The “Guarding the Guardians” Problem: An Analysis of the Organizational Performance of an Internal Affairs Division

Sandro Cabral*
School of Management, Federal University of Bahia (UFBA)
Av. Reitor Miguel Calmon, s/n – 3. andar
Salvador, Bahia, Brazil, 40110-903
Phone/Fax : +55 71 3283 7302/7659
E-mail: scabral@ufba.br

Sérgio G. Lazzarini
Insper Institute of Education and Research
R. Quatá, 300
São Paulo, SP Brazil 04546-042
E-mail: SergioGL1@insper.edu.br

Paper accepted for publication in the Journal of Public Administration Research and Theory

(JPART) – October, 2013

* Corresponding Author

FUNDING: This article was supported by the Brazilian National Council for Scientific and Technological Development (CNPq) under grants 401319/2010-5 and 306378/2012-4, by CAPES Foundation under the travel grant AEX 0711/10-1, and by Bahia State Research Foundation (FAPESB) under the travel grant 2000/2010

ACKNOWLEDGEMENTS: We thank the suggestions by Andrew Whitford (editor) and there anonymous referees. We also thank the comments by Dan Elfenbein, Felipe Fernandes, Ricard Gil, Pierre Kopp, Tammy Landau, Adam McCaulay, Anita McGahan, Gary Miller, Aldo Musacchio, Armando C. Pinheiro, Ben Ross Schneider, Judith Tendler, Graham Willis, Todd Zenger, and seminar participants at Academy of Management (Montreal), EESP/FGV, Federal University of Bahia, Federal University of Alagoas, Harvard-MIT, Insper, New York University, RWIO (Rio de Janeiro), University of Brasilia, University of Paris 1 – Sorbonne, University of Toronto, and Washington University. Special thanks to the police officers who helped us in our data collection. All remaining errors and omissions are our own
The “Guarding the Guardians” Problem: An Analysis of the Organizational Performance of an Internal Affairs Division

ABSTRACT

Public administration scholars have proposed that efficient monitoring and accountability are key performance elements in public services. We contribute to this debate by focusing on the so-called “guarding the guardians” dilemma: monitoring commissions made up of public officials who may refrain from punishing their peers and turn a blind eye to certain deviations. In the police, this is the case of internal affairs divisions (IADs), which are organizations crafted to monitor the behavior of police officers. We expand the existing literature by proposing a set of hypotheses that are tested using a distinct dataset containing detailed information on investigations against police officers performed by an IAD. We assess the organizational performance of the IAD by examining how various factors influence competing hazards of acquittal, mild and harsh punishment of the implicated officers. We find that high-ranking, high-tenured officers are less harshly punished with termination of duty, while the opposite effect is found with respect to cases with extensive media coverage. We also find that cases judged by specialized investigation officers increase the likelihood of punishment compared to cases assigned to temporary, rotating commissions, especially for cases with a long time span.

Key words: guarding the guardians, organizational performance, internal affairs, police

“I don’t see how the division can investigate itself”

Al Pacino’s line in the movie Serpico

“… it would be ridiculous for a guardian to need a guardian himself!”

Plato, Republic (translated by C. Reeve)

1. Introduction

Public bureaucracies have increasingly been called upon to pursue the appropriate execution and continuous improvement of their services. However, although the vast majority of citizens throughout the world are in close contact with public services (such as education, health care, and law enforcement), the literature on public service performance is in its infancy (Walker, Boyne and Brewer 2010). Only recently have public administration scholars started analyzing the circumstances in which the strategies chosen by public managers are more likely to succeed (Boyne and Walker 2010). In particular, scholars have underscored the importance of measures to improve accountability, as well as internal procedures to guarantee
that the execution of public services does not violate civil rights, human rights, and probity (Walker et al 2010; Bovens 2010).

A strand in the literature on public administration deals more specifically on how organizational choices can improve monitoring and accountability. Moe (1984, 1994) pioneered the application of economic theories of organization to explain the behavior of public bureaucracies. His discussion, as well as subsequent work (Hart, Shleifer and Vishny 1997; Kim 2005; Miller and Whitford 2007), emphasizes two key difficulties in crafting effective control mechanisms in the context of public services. First, the assessment of performance is usually complex and subjective, therefore rendering incentive schemes such as performance-contingent payments more difficult to implement (Dixit 2002). Furthermore, the multiplicity and diffuse nature of “principals” (citizens, government officials, nongovernmental organizations and so on) greatly attenuates the intensity and effectiveness of direct monitoring of public “agents” (Kim 2005; Romzek 2000). Given these challenges, some argue that improved performance in the public sector requires oversight by independent governmental agencies responsible for monitoring and punishing deviations (Romzek and Ingraham 2000; Forrer, Kee, Newcomer and Boyer 2010).

Although this literature has improved our understanding of how to improve public sector performance, a critical dilemma arises: what guarantees that public bureaucrats will effectively monitor other public bureaucrats? This “guarding the guardians” problem is particularly relevant when monitoring is performed internally, through accountability bodies established within a given public bureaucracy—as is exemplified by the police and their so-called internal affairs divisions (IAD), which represent the empirical context of our paper. In several instances, judgment and punishment tend to be in the hands of the very officers who are supposed to be monitored (Prenzler 2000). To attenuate the problem, some police organizations appoint specialized officers to carry out IAD investigations. Because they work
separately from other officers working in regular police activities, there is a separation of roles between *agent-executers* and *agent-monitors*. However, even when IAD officers are specialized in their monitoring function, they may have been former police officers (agent-executers) or may return to normal police duties after leaving the IAD. Therefore, there are reasons to believe that investigations will be inherently biased towards certain officers and certain types of deviations (Liederbach, Boyd, Taylor and Kawucha 2007), thus undermining the performance of the accountability body.

Building on this dilemma, we outline a set of hypotheses that help expand the extant literature on the performance of oversight agencies of organizations affected by the guarding the guardians problem. We propose that the organizational performance of oversight units will crucially depend on three factors: how the unit is organized; the position of the accused officer in the hierarchy (which should influence his or her ability to influence outcomes); and accusation-specific factors whose effect will be dependent on internal norms and the salience of the deviation. We test our theory using a new and distinctive database containing detailed information on investigations against police officers in the IAD of a police organization located in a particular state in Brazil. The uniqueness of our database stems from its detailed information on internal investigations and its potential determinants. We also follow a competing risk approach by jointly modeling the duration of the investigation and the possible outcomes that can result from the final judgment (namely, whether the case resulted in acquittal of the implicated officer or not, and the harshness of the associated punishment).

In our empirical analysis, we operationalize *organizational performance* as the capacity of the oversight body to punish deviants when appropriate. The final judgment and punishment also depend on the extent to which police officers are unable to obstruct or thwart investigations. Therefore, the duration of the case and the final punishment outcome are key dimensions affecting the efficacy of oversight and accountability. Thus, they are related to
what Walker et al. (2010) refer to as the governance domain of organizational performance in which probity, accountability, and civil rights are important indicators of how governments provide goods and services addressing citizens’ needs.

In this sense, our contribution to the literature is twofold. First, by understanding the factors that affect the likelihood of sanctions against deviants we can shed some light on the debate on the design of oversight and control mechanisms in the public sector and their implications on organizational performance in its governance domain, which remains an underexplored field (Andres, Guasch, and Azumedi 2009; Walker et al 2010). In line with Rabovsky (2012), our study informs policymakers on the connections between accountability structures and the administrative responses of public bureaucracies. Second, although there is vast anecdotal evidence of police misconduct and biased judgment in several countries (Human Rights Watch 2009; Klockars, Kutnjak and Haberfeld 2006; Punch 2009), systematic empirical research on IAD and other watchdog agencies has been surprisingly scant. Even though our study focuses on a particular type of public bureaucracy, similar issues emanating from the guarding the guardians problem also appear in diverse organizational contexts such as legislatures (Capeletti 1983), self-management teams (Alchian and Demsetz 1972), regulatory agencies (Stigler 1971), governmental bureaus (Roberts, Dull and Choi, 2010), hybrid public-private arrangements (Cabral, Lazzarini and Azevedo 2010), among several others (see, for a general discussion, Hurwicz 2008).

Our paper is structured as follows. The next section discusses the theoretical underpinning of the guarding the guardians problem, which is then used to generate new hypotheses applied to the context of IADs. We then describe our empirical setting and data. The econometric tests are subsequently presented and discussed. In a nutshell, we find that officer-specific characteristics such as an officer’s position in the hierarchy and the nature of the accusation significantly affect the final verdict. We also find that the organizational
design of the investigation affects organizational performance: for cases with a long time
span, there is a positive effect on the likelihood of punishment associated with the separation
of roles between agent-executers and agent-monitors (i.e. officers specialized in the
monitoring task). Our results unveil a complex interplay of factors influencing organizational
performance and control mechanisms in public services plagued by the guarding the
guardians problem. We conclude with implications for theory and public policy.

2. Theory: Regulation, Delegation, and Organizational Performance of Public
Bureaucracies plagued by the Guarding the Guardians Problem

Public administration scholars have devoted increasing attention to the determinants of public
service performance (Behn 2003; Boyne, Meier, O’Toole and Walker 2006; Walker et al
2010). After several reforms in the public sector, research has evolved to analyze the impact
of the implemented strategies and to understand key factors driving successful initiatives
(Boyne 2003). In recent decades, governmental action to improve the performance of public
services has emerged in a scenario of increased calls for the effective regulation and oversight
of civil servants. We discuss below how organizational features of public bureaus should
increase the effectiveness of regulation and potentially lead to superior performance.

2.1. Regulation

The New Public Management (NPM) paradigm has been characterized by intense
decentralization, incentives for civil servants, increased competition between public and
private providers, and goal-oriented governments (Barzelay, 2000). Although the number of
civil servants involved in the delivery of public services has dropped, NPM proposes an
increase in the regulation of the government in terms of formality, complexity, and
specialization (Hood et al 1999). This process has given rise to an “audit society,” in which
the costs of regulation are non-negligible (Power, 1997).
Besides the traditional regulation apparatus through courts and legislatures, we have witnessed the emergence of specialized bureaus working as oversight agencies (secondary regulation; e.g. Hood, James, and Scott 2000). As a way to improve oversight, governments often create dedicated units or functional supervision tasks. In more sophisticated arrangements, such monitoring units will present an official mandate for the overseer; a clear orientation of the regulatory organization to influence public activities; and a clear separation of roles between the regulator and the subjects affected by the regulation (Hood et al 1999). Namely, some public agents will even specialize in monitoring other agents who are more directly involved in the execution of public services. We refer to this as a separation of roles between agent-monitors and agent-executers.

Examples of agent-monitors include inspectors supervising governmental agencies (Roberts, Dull and Choi 2010) and officers holding fixed, specialized positions in internal affairs divisions of the police (Stone 2007). With separated roles, agents-monitors with a strong sense of professionalism (Miller 2000) should more effectively develop oversight capabilities and become more insulated from other agents who may eventually be investigated and punished.

Yet this separation of roles is not enough to guarantee effective oversight. According to agency theory, when decisions are decentralized to agents (public or otherwise) whose objectives may not be aligned with principals, poor execution and performance will probably result (Fama and Jensen 1983; Jensen and Meckling 1976). If individual output can be objectively measured, principals can mitigate this conflict by crafting incentive contracts whereby agents are compensated or promoted according to some observable performance indicator. Another possibility to mitigate agency hazards is to directly monitor the behavior of agents and apply sanctions to those who deviate—including, in an extreme case, termination.
of the employment contract (for a review see Bolton and Dewatripont 2005). However, as we discuss below, these two mechanisms are more difficult to implement in public bureaucracies.

2.2. Delegation of the Oversight Function

Many activities in the public sector involve multiple principals, with multiple objectives and dispersed in various domains: the public in general, politicians, elected officials, non-governmental organizations, and so on (Dixit 2002). Unlike for-profit firms, there is no restricted group of principals, such as shareholders, with direct monitoring duties. Along these lines, Moe (1984) argues that external control mechanisms such as stock market evaluation and the threat of takeovers are absent for public bureaucracies, making the efficient governance of public bureaucracies even more difficult to obtain. In a context of delegation, supervision of public bureaucrats will therefore generally be performed by other public bureaucrats, with limited ability and even willingness to control the behavior of their peers (Wilson and Rachal 1977).

These features of public organizations will thus induce a fundamental “guarding the guardians” dilemma, originally addressed by Plato in *The Republic* (see Hurwicz 2008). With scant incentives and poor monitoring, who will guarantee that rules within the public sector will be effectively enforced? In other words, who will guard the guardians?

Efficient monitoring in public bureaucracies also requires special attention because the civil service may attract individuals who are not necessarily aligned with broader societal goals. Although Miller (2000) submits that delegation to bureaucrats with a sense of professionalism is an important feature of public organizations, there is a risk that adverse selection of unaligned officers and reduced effort due to moral hazard will severely undermine performance (Moe 1984). Furthermore, because in some cases agent-monitors are appointed from the pool of agent-executers and can possibly return to their normal execution activities in the future, they may refrain from punishing deviants and even be susceptible to
influence from their accused colleagues. Under these circumstances, the capture of the oversight unit will probably occur (Wilson 1989). In addition, norms and social ties within the organization may reduce the likelihood of conviction. In the police, for instance, some even talk about an implicit “code of silence” (Skolnick 2002). Officers who try to enforce the rule of law by “blowing the whistle” are often ostracized and subject to retaliation by colleagues (Punch 2000; Rotwell and Baldwin 2007). Thus, it is crucial to examine factors that could influence investigation outcomes, even in cases where specialized oversight bodies are in place.

2.3. Organizational performance

In public bureaucracies, the solution to the delegation problem becomes far more complicated. Performance metrics are oftentimes complex and multifaceted; they often involve elements that are difficult to objectively define and measure (Andrews, Boyne and Walker 2006). For instance, prisons must guarantee an adequate “treatment” of inmates and promote security (Cabral et al. 2010); public schools must strive for effective “education” (Acemoglu, Kremer and Mian 2008); police organizations must reduce crime rates while at the same time avoiding excessive use of “force” (Dick 2005). The presence of goal ambiguity in public organizations creates additional difficulties for crafting clear performance criteria (Chun and Rainey 2005). Complex bureaucracies, with a large number of different echelons (Aiken and Hage 1968), also magnify the risk of having different groups of officers with distinct opinions on sensitive aspects of the organization (Walker and Brewer 2008). These aspects typically induce low-powered incentives in public organizations (Miller and Whitford 2007; Williamson, 1999) and mandate increased oversight: when members of public organizations have different perceptions of the required goals and objectives, there is greater need for procedural regulations and centralization (Chun and Rainey 2005).
Although several dimensions of organizational performance have been proposed (see e.g. Boyne 2002, Boyne et al 2006), we focus on the governance performance domain, which refers to the ability of public bureaucracies to improve their accountability and monitor deviations from desired objectives as well as general principles of civil rights, human rights and probity (Walker et al 2010). The governance domain of organizational performance becomes particularly relevant in the current context of decentralization of decision making in public services and increased pressure towards accountability. Scholars emphasize that there is a trade-off between hierarchical control and participation in decisions: if public servants have autonomy to execute their services, it is necessary to guarantee that execution follows societal goals (Andrews 2010, Moore 1995).

The assessment of governance performance is crucial to verify if agents involved in oversight activities are carrying out the needs of principals, especially considering that the existing measures of governance are not entirely adequate (Fukuyama 2013). In the case of oversight agencies, governance performance should express characteristics that show that the agency under scrutiny is accountable to the public and to other public agencies (Andres et al 2009). One measure of governance performance seems to be particularly meaningful in oversight agencies in charge of investigations against civil servants who allegedly misbehaved: the resulting outcomes of the investigation procedures. Indeed, impunity could jeopardize the legitimacy of the oversight bodies and, in this case, the capacity of the agency to achieve a proper level of accountability in due time can be challenged.

Although the issues discussed above are general and applicable to various contexts for expositional simplicity we now describe our particular empirical context focused on internal affairs divisions (IADs) and then propose some context-specific hypotheses to be tested.

3. Internal Affairs Divisions and Organizational Performance
An evolving discussion has centered on how to increase performance and accountability in law enforcement (Brehm and Gates 1993; Butterfield, Edwards and Woodall 2005). The act of monitoring police officers’ conduct is essential as they face situations in their daily activities which may prompt wrongdoing. In fact, in order to prevent crime and to enforce the rule of law, a police officer’s duties frequently involve the use of coercive force (Dick 2005); and due to their inherent proximity with criminal activity, police officers may be particularly prone to illicit arrangements. A brief inspection of the daily newspapers throughout the world unveils several cases of crimes committed by “criminals in uniform” (Punch 2000).

Stone (2007) argues that there are three levels of police accountability: internal, external and societal. Internal affairs divisions (IADs) correspond to the first sphere, while independent governmental bodies and civilian “watchdog” agencies represent external and societal monitoring mechanisms. IADs throughout the world basically monitor the actions and/or omissions of police officers that may violate laws or departmental rules. Namely, IADs investigate complaints against police officers, which may originate from diverse sources such as citizens, police department internal agents, and defense attorneys. Allegations of misconduct may be sustained or not; and, according to the IAD’s specific rules, several possible punishments can be applied, including an oral reprimand, suspension without pay, loss of promotional opportunity and termination of duty, among others (Klockars et al. 2006; State of New Jersey 2000).

Despite the fact that IADs throughout the world differ in terms of internal rules and criteria for appointing their members, the guarding the guardians dilemma prevails: there is a concern that investigations will be perfunctory or conducted in such a way as to favor certain police officers under investigation (Lewis 1999). Not surprisingly, complaints of biased investigations have emerged in diverse countries such as Australia (Prenzler 2000), Brazil (Human Rights Watch 2009), Canada (Landau 1996), the United Kingdom (Punch 2009;
Waters and Brown 2000), the Netherlands (Punch 2009) and the United States (Klockars et al. 2006; Lasley 1994), among several others. Because poor accountability standards are known to undermine the performance of public services (Andrews et al 2012, Moore 1995), failure to craft effective IADs can be a fundamental source of failed law enforcement.

Within this context, in order to ensure accountability we need to understand which factors could potentially affect the ability of IADs to monitor and enforce internal rules. Below we propose a set of hypotheses focused on three general factors: officer-specific characteristics related to their position and history in the police bureaucracy; internal organizational procedures adopted by IADs; and aspects related to the nature of the accusation. As noted before, we operationalize the performance of monitoring units based on the likelihood of acquittal and the severity of the punishment experienced by the accused officer (mild, such as temporary suspension, or harsh, such as termination of duty). Given that these possible outcomes are not independent (e.g. an officer who is suspended cannot be at the same time fired), we propose hypotheses establishing how each particular factor should simultaneously affect the outcomes of the investigation.

### 3.1. Officer-Specific Factors

Organizational scholars have long advocated that an individual’s position and experience in the hierarchy could affect his or her ability to influence decisions (Crozier 1963; Milgrom and Roberts 1990). In this sense, we expect that police officers in high positions or positions appointed by other eminent individuals in the force have a superior ability to influence bureaucrats involved in the analysis of internal cases. Likewise, agent-monitors working in IADs may refrain from recommending harsh punishment against such top officers out of fear of some form of retaliation. As in any other public or private organization, people in charge of verifying and denouncing wrongdoing can be harassed and ostracized (Miceli and Near 1994). Although external observers tend to think that high-status wrongdoers deserve more
severe punishment than low-status wrongdoers (Fragale, Rosen, Xu and Meredith 2009),
punishment may be less likely when it involves high-status members of the organization
(Near and Miceli 1987) and when potential “whistle blowers” believe that wrongdoers are
more influential in the hierarchy (Gundlach, Douglas and Martinko 2003).

Age, ranking, and experience in the police force may also influence the outcomes of
investigations. Ratcliffe, Biles, Green and Miller (2005) examined the incidence of drug-
related complaints against police officers and found that such cases are most likely to involve
lower ranking, younger police officers. In another study, McElvain and Kposowa (2004)
found that officers with five to nine years in the force were eight times more likely to be
investigated for the use of force than those with twenty or more years of experience. Similarly
to higher-ranking officers, experienced agents may use internal tactics to postpone
investigations so as to deteriorate the quality of the incriminating evidence (Chappe 2012).
We can also expect that punishment against those officers, if any, will be relatively mild.
Their ability to influence outcomes should also negatively affect the incidence of harsh
punishment. In addition, investigating commissions themselves may refrain from severely
punishing officers with hierarchical status or a long history in the organization (Klockars et al.
2006). Thus, considering acquittal, mild and harsh punishment as competing events, we
propose:

$$H_1 \text{ Cases involving high-ranking, high-tenured officers are associated with an increased likelihood of acquittal, a reduced likelihood of mild punishment and an even more reduced likelihood of harsh punishment.}$$

3.2. Organizational Features of the Investigation

IAD investigators are normally officers who have previous field experience and who in some
cases work exclusively on investigations of misconduct. Yet, there is substantial
heterogeneity in the way IADs are organized, even within a given country. While some IADs
act by simply investigating complaints that emerge in the police force, other agencies play
more proactive roles in prevention of misconduct. For example, the New York Police Department (NYPD) has promoted organizational innovations, such as the Investigative Review Unit (IRU), to perform a constant evaluation of the investigations carried out by the Internal Affairs Bureau (CCPC 2000). These units are composed of officers specialized in the monitoring function, with accumulated expertise in investigations and hearing procedures. Such expertise should increase the effectiveness of the monitoring system (see Boyne, Day and Walker 2002). Using our previous terminology, in this case there is a separation of roles between agent-executers and agent-monitors.

In some contexts, however, internal supervisory bodies may be precarious. Kelly (2003) shows that 87% of police departments in the United States consist of fewer than 25 sworn officers. Some departments may lack the organizational capability to conduct effective investigations. In some cases, a formal separation between agent-monitors and agent-executers is not observed: officers who are not familiar with the peculiarities of internal affairs investigations may be assigned to temporary investigative roles. Thus, even in countries with external oversight and stronger legal procedures to punish wrongdoing, one might find IAD structures that are neither specialized nor completely insulated from internal influence (Klockars et al 2006; Punch 2009). Such lack of separation could negatively affect the outcome of investigations. Yet we do not expect any difference in terms of the relative incidence of mild or harsh punishment because these outcomes will probably depend on officer- and accusation-specific factors. In other words:

\[ H_{2a} \text{ Separation of roles between agent monitors and agent executors is associated with a decrease in the likelihood of acquittal, and an increase in the likelihood of mild and harsh punishment.} \]

However, it is possible that certain accused officers with past or current ties to the current members of the investigation commissions will try to influence decisions. Indeed, as in other public bureaucracies, capture may occur when regulators have close ties with
regulatees (Boyne et al 2002). Organizational scholars have underscored the importance of the informal organization: individuals may create distinctive influence channels through interpersonal relationships and social cliques (Krackhardt and Hanson 1993). These relationships can represent a source of power beyond what is defined solely by the attributes of the function or by the agent’s position of the individual in the hierarchy (Pfeffer, 1992). Similar to our previous logic for the effect of high-ranking, high-tenured officers, ties emanating from previous interactions can be a mechanism through which policemen get privileged information about the ongoing evolution of internal processes in order to craft defensive strategies and even obstruct investigations through interpersonal influence, especially when harsh punishment is a possibility:

\[ H_{2b} \] Previous ties between investigated officers and commission members are associated with an increased likelihood of acquittal, a reduced likelihood of mild punishment and an even more reduced likelihood of harsh punishment.

\[ H_{2c} \] Instances where investigated officers are former members of monitoring commissions are associated with an increased likelihood of acquittal, a reduced likelihood of mild punishment and an even more reduced likelihood of harsh punishment.

3.3. Accusation-Related Factors

The outcomes of the investigations may also depend on some factors related to the case under scrutiny. The severity of the case and its salience may affect the outcomes of bureaucratic agencies in charge of analyzing and deciding on critical issues (Carpenter 2002). We observe two critical features in particular: social norms and codes of conduct among police officers related to the type of deviation, and the salience of the accusation.

In the first case, tolerance of misconduct—both external and internal—is likely to be greater when colleagues want to preserve the integrity of the officers on duty and of the victims (Rappert 2002; Seron, Pereira and Kovath 2004). In fact, many police officers (and also some citizens) believe that “there is nothing wrong with police officers using foul,
insulting, abusive, or threatening (though not racist) language in response to citizens who insult, defy, or resist them” (Klockars et al 2006, 5). Abuse of “force” to arrest or extract information from suspects may also be seen as a legitimate way of effectively carrying out criminal investigations. In this sense, officers in charge of investigating colleagues implicated in deviations that are “socially accepted” by a given group may deliberately avoid severe punishment of officers implicated in violations legitimized by social norms. Yet, because by the rule of the law some offenses are expected to lead to some punishment, investigating officers may want to recommend mild (instead of harsh) punishment to avoid accusations of rampant impunity. Therefore:

\[ H_{3a} \] Deviations legitimized by internal norms and codes of conduct (such as excessive violence against suspects) will be associated with a reduced likelihood of acquittal, an increased likelihood of mild punishment, and a reduced likelihood of harsh punishment.

Second, although most cases of misconduct have low public visibility (Newburn 1999), certain deviations are also more salient (for example, brutal crimes against innocent citizens) and consequently they are more likely to be covered by the media. In fact, media portrayals of officer malpractice are thought to influence public opinion and shape the relationship between society and police departments (Graziano, Schuck and Martin 2010; Weitzer 2002). For example, the scenes of brutality and abuse of authority by officers of the Los Angeles Police Department against Rodney King in 1991 were considered the “most costly civil disturbance in the nation’s history” (Jacobs 1996, 1240) and tainted the image of the police substantially (Lasley 1994). External pressure from the public may thus prompt IAD officers to investigate such visible cases quickly (Bac 2001; Shi 2009); or, alternatively, extensive media coverage can be an indicator of how serious the deviation was, which can also exert an influence on the final judgment. All these factors should increase the likelihood that the IAD will recommend exemplary punishment for the deviant officers:
$H_{3b}$ Salient deviations, which are more likely to be covered in the media, are associated with a decreased likelihood of acquittal, an increased likelihood of mild punishment, and an even more increased likelihood of harsh punishment.

Finally, certain cases may be inherently more complex because they involve multiple accusations and multiple officers under investigation. For instance, several officers can be accused of kidnapping and murdering a citizen, and this could significantly increase the cost of collecting evidence, scheduling collective hearings, and so forth. Information gaps are expected to hamper the performance of the inspection system (Boyne et al 2002), even when agent-monitors have superior expertise and experience in the investigation task (Nutley et al 2012). Such complexity can increase the likelihood of acquittal, although no a priori effect is expected in terms of how harsh the punishment will be; it will depend on particular officer- or accusation-specific features. This logic leads to our last hypothesis:

$H_{3c}$ Complex cases involving multiple accusations and multiple accused officers are associated with an increased likelihood of acquittal, and a reduced likelihood of mild and harsh punishment.

4. Context and Data

We assess investigations against officers from the civil police force conducted by the General Internal Affairs Division (GIAD) of a Brazilian state. GIAD is a unified agency responsible for overseeing all specialized divisions in the state: the civil police, the military police, and the technical police. It is directly subordinate to the State Public Security Secretary, as are the other police divisions.

Although the investigation commissions set up by GIAD are not subordinate to any specific unit of the police, all commission members (agent-monitors) are former officers (agent-executers). Agent-monitors working for GIAD may also be reassigned to regular police activities and work again with the very agent-executers who they investigated in the past. In addition, some rotating, temporary GIAD commissions may be formed with current
police officers. In this case, there is no separation of roles between agent-monitors and agent-executors.

When a case is concluded, the final recommendation of the commission is then subject to final evaluation by the Chief of GIAD and by the State Secretary of Public Security, who have discretionary power to sanction the recommendation, or decide otherwise. If the final decision involves termination of employment, then the State General Attorney Office should also analyze the case before the State Governor sanctions the final punishment. In spite of these various phases and requirements, state law requires that investigations should be concluded within 120 days. However, as we shall see below, most cases last much longer.

4.1. Sampling Procedure

We accessed a list of all the investigation cases against police officers of the state civil police force submitted to GIAD between 1999 and 2006 (683 cases), and examined the outcomes of these cases up to December 2008. Our definition of “case” is rather specific: given an accusation, each observation corresponds to a particular implicated police officer. Thus, our unit of analysis is the investigation process against a particular officer who allegedly misbehaved.

Some accusations, as we discuss below, involve more than one police officer. Due to missing data in IAD’s internal archives for some variables—mainly the type of the accusation, records of previous accusations, age, and participation in previous investigation commissions—our sample was reduced to 639. Despite this reduction, our database still covers a substantial number (94%) of all cases that went to GIAD during our temporal window. For each case, we collected information on whether the judgment was concluded or not, the duration (in months) until its conclusion (if any), and the final outcomes of the concluded cases (i.e. whether the police officer was convicted or not). We were also granted access to officers’ individual files, with a track record of officers’ duties, personal data and
accusations within their specific departments. As a result, we built up a unique dataset on the cases investigated and under investigation in the state police. Our objective measures of performance also have the distinctive feature of avoiding response and measurement errors associated with survey-based research (Lee, Benoit-Bryan and Johnson 2012).

Admittedly, our database does not cover all possible accusations against police officers. Records provided by GIAD officers indicate that about 63% of the concluded investigations within local IADs escalated to GIAD in 2009. Unfortunately, we were unable to obtain reliable information from local IADs because, unlike GIAD, most units did not grant us access to information. Despite this limitation, we believe that our sample of GIAD cases is informative given the purposes of our research. Around 11% of the total population of officers in the State civil police were observed in our database, which indicates that GIAD is not investigating a negligible fraction of the force. Furthermore, there is substantial heterogeneity in officer- and case-specific characteristics, which allows us to examine how these factors influence final judgment outcomes.

The process of gathering information took more than 160 hours, between January and February 2009. During this period, one of us had the opportunity to share the same room as some GIAD officers, thus facilitating interaction and allowing for a deeper understanding of the mechanisms and procedures within the division. Although our empirical approach is essentially quantitative, we also conducted nine semi-structured interviews with GIAD members. Each interview took between 40 and 150 minutes. In addition, we organized a focus group with 30 officers (including detectives and chiefs) to identify their perceptions of police behavior and the work of the internal affairs divisions. Finally, we were allowed to observe two actual interrogation sessions, each lasting around 120 minutes. With this additional qualitative information, our goal was to improve our understanding of the local context of the GIAD so as to facilitate the interpretation of quantitative findings.
4.2. Our Data

Table 1 presents a description of our variables. We focus on the governance domain of organizational performance. Our performance (dependent) variables are measures coding the outcomes of the investigations. These outcomes are measured by three variables: Acquittal, Suspension and Termination, which are dummies coded 1 if the police officer was acquitted, punished through suspension and punished through termination of duty, respectively.\(^6\)

Termination is seen as a much harsher punishment than temporary suspension, which is considered as a mild sanction. Figure 1 provides a breakdown of concluded investigations according to the final judgment and the year when the investigation was started.\(^7\)

Because punishment depends on case conclusion, the incidence and severity of punishment can be greatly reduced if officers are able to postpone or undermine investigations. In our context, although by state law investigations are required to be concluded in four months, the modal conclusion length in our sample is around 18 months and we have instances of investigations that lasted more than 100 months (and even cases that were not concluded by the time we finished our data collection). In fact, in our database there are instances of pending cases dating back to the beginning of our temporal window, 1999 (Figure 2). The speed at which decisions are made is a key performance element of public bureaucracies throughout the world (Carpenter 2002). Thus, we jointly estimate the duration of the investigation given each potential outcome. Length is the time (in months) between the starting date of the case at GIAD and its conclusion date (or, in cases that were not concluded, the last date of our data collection, December 2008).

<<Figures 1 and 2 around here>>

In line with our proposed theory, we use three sets of explanatory variables: officer-specific characteristics related to their position and experience in the police; organizational features of the investigation to judge the case; and accusation-related variables of the case.
itself, such as the type of accusation and number of people implicated. All these variables are explained in detail in Table 1. As for the variables measuring the organizational features of the investigation, some further clarification is necessary. Cases within GIAD are judged following two distinct procedures. Permanent commissions are comprised of full-time police officers with longer tenure (generally more than twenty years) and with a good reputation in the force. These officers are responsible for judging various cases and are expected to follow the usual judicial procedures (scheduling hearings for the accused officers and witnesses, documenting the overall process and so on). Thus, members of permanent commissions are specialized agent-monitors, unlike other officers who remain in the force as agent-executers. The baseline case involves temporary commissions formed on a case-by-case basis, without any fixed appointment of the officials involved. Therefore, in this case, officers participating in a given judgment are not specialized in the task of judging others; in other words, they are both agent-monitors and agent-executers at the same time.

<<Table 1 around here>>

5. Estimation Methods

Given the nature of our hypotheses, assessing simultaneously the hazards of acquittal and the type of punishment, and the fact these events compete with one another, we employ a competing risk approach. Competing risks models have been used as an alternative to the analysis of survival data when there are multiple categories of failure (David and Moeschberger 1978)—or, in our case, alternative outcomes of case conclusion. Traditional survival analyses using proportional hazard models (see Cox 1972) fail to capture competing events, in which a given event may hinder the incidence of an alternative outcome and when an individual is exposed to more than one mutually exclusive event. Indeed, in the presence of competing risks and distinct failure types, the separate analysis of how each factor affects the hazard of each particular outcome is warranted (Prentice et al 1978).
A way to accommodate competing risks is to adapt the traditional proportional hazard model by allowing hazards to vary according to each type of outcome (e.g. Kalbfleisch and Prentice 2002: 255). Fine and Gray (1999) developed a model that incorporates time-dependent covariates, thus allowing the estimation of cause-specific “subdistribution hazards” assessing the risk of a given outcome given that other competing outcomes were not observed. More formally, we have the following specification:

\[ h_i(t, Z) = h_{0i}(t) \exp(Z^T \beta_i) \]

where \( h_i(t, Z) \) is the subdistribution hazard function for outcome \( i \), \( h_{0i}(t) \) is the “baseline” hazard function (which is not parameterized), \( Z_i \) is a vector of covariates (which can be time-varying) and \( \beta_i \) is a vector of their corresponding coefficients. In our particular context, we assess how covariates change the likelihood of a given outcome (e.g. termination) considering that other outcomes (e.g. suspension or acquittal) are also possible. Arguably, as per our hypotheses, these covariates may affect the rate of acquittal, suspension and termination differently. The Fine and Gray model has been widely used in several applications involving duration data with multiple outcomes (Beyersmann and Schumacher 2008).

Cases that had not been concluded by the end of the temporal window of the database are treated as censored observations. This feature of the model is particularly important given that we have a number of cases that were not concluded by the time we finished our data collection (see Figure 2). In addition, we add year dummies coding the starting year of the investigation to control for temporal effects associated with the timing of the case. Finally, because our unit of analysis is an accusation associated with a particular officer, and because multiple officers may be implicated in the same accusation, we adjust the estimates of the standard errors so as to consider any possible correlation between observations from officers associated with the same case.
A key assumption of the proportional hazards specification is that the ratio of hazard rates for two observations with distinct values of the covariates is constant over time. However, certain variables may have differential effects according to the observed length of the investigation. Our analysis of the data revealed that Permanent commission, in particular, has a marked time-varying effect. Indeed, the log-log graphs in Figure 3 show that the effect of a permanent commission on the conclusion of investigations (thus reducing their “survival” over time) varies according to the life span of the case, especially for the suspension and termination outcomes.\footnote{For instance, Figure 3b suggests that the Permanent Commission is more effective in increasing the likelihood of suspension for investigations that are expected to last for a long time. We thus proceed in some regressions by specifying the variable Permanent commission as a time-varying covariate, i.e. adding a multiplicative term Permanent commission×Time, which allows us to model how the effect of the variable changes according to the observed length (in months) of the investigation.}

6. Results and Discussion

6.1. General results

Table 2 reports the results of the proportional hazard regressions with competing risks. Model (1a), for instance, examines the likelihood of acquittal given that suspension and termination are competing risks. The only difference between regression sets (1) and (2) is that in the latter we specify the Tenure and Age variables as ranges based on some pre-specified levels so as to detect possible nonlinear effects (along these lines, see McElvain and Kposowa 2004). Regression set (3), in turn, adds the interaction term Permanent commission×Time, discussed before, to accommodate the time-varying effects of the permanent commission.
With respect to officer-specific characteristics, we find a strong, negative effect of Special chief ($p < 0.01$): top officers are generally not punished with termination of duty. Indeed, in our database we have no instance of a special chief who was fired. We also see that Tenure is a relevant covariate, especially for the risk of termination. Model (1c), for instance, shows that each additional year in the force reduces the hazard of termination by 5.7%. Moreover, as revealed by regression models (2c) and (3c), its effect seems to be nonlinear. According to model (3c), cases where $5 \leq \text{Tenure} \leq 9$ are also less likely to result in termination than cases where $10 \leq \text{Tenure} \leq 19$, although the difference is moderately significant ($p < 0.10$) and becomes not significant when we compare the latter with cases where $\text{Tenure} \geq 20$.\textsuperscript{10} Thus, it appears that some experience in the force is already sufficient to avoid a severe outcome: estimates from model (3c) indicate that after five years the hazard that the officer will be fired (rather than acquitted) falls by 55.8%. We also do not find significant effects for the other officer-specific variables Chief, Staff and Commissioned job.

Therefore, Hypothesis 1 is supported only with respect to harsh punishment (termination) and for officer-specific variables coding experience and distinguished ranking. Apparently, top experienced officers can use their superior knowledge and possible connections within the force to avoid decisions that involve harsh punishment; or, alternatively, that commissions are reluctant to severely punish more experienced officers.\textsuperscript{11} Our findings are consistent with previous anecdotal evidence in other countries. For instance, an officer from the Charlotte-Mecklenburg Police Department in the United States, interviewed by Klockars et al. (2006: 222), asserted that “the higher the rank – the lighter the punishment. There is little consistency in the punishment that is meted out.” The disadvantage of inexperienced officers in the investigation process is also seen in the following quote from one of our interviews:

“… a guy who is very young [in the force] does not yet know how to do things and ends up being caught…” (GIAD officer, pers. comm.)
As for the variables coding organizational features of the investigation, *Permanent commission* displays complex and, as previously discussed, time-varying effects. Models (1b) and (2b) indicate that the presence of the permanent commission approximately doubles the hazard of suspension. The main effect of *Permanent commission*, however, substantially changes when we add the time-varying term. The coefficient of *Permanent commission* × *Time* in model (3b), for instance, indicates that the positive effect of the permanent commission on the likelihood of suspension only holds for cases with a longer duration (*p* < 0.01). The effect of the commission on the hazard of termination is even more complex. Although the variable has a negative main effect on the likelihood of termination in model (3c), their effect becomes progressively positive as the duration of the case increases (*p* < 0.01). In other words, consistent with the patterns shown in Figure 3b and 3c, permanent commissions appear to increase the likelihood of suspension and termination but only in cases that would otherwise take longer than average.

Thus, Hypothesis 2a is supported only for cases with a long time span. Permanent commissions appear to accelerate sanctions when judging more difficult cases that would naturally last longer than average. However, our data also reveal that permanent commissions are associated with a lower hazard of termination for cases that can be concluded faster (see Figure 3c). A possible explanation for this is that before recommending harsh punishment, permanent commissions implement extra investigation procedures that help disclose information (e.g. more formal hearings) but that also have the downside of extending the duration of the case. This conclusion is supported by the following quote from our interviews:

“… the experience of the permanent commission members makes a difference…the quality of the investigation process carried out by the permanent investigation commissions is a hallmark” (Police Chief in the focus group, pers. comm.)
The coefficients of both Commission member and Tie with commission member, in turn, are generally not significant and therefore fail to provide support for Hypotheses 2b and 2c. Although Commission member shows a negative effect on the likelihood of suspension in model (3b), its significance is moderate ($p < 0.10$) and does not hold in alternative specifications. Therefore, accused officials who used to be commission members and who worked in the past with some current member are apparently unable to use their knowledge of judgment mechanisms and possible social ties with current commission members to affect investigation outcomes. Likewise, there is no evidence that commission members postpone the conclusion of the investigation when former colleagues are implicated.

A host of significant results, however, appear when we assess accusation-related variables. Cases involving extortion are more likely to lead to suspension and less likely to lead to acquittal ($p < 0.01$). Our interviews did reveal that extortion—such as when an officer requests money from entrepreneurs in return for protection—is a particularly abhorred type of behavior within the police force:

“Extortion is bad because the officer uses the organization for his own benefit... It stains the reputation of the whole police. A person who is victim of extortion generalizes the situation [for the whole police force].” (Head of GIAD, pers. comm.)

There are also some positive effects associated with disrespect, abuse and aggression, however with varying levels of significance. These cases tend to be less severely punished through termination. As indicated by one of our interviewed officers:

“If there is an internal investigation because the police officer beats up a bad guy, then folks [from IAD] don’t take that too seriously.” (Former investigation commission member 1, pers. comm.)

Yet IAD apparently does not turn a blind eye to such deviations, given that they are more likely to result in suspension. As mentioned before, abuse and aggression against suspects are apparently seen as behaviors that are more “acceptable” if the officer is acting so
as to actively solve a case and arrest the implicated criminals. Thus, IAD may use suspension as a way to avoid firing the deviant officers while at the same time deflecting allegations of impunity. In line with our theoretical discussion in section 3, these results appear to indicate that informal norms and codes of behavior within the police organization are important determinants of the final outcome of investigations in internal affairs divisions.12 Hypothesis 3a is therefore partially supported. Although such legitimised deviations are associated with mild punishment, no significant effect is found in terms of acquittal. Extortion, in turn, decreases the hazard of acquittal but does not affect termination.

We also find a significant effect of Media coverage. The effect of Media coverage on the likelihood of termination is also larger than its effect on suspension. According to the estimates of models (2b) and (2c), the difference between the coefficients of Media coverage is statistically significant \((p < 0.05)\) and marginal effects indicate that cases with media coverage increases the hazards of suspension and termination by 60.0% and 258.9% respectively. As discussed before, this result can be due either to external pressure to punish deviants or to the fact that more salient, shocking cases (naturally more prone to trigger punishment) tend to be covered by newspapers and other media. However, although the effect of Media coverage on the hazard of acquittal is generally negative, statistical significance is moderate across model specifications \((p < 0.10)\). We thus find partial support for Hypothesis 3b.

An increase in the number of accused officers increases the incidence of acquittal \((p < 0.01)\) and reduces the likelihood of punishment, especially with respect to suspension \((p < 0.01)\). Taking the estimates of models (2a) and (2b), each additional accused officer increases the hazard of acquittal by 25.5% and reduces the hazard of termination by 23.2%. However, the effect of the Number of accusations is significant only for the hazard of suspension \((p < 0.05)\). Thus, Hypothesis 3c is supported only with respect to the number of accused officers.
Apparently, the difficulty of collecting evidence and judging multiple individuals reduces the odds that the implicated officers will be punished.

6.2 Robustness check: permanent commission

The use of permanent commissions as a covariate also poses another problem: they may be endogenously chosen based on unobserved factors affecting the “difficulty” of the case. For instance, our interviews revealed that GIAD sometimes reassigns some delayed cases to permanent commissions established in the state capital. If this is true, then the coefficient of *Permanent commission* in the early regressions will be biased. For instance, in the process of case reassignment, permanent commissions may receive cases with officer- or accusation-specific traits that will affect case conclusion differently from the cases assigned to temporary, rotating commissions.

To circumvent this problem, we adopt the technique of propensity score matching (PSM; see Rosenbaum and Rubin 1983) to compare cases with and without the permanent commission better. PSM has been successfully used in studies with survival data (see e.g. Austin 2008). In particular, we use PSM to create “weights” reflecting the probability that a case, given its officer- and accusation-specific traits, will be assigned to a permanent commission. These weights are subsequently used in the hazard regression (e.g. Hernán, Brumback and Robins 2000). The objective of such weighting is to make the cases with and without a permanent commission more similar based on their basic characteristics, thus increasing the comparability among them. We also consider matched observations in regions of common support, i.e., cases within a similar range of propensity scores (Heckman et al., 1997).

The results are shown in Table 2, regression models (4a) to (4c). As before, the effect of the permanent commission is found to be dependent on the length of the case: despite its negative main effect on the likelihood of termination, the commission increases the hazard of
suspension and termination when cases have a long time span. Our earlier conclusion thus appears to hold: although at a potential cost of more formalized investigation procedures, permanent commissions appear to accelerate punishment in cases that would otherwise take longer.

Table 3 consolidates our main findings, related to our previously outlined hypotheses.

<<Table 3 around here>>

7. Conclusion

Scholars have proposed that monitoring and accountability are key factors in improving the performance of public services. Yet, given the complexity of performance objectives as well as the dispersed nature of the relevant stakeholders affected by public services, effective monitoring becomes a formidable challenge. We contribute to this debate by focusing on the so-called guarding the guardians dilemma, which occurs when the very agents to be monitored are appointed as monitors. Whenever agents become responsible for accountability functions in bureaucracies, there is a concern that they will bias their judgment in favor of certain colleagues or certain types of deviations which they deem “acceptable”. Agents may exert suboptimal effort in the investigation, thereby resulting in impunity and poor organizational performance.

Our theory and empirical analysis help expand the extant literature on the efficient design of public bureaucracies by proposing a host of factors that can influence the governance performance of public agencies subject to the guarding the guardians problem. Our study also sheds some light on the underlying mechanisms that shape accountability and probity measures of governance performance, which is a promising venue in the public service performance literature (Walker et al 2010). We show that the complex interplay of management decisions and contingent variables may affect the outcomes of oversight organizations in different ways. By scrutinizing potential factors influencing the investigation
process and its outcomes, our work also provides several insights for public managers involved in oversight bodies.

In particular, we show that high-ranking, experienced officers (with longer tenure in the police bureaucracy) are punished less harshly than their younger colleagues, thus confirming that the position and tenure of the officer in the force matter. We also observe that certain types of deviations interpreted as a normal “side effect” of police action (such as aggression against suspects) are treated less harshly than other deviations. This result suggests that norms and codes of conduct can attenuate sanctions against deviants. However, we found that extensive media coverage of officers’ deviations increases the odds that the implicated officers will be convicted, which demonstrates the positive effect of media on the governance performance of monitoring agencies. This finding thus suggests that increasing media access to events of malpractice can be an important way to prompt punishment in settings involving the guarding the guardians problem.

Our analysis of permanent commissions also provides some insights into the separation of roles between agent-executers and agent-monitors. In light of our finding that such commissions are particularly helpful in long running investigations, another policy implication is that the creation of units with specialized agent-monitors should be targeted to more complex cases that require detailed scrutiny and multiple investigation steps. Although we found no evidence that commissions are prone to pressure from the investigated officers who have possible ties with commission members, the careful selection of reputable officers relatively insulated from internal influence remains a point of attention.

Of course our paper has some limitations, which create several avenues for future research. We focused on a single oversight unit of a specific geographic location and the results we obtained might not hold in distinct organizations or even in other internal affairs divisions located in other developing or developed countries. Further research can be carried
out in other institutional contexts or in a cross-country setting to verify the generalizability of our findings. This will allow us to learn from the practices adopted by other watchdog agencies and other public bureaus where agents are appointed to monitor other agents. Future studies should address the impact of organizational and individual characteristics on the governance performance of other types of public services plagued by the guarding the guardians dilemma (such as legislatures or other regulatory bodies).

There is also room to examine other dimensions of public service performance, perhaps at a more aggregated level. For instance, scholars could also assess if improved governance performance of internal affairs divisions is associated with reduced criminality or improved public perceptions of security and fair treatment. Our data collection also indicates that police divisions as well as other public service organizations can open up their data to public scrutiny. It is also possible to examine how improved transparency of public sector performance can trigger improvements in the governance of public bureaus or in other relevant dimensions. We hope that new studies can address these remaining issues.

References


Hood, Christopher., James, O., & Scott, C. 2000. Regulation of government: Has it increased, is it increasing, should it be diminished? *Public Administration*, 78(2), 283-304.


Notes

1 Public administration scholars (Flemming 2008; Jobson and Schneck 1982; Nicholson-Crotty and O’Toole 2004; Needham 2009) and economists (Glaeser & Sacerdote 1999; Levitt and Donohue III 2001) have long studied criminal justice and law
enforcement; however, issues related to the internal monitoring of the police remain relatively unaddressed. Also, although sociologists (Mulcahy 1995) and criminologists (Klockars et al. 2006; Liederbach et al. 2007; Prenzler 2000; McElvain and Kposowa 2004; Prenzler and Lewis 2005; Walker 2006) have examined the functioning of internal divisions, to the best of our knowledge there is no specific study focusing on the factors that may facilitate or thwart the effective judgment of charges against police officers who engage in deviations.

2 The name of the State is not revealed for reasons of confidentiality. However, this Brazilian state is among the five most populated states in the country and its capital is one of the most important cities in Brazil.

3 In Brazil, at State government level there are two types of law enforcement agencies: military police and civil police. While the former is responsible for preserving social order with a militarized approach, the latter is in charge of investigating crimes, preparing inquiries, and supporting subsequent legal sanctions against the accused (Paes-Machado and Noronha 2003). Normally, State military police forces are also responsible for fire brigade departments, while the civil police conduct on-site investigations, which is a crucial task in law enforcement. In the context of the State in this study, a third law enforcement body runs part of the forensic and legal medicine activities: the technical police. The three police forces of the state have their own internal affairs divisions.

4 We were unable to gain access to investigation processes against military police officers. The civil police force of the State under analysis here had as of 2010 5,506 police officers (16.1% are police chiefs, 68.3% are detectives and 15.6% are staff). From Table 1 we can observe that in our sample 81% of the investigations involve detectives, which is fairly above the proportion of detectives in the police force. The main reason for this difference is related to the fact that detectives are more prone to engage in misconduct given the nature of their activity, which is normally conducted outside police facilities.

5 We interviewed six officers directly involved the investigations, which represented 80% of GIAD staff. The remaining 20% were on vacation at the time we performed our field work. We also interviewed two former investigation commission members and the head of GIAD, who is not member of the civil police. The main open questions covered the difficulties when performing the investigations, the role of external pressure, the main strategies used by the accused officers during the hearings, the organizational design of GIAD, the role of the type of the accusation, among other issues. The questions are not reported due to space limits, but they are available upon request. We also ran a focus group with randomly assigned officers. Basically using the same open questions as in our interviews, we approached these officers to obtain their perceptions of IAD and its monitoring function.

6 Suspended officers are not allowed to perform their activities. Their guns, badges and internal identification codes are removed. In addition, during the penalty period, the suspended officers do not get paid, and the suspension may negatively affect future job promotions within the force.

7 Notice, again, that our database covers only cases that reached GIAD, which means that local (departmental) IADs had already recommended conviction for the accused officers. Unfortunately, we were not granted access to investigations initiated (and terminated) at local IADs.

8 Gordon (2002) also presents an alternative model with two competing risks and stochastic dependence across those risks (i.e. sources of unobserved heterogeneity that may affect multiple conclusion outcomes).

9 To understand the log-log specification, notice that the proportional hazards assumption implies that \( \ln[S(t)]/\ln[S'(t)] = \exp[(Z – Z')\beta] \) where \( Z \) and \( Z' \) are distinct vectors of the covariates (observations with and without a permanent commission) and \( S(t) \) and \( S'(t) \) are the associated survival functions (the cumulative probability that processes will not be concluded beyond month \( t \)). Taking logs at both sides we obtain \( \ln[-\ln[S(t)]]/\ln[-\ln[S'(t)]] + [(Z – Z')\beta]. \) Thus, if the proportional hazards assumption holds, then the probability of non-conclusion (“survival”) as a function of time when the commission is permanent should be graphically parallel to the function when the commission is not permanent. If these functions are not parallel, as shown in Figure 3, then there is probably violation of the proportional hazards assumption. However, as we explain in the text, our time-varying covariate will fix this effect. See, for instance, Cleves et al. (2004).

10 There is, however, an unexpected positive effect on suspension but only for cases where \( 10 \leq \text{Tenure} \leq 19. \)

11 One might argue that a form of “survival bias” can drive this result: over time, only the best officers are promoted and remain in the force. However, we control for this potential bias by including variables that code the history of the officer (number of past accusations and convictions).

12 It is interesting to note the absence of significance involving the variables Previous accusations and Previous convictions, even though some authors have observed that internal affairs divisions may be more willing to convict “problem officers” who normally receive repeated complaints (Terrill & McCluskey 2002).
FIGURE 1
Final outcomes of the concluded cases according to the year when they were started

Year when the process was started

- Acquittal
- Suspension
- Termination
FIGURE 2
Number of investigations according to the year when they were started in the IAD

![Bar graph showing number of investigations by year.](Figure_2)

FIGURE 3
Log-log graphs indicating how the likelihood of conclusion, according to each possible outcome, varies with time (months) and type of investigation commission

a) Acquittal

![Log-log graph for acquittal.](Figure_3a)

b) Suspension

![Log-log graph for suspension.](Figure_3b)
c) Termination

![Graph showing survival probability vs. ln(Length) for perm. commission = 0 and perm. commission = 1.]

- Perm. commission = 0
- Perm. commission = 1
TABLE 1
Summary statistics and description of variables

<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
<th>Mean (std. dev.)</th>
<th>Min.</th>
<th>Max.</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variables</td>
<td>Length</td>
<td>31.105 (22.294)</td>
<td>4</td>
<td>118</td>
<td>Length (in months) between the starting date of the case and either its conclusion date or the final date of data collection (December 2008).</td>
</tr>
<tr>
<td></td>
<td>Acquittal</td>
<td>0.258 (0.438)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the case resulted in acquittal and 0 otherwise.</td>
</tr>
<tr>
<td></td>
<td>Suspension</td>
<td>0.288 (0.453)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the case resulted in suspension and 0 otherwise.</td>
</tr>
<tr>
<td></td>
<td>Termination</td>
<td>0.135 (0.342)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the case resulted in termination of duty and 0 otherwise.</td>
</tr>
<tr>
<td>Officer-specific features</td>
<td>Chief</td>
<td>0.116 (0.320)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the police officer is a chief and 0 otherwise.</td>
</tr>
<tr>
<td></td>
<td>Special chief</td>
<td>0.025 (0.156)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the police officer is a &quot;special&quot; chief (with a higher position in the hierarchy) and 0 otherwise.</td>
</tr>
<tr>
<td></td>
<td>Staff</td>
<td>0.074 (0.261)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the police officer holds a staff (clerical) position.</td>
</tr>
<tr>
<td></td>
<td>Commission job</td>
<td>0.336 (0.473)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the police officer was appointed by other top individuals in the hierarchy.</td>
</tr>
<tr>
<td></td>
<td>Tenure</td>
<td>11.114 (8.358)</td>
<td>1</td>
<td>39</td>
<td>Number the years since the officer joined the police force at the case starting date</td>
</tr>
<tr>
<td>Organizational features</td>
<td>Permanent commission</td>
<td>0.465 (0.499)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the case was judged by a commission composed of officers responsible for evaluating multiple cases.</td>
</tr>
<tr>
<td></td>
<td>Commission member</td>
<td>0.128 (0.335)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the officer under investigation was him or herself also an internal affairs commission member in the past.</td>
</tr>
<tr>
<td></td>
<td>Tie with commission member</td>
<td>0.156 (0.364)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the officer under investigation worked in the past in the same police station of any member of the investigation commission responsible for the case,</td>
</tr>
<tr>
<td>Accusation-related features</td>
<td>Homicide</td>
<td>0.139 (0.347)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the accusation involves homicide of any kind.</td>
</tr>
<tr>
<td></td>
<td>Extortion</td>
<td>0.297 (0.457)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the accusation involves extortion of any kind.</td>
</tr>
<tr>
<td></td>
<td>Trafficking</td>
<td>0.030 (0.170)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the accusation refers to the officer's involvement in drug trafficking.</td>
</tr>
<tr>
<td></td>
<td>Disrespect</td>
<td>0.205 (0.404)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the accusation involves disrespect of internal rules and insubordination.</td>
</tr>
<tr>
<td></td>
<td>Abuse</td>
<td>0.274 (0.446)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the accusation involves abuse of power or authority</td>
</tr>
<tr>
<td></td>
<td>Aggression</td>
<td>0.275 (0.447)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the accusation involves aggression towards suspects.</td>
</tr>
<tr>
<td></td>
<td>Media coverage</td>
<td>0.097 (0.296)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the incident was covered by local newspapers or other media.</td>
</tr>
<tr>
<td></td>
<td>Number of</td>
<td>2.243</td>
<td>1</td>
<td>8</td>
<td>Number of police officers implicated in</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev</td>
<td>Min</td>
<td>Max</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------</td>
<td>----------</td>
<td>-----</td>
<td>-----</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>accused</td>
<td>(1.487)</td>
<td>(0.533)</td>
<td>0</td>
<td>4</td>
<td>Number of deviations included in the case against the officer.</td>
</tr>
<tr>
<td>Number of accusations</td>
<td>1.330</td>
<td>(0.533)</td>
<td>0</td>
<td></td>
<td>Number of violations included in the same case.</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.944</td>
<td>(0.231)</td>
<td>0</td>
<td>1</td>
<td>Dummy variable coded 1 if the officer is male and 0 otherwise.</td>
</tr>
<tr>
<td>Age</td>
<td>40.690</td>
<td>(8.164)</td>
<td>25</td>
<td>69</td>
<td>Age of the police officer, in years at the case starting date.</td>
</tr>
<tr>
<td>Previous accusations</td>
<td>0.912</td>
<td>(2.960)</td>
<td>0</td>
<td>50</td>
<td>Number of previous accusations that the officer received.</td>
</tr>
<tr>
<td>Previous convictions</td>
<td>1.235</td>
<td>(3.218)</td>
<td>0</td>
<td>53</td>
<td>Number of previous convictions.</td>
</tr>
</tbody>
</table>

$N = 639$. Each unit of observation is an accusation against a particular police officer; a particular case may involve multiple accusations towards more than one officer.
### TABLE 2
Factors affecting the final judgment of investigations (proportional hazard regressions with competing risks)

<table>
<thead>
<tr>
<th></th>
<th>Acquittal (1a)</th>
<th>Suspension (1b)</th>
<th>Termination (1c)</th>
<th>Acquittal (2a)</th>
<th>Suspension (2b)</th>
<th>Termination (2c)</th>
<th>Acquittal (3a)</th>
<th>Suspension (3b)</th>
<th>Termination (3c)</th>
<th>Acquittal (4a)</th>
<th>Suspension (4b)</th>
<th>Termination (4c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chief</td>
<td>-0.165</td>
<td>0.208</td>
<td>-0.281</td>
<td>-0.272</td>
<td>0.233</td>
<td>-0.487</td>
<td>-0.273</td>
<td>0.221</td>
<td>-0.485</td>
<td>-0.058</td>
<td>0.222</td>
<td>-0.635</td>
</tr>
<tr>
<td></td>
<td>(0.319)</td>
<td>(0.293)</td>
<td>(0.440)</td>
<td>(0.319)</td>
<td>(0.296)</td>
<td>(0.510)</td>
<td>(0.320)</td>
<td>(0.293)</td>
<td>(0.502)</td>
<td>(0.469)</td>
<td>(0.333)</td>
<td>(0.654)</td>
</tr>
<tr>
<td>Special chief</td>
<td>0.487</td>
<td>-0.758</td>
<td>-13.677***</td>
<td>0.732</td>
<td>-0.758</td>
<td>-14.074***</td>
<td>0.726</td>
<td>-0.761</td>
<td>-13.888***</td>
<td>0.779</td>
<td>-0.894</td>
<td>-22.331***</td>
</tr>
<tr>
<td></td>
<td>(0.498)</td>
<td>(0.679)</td>
<td>(0.589)</td>
<td>(0.508)</td>
<td>(0.652)</td>
<td>(0.570)</td>
<td>(0.509)</td>
<td>(0.666)</td>
<td>(0.580)</td>
<td>(0.584)</td>
<td>(0.702)</td>
<td>(0.779)</td>
</tr>
<tr>
<td>Staff</td>
<td>-0.201</td>
<td>0.203</td>
<td>-0.003</td>
<td>-0.042</td>
<td>0.161</td>
<td>-0.039</td>
<td>-0.049</td>
<td>0.127</td>
<td>-0.053</td>
<td>-0.582</td>
<td>-0.530</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>(0.418)</td>
<td>(0.342)</td>
<td>(0.567)</td>
<td>(0.405)</td>
<td>(0.352)</td>
<td>(0.562)</td>
<td>(0.403)</td>
<td>(0.363)</td>
<td>(0.549)</td>
<td>(0.610)</td>
<td>(0.498)</td>
<td>(0.694)</td>
</tr>
<tr>
<td>Commissioned</td>
<td>-0.164</td>
<td>0.184</td>
<td>-0.078</td>
<td>-0.148</td>
<td>0.218</td>
<td>-0.014</td>
<td>-0.159</td>
<td>0.204</td>
<td>-0.036</td>
<td>0.016</td>
<td>0.302</td>
<td>0.503</td>
</tr>
<tr>
<td>job</td>
<td>(0.197)</td>
<td>(0.194)</td>
<td>(0.286)</td>
<td>(0.201)</td>
<td>(0.189)</td>
<td>(0.308)</td>
<td>(0.200)</td>
<td>(0.189)</td>
<td>(0.313)</td>
<td>(0.248)</td>
<td>(0.228)</td>
<td>(0.435)</td>
</tr>
<tr>
<td>Tenure</td>
<td>0.036*</td>
<td>0.016</td>
<td>-0.059**</td>
<td>0.122</td>
<td>0.180</td>
<td>-0.895**</td>
<td>0.131</td>
<td>0.162</td>
<td>-0.817**</td>
<td>0.717*</td>
<td>-0.097</td>
<td>-0.701</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.019)</td>
<td>(0.027)</td>
<td>(0.314)</td>
<td>(0.284)</td>
<td>(0.377)</td>
<td>(0.310)</td>
<td>(0.284)</td>
<td>(0.381)</td>
<td>(0.411)</td>
<td>(0.346)</td>
<td>(0.429)</td>
</tr>
<tr>
<td>5 ≤ Tenure ≤ 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 ≤ Tenure ≤ 19</td>
<td>-0.192</td>
<td>0.840****</td>
<td>-1.642****</td>
<td>-0.175</td>
<td>0.831****</td>
<td>-1.596****</td>
<td>-0.648</td>
<td>0.632*</td>
<td>-2.223***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.351)</td>
<td>(0.291)</td>
<td>(0.392)</td>
<td>(0.351)</td>
<td>(0.294)</td>
<td>(0.399)</td>
<td>(0.522)</td>
<td>(0.336)</td>
<td>(0.462)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tenure ≥ 20</td>
<td>0.383</td>
<td>0.561</td>
<td>-1.438****</td>
<td>0.402</td>
<td>0.571</td>
<td>-1.375****</td>
<td>1.302**</td>
<td>-0.243</td>
<td>-2.273***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.385)</td>
<td>(0.355)</td>
<td>(0.473)</td>
<td>(0.385)</td>
<td>(0.361)</td>
<td>(0.485)</td>
<td>(0.571)</td>
<td>(0.447)</td>
<td>(0.644)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perm. commission</td>
<td>0.082</td>
<td>0.768****</td>
<td>-0.448</td>
<td>0.109</td>
<td>0.671****</td>
<td>-0.293</td>
<td>-0.296</td>
<td>-0.213</td>
<td>-2.438****</td>
<td>-0.184</td>
<td>-0.610</td>
<td>-2.435****</td>
</tr>
<tr>
<td></td>
<td>(0.248)</td>
<td>(0.223)</td>
<td>(0.285)</td>
<td>(0.259)</td>
<td>(0.231)</td>
<td>(0.280)</td>
<td>(0.370)</td>
<td>(0.387)</td>
<td>(0.744)</td>
<td>(0.426)</td>
<td>(0.410)</td>
<td>(0.722)</td>
</tr>
<tr>
<td>Commission</td>
<td>-0.165</td>
<td>-0.360</td>
<td>0.191</td>
<td>-0.063</td>
<td>-0.429</td>
<td>0.347</td>
<td>-0.046</td>
<td>-0.461*</td>
<td>0.338</td>
<td>-0.329</td>
<td>-0.152</td>
<td>0.704</td>
</tr>
<tr>
<td>member</td>
<td>(0.274)</td>
<td>(0.260)</td>
<td>(0.411)</td>
<td>(0.272)</td>
<td>(0.265)</td>
<td>(0.423)</td>
<td>(0.272)</td>
<td>(0.264)</td>
<td>(0.423)</td>
<td>(0.356)</td>
<td>(0.286)</td>
<td>(0.442)</td>
</tr>
<tr>
<td>Tie with comm.</td>
<td>0.039</td>
<td>0.059</td>
<td>-0.352</td>
<td>0.098</td>
<td>-0.003</td>
<td>-0.369</td>
<td>0.098</td>
<td>-0.000</td>
<td>-0.379</td>
<td>0.140</td>
<td>0.070</td>
<td>-0.262</td>
</tr>
<tr>
<td>member</td>
<td>(0.266)</td>
<td>(0.233)</td>
<td>(0.399)</td>
<td>(0.261)</td>
<td>(0.230)</td>
<td>(0.405)</td>
<td>(0.259)</td>
<td>(0.231)</td>
<td>(0.388)</td>
<td>(0.279)</td>
<td>(0.268)</td>
<td>(0.421)</td>
</tr>
<tr>
<td>Homicide</td>
<td>-0.506</td>
<td>0.579</td>
<td>-0.271</td>
<td>-0.429</td>
<td>0.547</td>
<td>-0.317</td>
<td>-0.460</td>
<td>0.529</td>
<td>-0.308</td>
<td>-0.167</td>
<td>0.707</td>
<td>-1.198**</td>
</tr>
<tr>
<td></td>
<td>(0.313)</td>
<td>(0.527)</td>
<td>(0.439)</td>
<td>(0.320)</td>
<td>(0.530)</td>
<td>(0.449)</td>
<td>(0.322)</td>
<td>(0.527)</td>
<td>(0.451)</td>
<td>(0.381)</td>
<td>(0.612)</td>
<td>(0.499)</td>
</tr>
<tr>
<td>Extortion</td>
<td>-1.336***</td>
<td>1.557***</td>
<td>0.216</td>
<td>-1.248***</td>
<td>1.558***</td>
<td>0.150</td>
<td>-1.270***</td>
<td>1.568***</td>
<td>0.187</td>
<td>-1.224**</td>
<td>1.542***</td>
<td>-0.437</td>
</tr>
<tr>
<td>Variable</td>
<td>Coefficients (SE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trafficking</td>
<td>-0.238 (0.399) -0.584 (0.455) -0.508 (0.376)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disrespect</td>
<td>-0.802** (0.330) 1.639*** (0.448) -1.831*** (0.523)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abuse</td>
<td>-0.235 (0.360) 1.595*** (0.478) -0.988** (0.483)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggression</td>
<td>-0.248 (0.365) 1.581*** (0.478) -1.546*** (0.483)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media coverage</td>
<td>-1.041* (0.568) 0.455* (0.249) 1.254*** (0.353)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of accused</td>
<td>0.218*** (0.077) -0.253*** (0.071) -0.300* (0.171)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of accusations</td>
<td>-0.123 (0.422) -1.102** (0.508) 0.581 (0.425)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>-0.275 (0.379) 0.368 (0.480) 0.759 (0.800)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.003 (0.017) -0.021 (0.018) 0.005 (0.023)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 ≤ Age ≤ 40</td>
<td>-0.114 (0.329) -0.069 (0.289) -0.452 (0.399)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41 ≤ Age ≤ 50</td>
<td>0.337 (0.420) -0.758** (0.328) 0.036 (0.420)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age ≥ 51</td>
<td>-0.175 (0.465) -0.328 (0.410) 0.042 (0.540)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous convictions</td>
<td>-0.008 (0.144) 0.085 (0.133) 0.101 (0.220)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous accusations</td>
<td>-0.038 (0.137) -0.055 (0.129) -0.038 (0.208)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Permanent comm.</td>
<td>0.015 (0.133) 0.043*** (0.141) 0.094*** (0.211)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The table shows parameter estimates and robust standard errors in parenthesis (clustered in each case involving investigations against multiple officers). All models include dummy variables indicating the year when the investigation started (not reported here). Model (1a), for instance, assesses the likelihood of acquittal treating the likelihood of suspension and termination as competing risks. Models (4a) to (4c) consider weights based on a propensity score matching (PSM) model specifying the likelihood that a given case, given its observable traits, will be investigated by a permanent commission.

<table>
<thead>
<tr>
<th>×Time</th>
<th>Weights using PSM?</th>
<th>(0.011)</th>
<th>(0.016)</th>
<th>(0.027)</th>
<th>(0.012)</th>
<th>(0.019)</th>
<th>(0.029)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N (cases)</td>
<td>639</td>
<td>639</td>
<td>639</td>
<td>639</td>
<td>639</td>
<td>617</td>
<td>617</td>
</tr>
<tr>
<td>p (Wald test)</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>
## TABLE 3
Summary of results

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1$ Cases involving high-ranking, high-tenured officers are associated</td>
<td>Supported only for harsh punishment (termination). Top, experienced officers apparently exert influence on the final judgment or use their evolving internal contacts to help obstruct incriminating information.</td>
</tr>
<tr>
<td>with an increased likelihood of acquittal, a reduced likelihood of mild</td>
<td></td>
</tr>
<tr>
<td>punishment and an even more reduced likelihood of harsh punishment.</td>
<td></td>
</tr>
<tr>
<td>$H_{2a}$ Separation of roles between agent monitors and agent executors</td>
<td>Supported for cases with a long time span. Permanent commissions seem to increase the likelihood of suspension and termination in cases of conviction that would take longer than average.</td>
</tr>
<tr>
<td>is associated with a decrease in the likelihood of acquittal, and an</td>
<td></td>
</tr>
<tr>
<td>increase in the likelihood of mild and harsh punishment.</td>
<td></td>
</tr>
<tr>
<td>$H_{2b}$ Previous ties between investigated officers and commission</td>
<td>Not supported.</td>
</tr>
<tr>
<td>members are associated with an increased likelihood of acquittal, a</td>
<td></td>
</tr>
<tr>
<td>reduced likelihood of mild punishment and an even more reduced likelihood</td>
<td></td>
</tr>
<tr>
<td>of harsh punishment.</td>
<td></td>
</tr>
<tr>
<td>$H_{2c}$ Instances where investigated officers are former members of</td>
<td>Not supported.</td>
</tr>
<tr>
<td>monitoring commissions are associated with an increased likelihood of</td>
<td></td>
</tr>
<tr>
<td>acquittal, a reduced likelihood of mild punishment and an even more</td>
<td></td>
</tr>
<tr>
<td>reduced likelihood of harsh punishment.</td>
<td></td>
</tr>
<tr>
<td>$H_{3a}$ Deviations legitimized by internal norms and codes of conduct</td>
<td>Partially supported. Abuse and aggression are associated with an increased likelihood of suspension and a reduced likelihood of termination (no effect is found on acquittal). Deviations not legitimized by social norms (such as extortion) are associated with a reduced likelihood of acquittal and an increased likelihood of suspension (no effect is found on termination).</td>
</tr>
<tr>
<td>(such as excessive violence against suspects) will be associated with a</td>
<td></td>
</tr>
<tr>
<td>reduced likelihood of acquittal, an increased likelihood of mild</td>
<td></td>
</tr>
<tr>
<td>punishment, and a reduced likelihood of harsh punishment.</td>
<td></td>
</tr>
<tr>
<td>$H_{3b}$ Salient deviations, which are more likely to be covered in the</td>
<td>Partially supported: media coverage appears to trigger more severe punishment, although the predicted negative effect on acquittal shows moderate significance. This result is probably because salient deviations are perceived to be more “serious” and/or because external pressure will prompt harsh punishment.</td>
</tr>
<tr>
<td>media, are associated with a decreased likelihood of acquittal, an</td>
<td></td>
</tr>
<tr>
<td>increased likelihood of mild punishment, and an even more increased</td>
<td></td>
</tr>
<tr>
<td>likelihood of harsh punishment.</td>
<td></td>
</tr>
<tr>
<td>$H_{3c}$ Complex cases involving multiple accusations and multiple</td>
<td>Generally supported with respect to the number of accused. Cases involving a large number of officers are apparently more likely to result in acquittal and less likely to result in punishment, possibly because of the increased cost in collecting incriminating evidence and performing collective hearings.</td>
</tr>
<tr>
<td>accused officers are associated with an increased likelihood of</td>
<td></td>
</tr>
<tr>
<td>acquittal, and a reduced likelihood of mild and harsh punishment.</td>
<td></td>
</tr>
</tbody>
</table>