniform Crime Reporting (UCR) city-level arrest data prior to 1980 tend to fluctuate significantly, raising concerns of incorrect reporting and/or measurement error.
8. The survey was created and administered through the Qualtrics software.
9. Separate and individual surveys were sent to those jurisdictions in which the board and the COA are separate entities.
10. Names of the municipalities whose COAs responded to our survey have been kept confidential due to Institutional Review Board (IRB) requirements.
11. The 1980, 1990, 2000 and 2010 values of these variables were obtained from US Census reports, while the values for 2015 were obtained from the ACS 5-year estimates.
12. A test of significance comparing the $\beta_{Investigation\ score \times Age}$ ($= 7.07$) and β_{Age} ($= -20.67$) coefficients shows that the latter is statistically larger than the former ($p=0.0930$). This, in turn shows that the $\beta_{Investigation\ score \times Age}$ coefficient only partially offsets the β_{Age} coefficient.

13. In an alternate specification, we use the percentage of a municipality's population that is black as a measure of voice. The interaction term of teeth with this alternative measure of voice had a similar impact on the dependent variable.
14. We also estimated the impact of COAs on DCA rates of blacks and whites separately. Those results were not included in the article to conserve space and will gladly be provided by the first author upon request.
15. Estimates for these models were not included in the article to conserve space and will gladly be provided by the first author upon request.

Bibliography

- Albalade, D. 2008. "Lowering Blood Alcohol Levels to Save Lives: The European Experience." *Journal of Policy Analysis and Management* 27 (1):20–39.
- Alexander, M. 2012. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. The New Press.
- Altman, David. 2002. "Prospects for E-Government in Latin America: Satisfaction With Democracy, Social Accountability, and Direct Democracy." *International Review of Public Administration* 7 (2):5–20. <https://doi.org/10.1080/12294659.2002.10805002>.
- American Civil Liberties Union. 2016. "Government Report Shows Broken Windows Policing Doesn't Make New York Safer." ACLU Website. 2016. <https://www.aclu.org/news/government-report-shows-broken-windows-policing-doesnt-make-new-york-safer>.
- Angelis, J. De, R. Rosenthal, and B. Buchner. 2016. "Civilian Oversight of Law Enforcement: Assessing the Evidence."
- Angelis, Joseph De, Richard Rosenthal, and Brian Buchner. 2016. "Civilian Oversight of Law Enforcement: Assessing the Evidence."
- Baldwin, Elizabeth. 2018. "Exploring How Institutional Arrangements Shape Stakeholder Influence on Policy Decisions: A Comparative Analysis in the Energy Sector." *Public Administration Review* 00:1–10. <https://doi.org/10.1111/puar.12953>.
- Barber, Catherine, Deborah Azrael, Amy Cohen, Matthew Miller, Deonza Thymes, David Enze Wang, and David Hemenway. 2016. "Homicides by Police: Comparing Counts From the National Violent Death Reporting System, Vital Statistics, and Supplementary Homicide Reports." *American Journal of Public Health* 106 (5). American Public Health

- Association:922–27. <https://doi.org/10.2105/AJPH.2016.303074>.
- Brereton, D. 2000. “Evaluating the Performance of External Oversight Bodies.” In *Civilian Oversight of Policing: Governance, Democracy and Human Rights*, 156–83.
- Brinkerhoff, Derick W., and Anna Wetterberg. 2016. “Gauging the Effects of Social Accountability on Services, Governance, and Citizen Empowerment.” *Public Administration Review* 76 (2):274–86. <https://doi.org/10.1111/puar.12399>.
- Cárdenas, Sergio, and Edgar E. Ramírez de la Cruz. 2017. “Controlling Administrative Discretion Promotes Social Equity? Evidence from a Natural Experiment.” *Public Administration Review* 77 (1):80–89. <https://doi.org/10.1111/puar.12590>.
- Cordner, G., and M.S. Scott. 2014. “Police Discretion and Its Control.” In *Encyclopedia of Criminology and Criminal Justice*, edited by G. Bruinsma and David L. Weisburd, 3875–3595. New York: Springer.
- DiMaggio, and W Powell. 1983. “The Iron Cage Revisited: Collective Rationality and Institutional Isomorphism in Organizational Fields.” *American Sociological Review* 48 (2):147–60.
- DiMaggio, and WW Powell. 1991. “The New Institutionalism in Organizational Analysis.” In *The New Institutionalism in Organizational Analysis*, edited by PJ DiMaggio and WW Powell. University of Chicago Press.
- Dobbin, F. 1994. *Forging Industrial Policy: The United States, Britain, and France in the Railway Age*.
https://books.google.com/books?hl=en&lr=&id=y5QS3cDJUT0C&oi=fnd&pg=PR11&dq=Dobbin,+Frank.+1994.+Forging+Industrial+Policy:&ots=Q6qLQPMR3Y&sig=mTUuRpb_XWdECf63Vm68nIU2bCA.

- Epp, Charles R., Steven Maynard-Moody, and Donald Haider-Markel. 2017. "Beyond Profiling: The Institutional Sources of Racial Disparities in Policing." *Public Administration Review* 77 (2). Wiley Subscription Services, Inc.:168–78. <https://doi.org/10.1111/puar.12702>.
- Eterno, John A., Christine S. Barrow, and Eli B. Silverman. 2017. "Forcible Stops: Police and Citizens Speak Out." *Public Administration Review* 77 (2):181–92. <https://doi.org/10.1111/puar.12684>.
- Fagan, J, A Geller, G Davies, and V West. 2010. "Street Stops and Broken Windows Revisited." In *Race, Ethnicity, and Policing*, edited by KR Rice and MD White, 1st ed., 309–48. New York University Press.
- Fallik, Seth W., and Kenneth J. Novak. 2014. "Biased Policing." In *Encyclopedia of Criminology and Criminal Justice*, edited by G. Bruinsma and David L. Weisburd, 154–62. New York: Springer.
- Ferdik, Frank V., Jeff Rojek, and Geoffrey P. Alpert. 2013. "Citizen Oversight in the United States and Canada: An Overview." *Police Practice and Research* 12 (2):104–16. <https://doi.org/10.1016/B978-0-08-098222-9.00004-2>.
- Fox, Jonathan A. 2015. "Social Accountability: What Does the Evidence Really Say?" *World Development* 72. Elsevier Ltd:346–61. <https://doi.org/10.1016/j.worlddev.2015.03.011>.
- Gelman, Andrew, Jeffrey Fagan, and Alex Kiss. 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*. <https://doi.org/10.1198/016214506000001040>.
- Gooden, ST. 2014. *Race and Social Equity: A Nervous Area of Government*. ME Sharpe.
- Hannan, MT, and J Freeman. 1989. *Organizational Ecology*. Cambridge, Harvard, U. Press.

- Cambridge, MA: Harvard University Press.
- Hickman, Matthew J. 2006. "Bureau of Justice Statistics Citizen Complaints about Police Use of Force." *Bureau of Justice Statistics*.
- Hong, Sounman. 2017. "Does Increasing Ethnic Representativeness Reduce Police Misconduct?" *Public Administration Review* 77 (2):195–205.
<https://doi.org/10.1111/puar.12629>.
- Jennings, Jay T., and Meghan E. Rubado. 2017. "Preventing the Use of Deadly Force: The Relationship between Police Agency Policies and Rates of Officer-Involved Gun Deaths." *Public Administration Review* 77 (2):217–26. <https://doi.org/10.1111/puar.12738>.
- Kim, Jaeok, and André Kiesel. 2018. "The Long Shadow of Police Racial Treatment: Racial Disparity in Criminal Justice Processing." *Public Administration Review* 78 (3):422–31.
<https://doi.org/10.1111/puar.12842>.
- Kochel, Tammy R, David B Wilson, and Stephen D Mastrofski. 2011. "Effect of Suspects Race on Officers' Arrest Decisions." *Criminology* 49 (2):473–512.
<https://doi.org/10.1111/j.1745-9125.2011.00230.x>.
- Lewis, Colleen. 1999. *Complaints against Police: The Politics of Reform*. Hawkins Press.
- Lipsky, M. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Service*. New York: Russell Sage Foundation.
- Ma, Jun. 2012. "The Rise of Social Accountability in China." *Australian Journal of Public Administration* 71 (2):111–21. <https://doi.org/10.1111/j.1467-8500.2012.00770.x>.
- Mahoney, J., and K. Thelen. 2010. "A Theory of Gradual Institutional Change." In *Explaining Institutional Change: Ambiguity, Agency, and Power*, edited by J Mahoney and K Thelen. Cambridge University Press.

- Maynard-Moody, Steven, and Michael Musheno. 2012. "Social Equities and Inequities in Practice: Street-Level Workers as Agents and Pragmatists." *Public Administration Review* 72 (SUPPL.1):16–23. <https://doi.org/10.1111/j.1540-6210.2012.02633.x>.
- Menifield, Charles E., Geiguen Shin, and Logan Strother. 2018. "Do White Law Enforcement Officers Target Minority Suspects?" *Public Administration Review* 00:1–13. <https://doi.org/10.1111/puar.12956>.
- Neil, R, and C Winship. 2019. "Methodological Challenges and Opportunities in Testing for Racial Discrimination in Policing." *Annual Review of Criminology* 2:73–98.
- Nicholson-Crotty, Sean, Jill Nicholson-Crotty, and Sergio Fernandez. 2017. "Will More Black Cops Matter? Officer Race and Police-Involved Homicides of Black Citizens." *Public Administration Review* 77 (2):206–16. <https://doi.org/10.1111/puar.12734>.
- Nix, Justin, Bradley A. Campbell, Edward H. Byers, and Geoffrey P. Alpert. 2017. "A Bird's Eye View of Civilians Killed by Police in 2015." *Criminology & Public Policy* 16 (1):309–40. <https://doi.org/10.1111/1745-9133.12269>.
- Police Assessment Resource Center. 2005. "Review of National Police Oversight Models." <https://static1.squarespace.com/static/5498b74ce4b01fe317ef2575/t/54caf3abe4b04c8e2a3b6691/1422586795583/Review+of+National+Police+Oversight+Models+%28Feb.+2005%29.pdf>.
- President's Task Force on 21st Century Policing. 2015. "Final Report of the President's Task Force on 21st Century Policing." Washington, DC.
- Rivera, Mario A., and James D. Ward. 2017. "Toward an Analytical Framework for the Study of Race and Police Violence." *Public Administration Review* 77 (2):242–50. <https://doi.org/10.1111/puar.12748>.

- Rocco, P, and C Thurston. 2014. "From Metaphors to Measures: Observable Indicators of Gradual Institutional Change." *Journal of Public Policy* 34 (1):35–62.
- Schatz, Florian. 2013. "Fighting Corruption with Social Accountability: A Comparative Analysis of Social Accountability Mechanisms' Potential to Reduce Corruption in Public Administration." *Public Administration and Development* 33 (3):161–74.
<https://doi.org/10.1002/pad.1648>.
- Schneiberg, M, and M Lounsbury. 2017. "Social Movements and the Dynamics of Institutions and Organizations." In *The SAGE Handbook of Organizational Institutionalism*, edited by R Greenwood, C Oliver, TB Lawrence, and RE Meyer, 2nd ed., 281–310. SAGE Publications.
- Scott, WR. 1995. *Organizations and Institutions*. Thousand Oaks, CA: Sage.
- Shjarback, J, S Decker, ... JJ Rojek - ... An International Journal, and Undefined 2017. 2017. "Minority Representation in Policing and Racial Profiling: A Test of Representative Bureaucracy vs Community Context." *Policing: An International Journal of Police Strategies & Management* 40 (4):748–67.
- Terrill, W, and Jason R Ingram. 2016. "Citizen Complaints Against the Police: An Eight City Examination." *Police Quarterly* 19 (2):150–79. <https://doi.org/10.1177/1098611115613320>.
- Thelen, K., and J. Conran. 2016. "Institutional Change." In *The Oxford Handbook of Historical Institutionalism*, edited by O Fioretos, TG Falleti, and A Sheingate, 51–70. Oxford University Press.
- U.S. Census Bureau. 2014. "American FactFinder 5-Year Estimates." American Community Survey. 2014.
- Vitale, AS. 2017. *The End of Policing*. Verso Books.
- Walker, S. 2001. *Police Accountability: The Role of Citizen Oversight*. Belmont, CA:

- Wadsworth Publishing.
- . 2006. “The History of the Citizen Oversight.” In *Citizen Oversight of Law Enforcement*, edited by JC Perino, 1–10. American Bar Association.
- Walker, S, and C Archbold. 2014. *The New World of Police Accountability*. SAGE Publications.
- Walker, S, and V Bumphus. 1992. “The Effectiveness of Civilian Review: Observations on Recent Trends and New Issues Regarding the Civilian Review of the Police.” *American Journal of Police* 11 (1).
- Walker, S, and E Luna. 2000. “Institutional Structure vs. Political Will: Albuquerque as a Case Study in the Effectiveness of the Civilian Oversight of the Police.” In *Civilian Oversight of Policing: Governance, Democracy, and Human Rights*, 83–104. Hart Pub.
- Walker, Samuel. 2006. “Police Accountability: Current Issues and Research Needs.” *NIJ Policing Research Workshop: Planning for the Future*.
- Ward, JD, and MA Rivera. 2014. “Institutional Racism, Organizations & Public Policy.”
- Weisburd, David, and Malay K. Majmundar. 2018. “Proactive Policing: Effects on Crime and Communities.” *The National Academies Press*. <https://doi.org/10.17226/24928>.
- Wilkins, Vicky M., and Brian N. Williams. 2008. “Black or Blue: Racial Profiling and Representative Bureaucracy.” *Public Administration Review* 68 (4):654–64.
<https://doi.org/10.1111/j.1540-6210.2008.00905.x>.
- Wooldridge, JM. 1999. “Distribution-Free Estimation of Some Nonlinear Panel Data Models.” *Journal of Econometrics* 90 (1):77–97.
- Worden, RE. 2008. “Citizen Oversight of the Ashton Police : Perceptions of Police Clients and Complainants, 2003.”
- Zhao, Jihong, Ni He, and Nicholas Lovrich. 2006. “The Effect of Local Political Culture on

Policing Behaviors in the 1990s: A Retest of Wilson's Theory in More Contemporary Times." *Journal of Criminal Justice* 34 (6):569–78.

<https://doi.org/10.1016/j.jcrimjus.2006.09.011>.

Zhao, Jihong, Ling Ren, and Nicholas Lovrich. 2012. "Political Culture Versus Socioeconomic Approaches to Predicting Police Strength in U.S. Police Agencies: Results of a Longitudinal Study, 1993 to 2003." *Crime & Delinquency* 58 (2):167–95.

<https://doi.org/10.1177/0011128708318947>.

Figures

		Discretion Afforded by Existing Institutions	
		Low	High
Characteristics of the Political Context	Strong Veto Possibilities	Drift	Layering
	Weak Veto Possibilities	Displacement	Conversion

Figure 1: Types of Institutional Change under various combinations of political contexts and discretion afforded by existing institutions

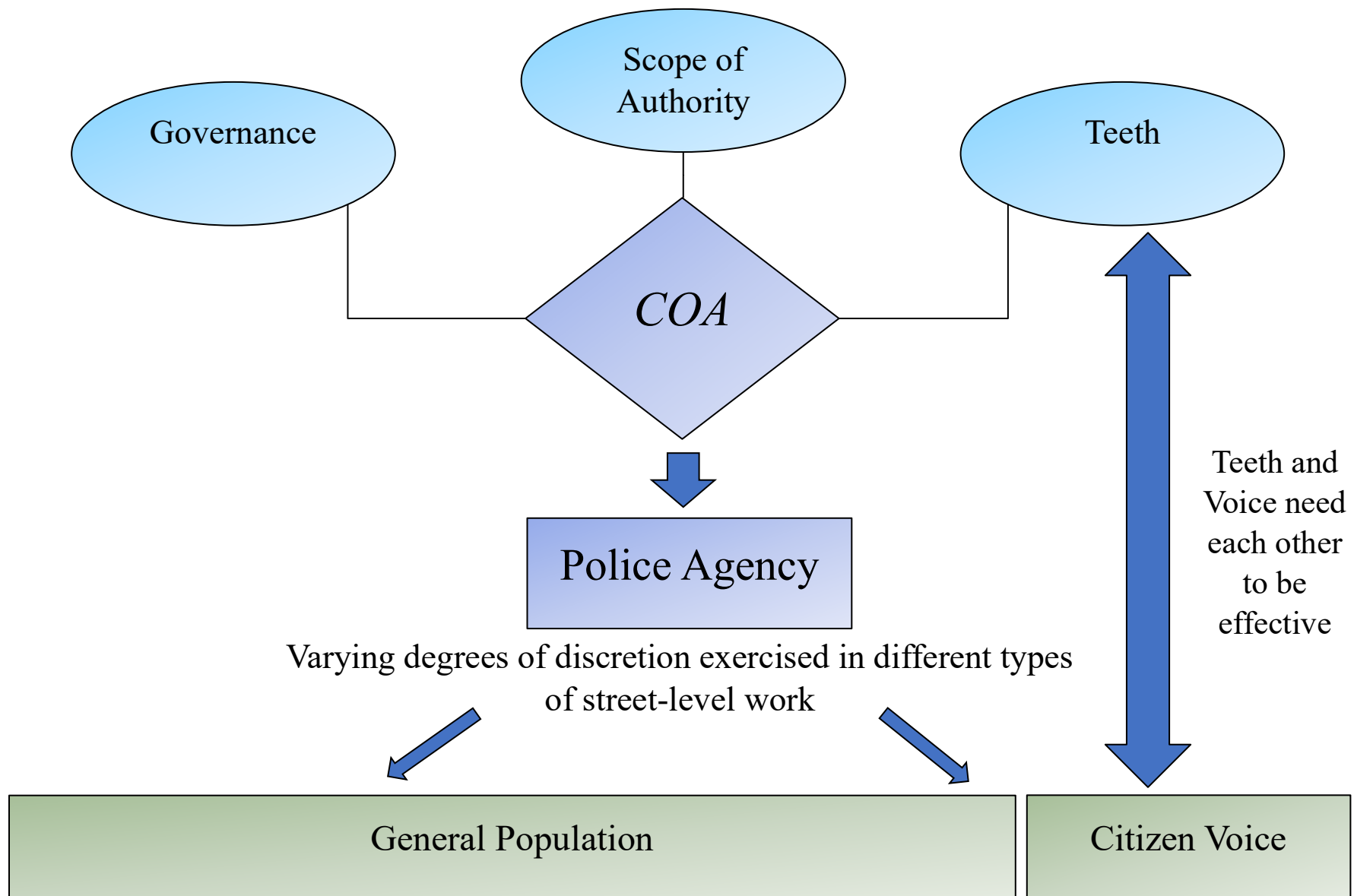


Figure 2: Schematic diagram of the factors hypothesized to affect COA effectiveness

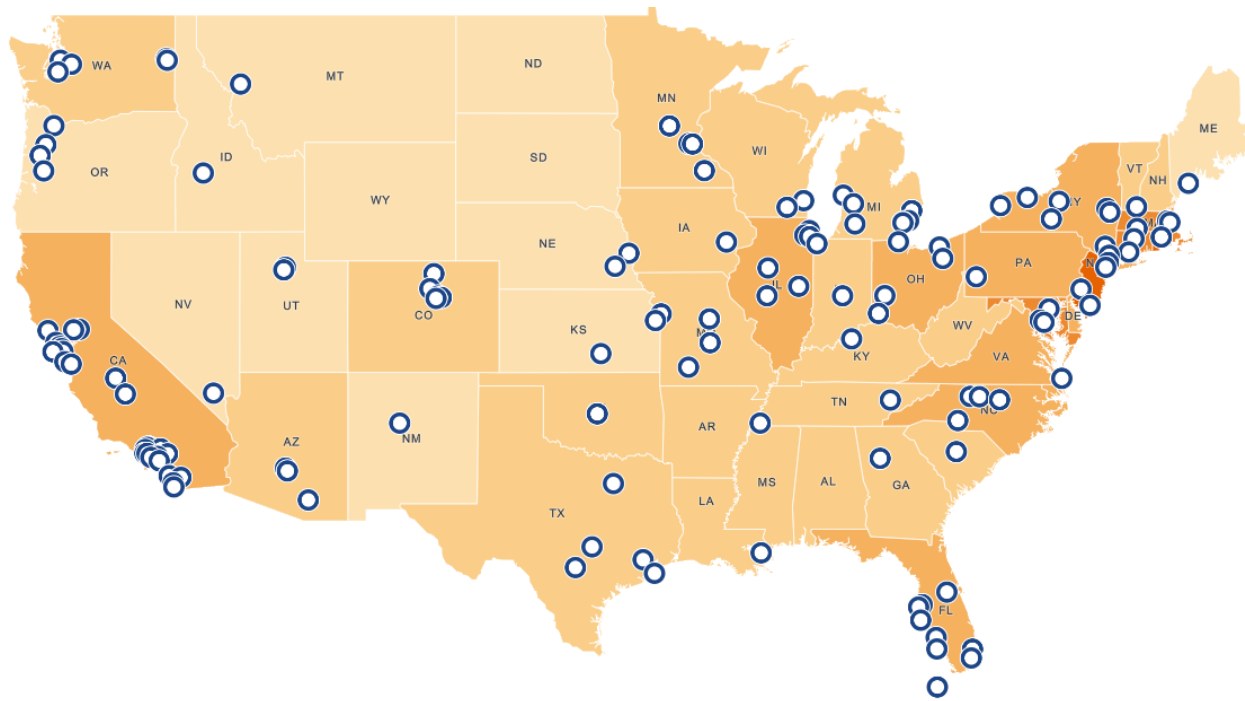


Figure 3: Geographic Dispersion of COAs on contiguous USA

Tables

Table 1: Variables and data sources

Variable	Type	Source
Disorderly conduct arrest rates by race	Panel; 1980-2014	Uniform Crime Reports (UCR), and through freedom of information (FOIA) requests from Illinois, Washington D.C., and selected cities in Washington and Massachusetts.
Police homicides of citizens by race	Panel; 2000-2014	Fatalencounters.org
COA year of creation, scope of authority, teeth, and governance	Cross-sectional. Assumed to remain time-invariant	Online Survey, administered through Qualtrics.
Voice (i.e., number of civil rights organizations in jurisdiction)	Panel; 1980-2014	Internal Revenue Service Masterfile
Alternative accountability mechanism dummy	Cross-sectional. Assumed to remain time-invariant	Online Survey, administered through Qualtrics.
Consent decree dummy	Panel; 1980-2014	U.S. Department of Justice website
Per capita income	Panel; 1980-2014	U.S. Census, American Community Survey 5-year estimates
Unemployment	Panel; 1980-2014	U.S. Census, American Community Survey 5-year estimates
Percentage of 25+ population with bachelors degree	Panel; 1980-2014	U.S. Census, American Community Survey 5-year estimates
Percentage of population that is black	Panel; 1980-2014	U.S. Census, American Community Survey 5-year estimates
Violent Crime Rate	Panel; 1980-2014	UCR

Table 2 Descriptive Statistics

Variable	Mean	SD	Min	Max
Racial disparity in disorderly conduct arrests in pre-period (per 100,000)	487.49	722.26	-654.16	8615.85
Racial disparity in disorderly conduct arrests in post-period (per 100,000)	352.24	695.87	-3533.90	5357.12
Racial disparity in police homicides in pre-period (per 100,000)	0.42	2.32	-8.89	27.80
Racial disparity in police homicides in post-period (per 100,000)	0.51	2.76	-5.08	58.86
Percentage of COAs with authority to classify citizen-initiated complaint	48.89	50.26		
Percentage of COAs with authority to review police complaint investigations	83.69	37.12		
Percentage of COAs with authority to conduct independent investigations	46.73	50.16		
Percentage of COAs with authority to audit/monitor police complaint investigations	57.30	49.74		
Percentage of COAs with authority to recommend/issue investigation findings to police	68.90	47.60		
Percentage of COAs with authority to access IA electronic databases	37.36	48.64		
Percentage of COAs with authority to recommend discipline and policy change	20.91	40.85		
Percentage of COAs with by-district governance	25.00	39.73		
Percentage of jurisdictions with at least one alternate accountability mechanism	64.54	48.05		
Percentage of jurisdictions under consent decree or court oversight	9.09	28.87		
COA budget (\$)	523,528.50	1,223,673.00	0.00	8,460,483.00
Number of full-time paid staff	9.18	14.32	0.00	70.00
Percent population black	20.48	18.19	.0026	85.24
Per capita income (\$)	19,264.65	8,691.47	5352.00	83387.00
Unemployment rate	7.18	3.38	1.00	32.50
Percentage of population 25+ with bachelors degree	29.59	14.15	6.37	81.60
Violent crime rate	990.70	671.04	16.70	4352.80

Table 3: Racial Disparity in Disorderly Conduct Arrest Rates - Model Estimates

Independent Variables	Racial Disparity in Disorderly Conduct Arrest Rates					
	(1)	(2)	(3)	(4)	(5)	(6)
Created	-64.04					
Age	-16.41**	-14.76	-15.68*	-16.24*	-18.80**	-20.67**
<i>Scope of Authority</i>						
Investigation score		-124.48^	-104.37^	-101.52^	-150.28**	-121.80*
Audit-Investigation score		11.62	12.07	-5.12	37.46	89.14
Review-Audit score		54.88	58.90	66.14	88.36	-25.73
Investigation score \times Age					6.12**	7.07**
Audit/Investigation score \times Age					-5.69^	-6.30**
Review-Audit \times Age					-3.49	4.90^
<i>Teeth-Voice Interaction Terms</i>						
Teeth Level-I Dummy \times Voice			30.72	34.43	43.22	-1.24
Teeth Level-II Dummy \times Voice			19.5	22.84	24.15	-11.41
<i>Voice</i>						
Number of Civil Rights Organizations			-37.72	-41.21	-41.72	-2.46
<i>Governance</i>						
By-district governance				-171.67*	-164.86*	-200.28**
<i>All Other Covariates Included</i>						
	No	No	No	No	No	Yes
<i>Year and Jurisdiction-specific Fixed Effects Included</i>						
	Yes	Yes	Yes	Yes	Yes	Yes
<i>Standard Errors clustered at Jurisdiction level</i>						
	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,226	2,949	2,949	2,949	2,949	2,770
BIC	49,650.1	45,455.98	45,468.69	45,468.75	45,475.65	42,444.13

^ $p < .15$; * $p < .1$; ** $p < .05$; *** $p < .01$

Table 4: Racial Disparity in Police Homicides of Citizens per 100,000 persons- Model Estimates

Independent Variables	Racial Disparity in Police Homicides of Citizens per 100,000 persons					
	(1B)	(2B)	(3B)	(4B)	(5B)	(6B)
Created	-.52					
Age	.03	.02	.02	-.03	-.01	-.03
<i>Scope of Authority</i>						
Investigation score		-.49	-.66 [^]	-.69	-.57	-.37
Monitoring score		-.40	-.69 [^]	-.61 [^]	-.57 [^]	-.99 [^]
Review/Audit score		-.03	.04	.20	.21	.71
Investigation score × Age					-.06***	-.06**
Monitoring score × Age					-.03 [^]	-.03
Review/Audit × Age					-.02	-.04
<i>Teeth-Voice Interaction Terms</i>						
Teeth Level-I Dummy × Voice			.13	.29*	.26 [^]	.40 [^]
Teeth Level-II Dummy × Voice			.18**	.25**	.26**	.29
<i>Voice</i>						
Number of Civil Rights Organizations			-.17 [^]	-.28*	-.25*	-.17
<i>Governance</i>						
By-district governance dummy variable				-1.83*	-1.76 [^]	-1.16
<i>All Other Covariates Included</i>	No	No	No	No	No	Yes
<i>Year and Jurisdiction-specific Fixed</i>	Yes	Yes	Yes	Yes	Yes	Yes
<i>Standard Errors clustered at Jurisdiction level</i>	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,110	895	895	895	895	671
BIC	2,966.63	2,116.34	2,128.14	2,100.37	2,105.91	1,666.33

[^] $p < .15$; * $p < .1$; ** $p < .05$; *** $p < .01$

Table 5: Summary of COA Impacts on Racial Disparity in Disorderly Conduct Arrests and Police Homicides of Citizens

		Dependent Variable	
		Reduces racial disparity in DCA?	Reduces racial disparity in PHC?
Scope of Authority	Investigation-focused	Yes, over time. Potentially in the year of creation as well.	Yes, over time
	Monitoring-focused	Yes, over time	No
	Review/Audit-focused	Yes, over time	No
Teeth-Voice Interaction	Level-I teeth × Voice	No	No
	Level-II teeth × Voice	No	No
Governance	By municipal district	Yes	No

Appendix-I

The various types of institutional change are described as follows:

Displacement: The lower-left quadrant, displacement, refers to the institutional change that is likely to occur when the political context is conducive, and existing institutions afford agents limited discretion. The source of change is the political context rather than ambiguous existing institutions. If stakeholders in the political environment do not offer sufficient resistance to institutional challengers, then change is most likely to occur due to existing institutions being simply replaced by new institutions.

Conversion: The lower-right quadrant, refers to the type of institutional change that is likely to occur when the political context is conducive, and existing institutions afford agents discretion to reinterpret institutions in a different way. In such a scenario, existing institutions do not change, however, change agents exploit the inherent ambiguity in the meaning of the institutions to enforce them in different ways, which in turn changes institutional outcomes.

Layering: The upper-right quadrant, refers to the institutional change that is likely to occur when veto points are strong, and existing institutions afford agents broad discretion. In this scenario, change may occur in a piecemeal fashion by additions or modifications at the margins of existing institutions. Institutional outcomes change because of interaction between existing institutions and amendments/additions to those institutions.

Drift: In the upper-left quadrant, drift refers to the type of institutional change that is likely to occur when veto points are strong, and existing institutions afford actors little discretion to reinterpret rules. In such a scenario, the formal institution again does not change, but its substantive impact changes due to shifts in the external environment and the failure to update institutions in accordance with changes in the sociopolitical environment.

Appendix-II

To determine whether specific COA authorities load on to specific latent factors, we conducted a factor analysis on responses to certain questions in the survey. Our goal was to classify agencies into categories based on the type of authorities they possessed. Varimax rotation was used to determine final factor loadings. We ran the factor analysis on the responses to the following thirteen questions/variables:

(1) Does your agency ever have the authority to:

- i. Classify the nature of a citizen-initiated complaint?
- ii. Review police complaint investigations (e.g., for thoroughness, completeness, accuracy)?
- iii. Conduct investigations that are independent of the police?
- iv. Audit and/or monitor police complaint investigations for compliance with investigative standards?
- v. Send complaint investigations back to the police for further investigation?
- vi. Recommend and/or issue investigation findings to the police?
- vii. Recommend discipline?
- viii. Evaluate and/or recommend changes in police policies, procedures, training, and/or management practice?
- ix. Analyze data to identify trends and patterns in police misconduct?
- x. Access closed internal affairs police files?
- xi. Access ALL internal affairs police files?
- xii. Access internal affairs electronic databases?

- (2) A recoded indicator that equaled one if a respondent reported that the COA had at least one paid full-time employee, and zero otherwise.

Table A2.1 and A2.2 show the output obtained of the factor analysis procedure. We retained the first three factors as their eigenvalues are greater than 1, and they cumulatively explain 92% of the variation in the thirteen variables. The output shows that the above authorities load on to the following three factors:

- a. Factor 1: Classify citizen-initiated complaints (factor loading=0.76), conduct investigations that are independent of the police (0.69), recommend and/or issue investigation findings to the police (0.55), recommend discipline (0.46), and the indicator for paid, full-time staff (0.54).
- b. Factor 2: Audit/monitor police complaint investigations for compliance with investigative standards (0.26), analyze data on trends and patterns in police misconduct (0.26), access all internal affairs files (0.74), access closed internal affairs files (0.45), access internal affairs electronic databases (0.69), evaluate and/or recommend changes in police policies (0.38), and the indicator for full-time paid staff (0.30).
- c. Factor 3: Review police complaint investigations (0.63), audit/monitor police complaint investigations for compliance with investigative standards (0.66), access closed internal affairs files (0.52).

Based on our reading of the literature, especially Angelis, Rosenthal, and Buchner (2016), the first factor appears to be a typical representation of an investigative COA, while the second and third factors appear to be consistent with monitoring and review/audit COA. Thus, we named the latent factors for the authorities which seemed to load onto them. We then predicted the factor scores for the above three factors using all variables, except for the ability to recommend

discipline, and the ability to evaluate and/or recommend changes in police policies, procedures, training, and/or management practice. We did not use the above two variables while predicting the factors scores because these variables comprise agency teeth, whose impact we wanted to test separately.

The predicted factor scores for each agency, thus represent an index that denotes the extent to which it emphasizes an investigatory, monitoring, or review/audit focused approach towards police oversight. These predicted factors scores were used in the regression models as independent variables.

Table A2.1 Factor Analysis

Method: Principal Factors

Number of obs: 80

Retained Factors 8

Number of parameters 76

Factor	Variance	Difference	Proportion	Cumulative
Factor 1	2.04	.32	.37	.37
Factor 2	1.72	.42	.31	.69
Factor 3	1.30	.57	.24	.92
Factor 4	.73	.40	.13	1.06
Factor 5	.33	.16	.06	1.12
Factor 6	.17	.06	.03	1.15
Factor 7	.10	.03	.02	1.17
Factor 8	.07		.01	1.18

LR Test: independent vs. saturated: Chi-squared(78)=320.93 Prob>Chi-squared=.0000

Table A2.2 Rotated Factor Loadings (pattern matrix) and unique variances

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Uniqueness
Classify citizen-initiated complaints	.76	.05	.02	.10	.02	-.04	-.02	-.06	.40
Review police complaint investigations	-.05	.08	.63	-.02	.18	-.02	-.02	-.06	.55
Conduct investigations that are independent of police	.69	.18	.08	.18	-.13	.08	-.07	.16	.40
Audit/monitor police complaint investigations for compliance with investigative standards	.10	.26	.66	.21	-.05	.03	.06	.06	.44
Send complaint investigations back to the police for further investigation	-.01	.13	.23	-.04	.42	-.01	.00	-.01	.75
Recommend and/or issue investigation findings to the police	.55	.07	-.05	-.01	.27	.03	.23	-.02	.56
Analyze data on trends and patterns in police misconduct	.27	.26	.13	.60	-.05	.01	-.03	.04	.47
Access all IA files	.08	.74	.17	.11	-.00	.07	-.04	-.08	.39
Access closed IA files	.18	.45	.52	-.00	-.06	.12	-.12	-.00	.47
Access IA electronic databases	.20	.69	.14	.16	.07	-.04	.09	.11	.41
Evaluate and/or recommend changes in police policies, procedures, training and/or management practices	.11	.38	.22	.12	.04	.29	-.02	.02	.69
Recommend discipline	.46	.24	.10	-.21	-.09	.23	.09	-.02	.61
Indicator for paid full-time staff	.54	.30	0.12	.43	.09	.01	.08	-.13	.39

Appendix-III

To test whether cities established COAs in response to sudden spikes in the dependent variables, we regressed equation (1), but instead of study variables of interest, we included a series of four indicator variables that indicated the time distance from the year in which the COA was created. None of the time leads were significant at the 5% level in the regression for either dependent variable. However, the one-period time lead was negative and significant at 10% in the equation for racial disparity in DCA. However, significance tests to check whether the time leads were simultaneously different from zero failed to reject the null that the time leads were simultaneously equal to zero ($p=0.218$ for the disparity in DCA, and $p=0.731$ for the disparity in PHC). This suggests that, on average, jurisdictions did not self-select into establishing COAs due to a sudden change in the outcomes of interest for this study. Estimates from these models are shown below.

Table A3.1 Test for Relationship between Time Distance of COA Creation and Racial disparity in Disorderly Conduct Arrests and Police Homicides of Citizens

	Black-White Disparity in Disorderly Conduct Arrest Rate	Black-White Disparity in Police Homicide Rate
Time Distance Dummies		
<i>T minus 1</i>	-193.44*	-1.10
<i>T minus 2</i>	-127.53	-.63
<i>T minus 3</i>	-2.84	-1.06
<i>T minus 4</i>	-29.86	-.16
Percent Population Black	-7.99	-0.15
Per Capita Income	-0.01	-0.0004
Unemployment	-37.40*	-0.05
Percentage 25+ w/ Bachelors Degree	-32.35	-0.11
Violent Crime Rate	0.08	.001
Investigation or Oversight Decree	-213.89	15.27***
Black White Disparity in DCA		-.0009
Year and Jurisdiction-specific Fixed Effects	Yes	Yes
Standard errors clustered at Jurisdiction level	Yes	Yes
<i>N</i>	1,636	117

$\wedge p < .15$; * $p < .1$; ** $p < .05$; *** $p < .01$

Appendix-IV

We attempted to verify the fundamental identifying assumption of a fixed effects design, per which pre-treatment trends (in the dependent variables) in municipalities that established a COA should be the same as the pre-treatment trends in cities that did not establish a COA. While the pre-treatment trends in DCA were different for treated and untreated cities in six out of thirty-five years, a joint test comparing pre-treatment trends of treated and untreated cities found that trends in racial disparity in DCAs were statistically indistinguishable ($p=0.204$). Regarding PHCs, we found that trends were statistically indistinguishable in treated and untreated cities in fourteen out of fifteen years at the 10% significance level, which is lower than what would be expected by chance alone. Overall, the above two tests indicate that there is little or no self-selection into the creation of COAs. Estimates from these models are shown below.

Table A4.1 Test for Difference in Trends of Disorderly Conduct Arrests and Police Homicide Rate Between Treated and Untreated Jurisdictions

	Black-White Disparity in Disorderly Conduct Arrests	Black-White Disparity in Police Homicide Rate
<i>Treatment Dummy-Year interaction</i>		
Treated × 1980	9.28	
Treated × 1981	54.02	
Treated × 1982	43.13	
Treated × 1983	-29.43	
Treated × 1984	-15.71	
Treated × 1985	-22.11	
Treated × 1986	-1.03	
Treated × 1987	-23.20	
Treated × 1988	50.32	
Treated × 1989	37.99	
Treated × 1990	-18.96	
Treated × 1991	78.93	
Treated × 1992	44.31	
Treated × 1993	48.30	
Treated × 1994	-23.85	
Treated × 1995	152.46	
Treated × 1996	73.50	
Treated × 1997	20.81	
Treated × 1998	85.46	
Treated × 1999	154.60	
Treated × 2000	147.65	1.35
Treated × 2001	248.99	-.12
Treated × 2002	216.04	.16
Treated × 2003	375.89*	.65
Treated × 2004	654.24**	-1.23
Treated × 2005	395.55*	-1.28
Treated × 2006	310.66	-1.40
Treated × 2007	221.32	-1.97
Treated × 2008	705.60	.74
Treated × 2009	563.31	-2.45*

Treated × 2010	448.53*	-1.25
Treated × 2011	379.99*	-.06
Treated × 2012	495.49*	-.70
Treated × 2013	466.38	-.13
Treated × 2014	827.94	.73
Black White Disparity in DCA		-.0004
All Covariates Included	Yes	Yes
Year and Jurisdiction-specific Fixed Effects	Yes	Yes
Standard errors clustered at Jurisdiction level	Yes	Yes
<i>N</i>	6,111	2,112

*^p<.15; *p<.1; **p<.05; ***p<.01*