H i C N Households in Conflict Network

The Institute of Development Studies - at the University of Sussex - Falmer - Brighton - BN1 9RE
www.hicn.org

CSI in the tropics
Experimental evidence of improved public service delivery through coordination

Daniela Collazos[*], Leopoldo Fergusson[†], Miguel La Rota[‡], Daniel Mejía[§] and Daniel Ortega[**]

HiCN Working Paper 342

November 2021

Abstract: This paper evaluates the impacts of increased coordination, accountability, and leadership among teams of responsible public officials, with evidence from homicide investigations in Colombia. We randomly assigned the investigations of 66% of the 1,683 homicides occurring in Bogotá, Colombia, during 2016 to a new investigation procedure emphasizing these features. We find a statistically significant 30% increase in the conviction rate in the treatment group relative to the control group. Indicators of the quality of the investigative process also improve, as well as the rate at which a formal accusation is presented before a court. Complementary findings suggest that the treatment produces well-coordinated teams that can communicate more fluently. Also, a survey of investigative team members reveal that work motivation, the extent to which they receive feedback on their performance, the pertinence and effectiveness of their roles, and the perceived quality and coordination of the team all improve under the new scheme.

Key words: Crime, Homicides, Team work, Public sector

JEL Codes: C93, D73, J45, K14, K42

Acknowledgements: We thank the Open Society Foundations and the Latin America Development Bank - CAF- for funding. Daniel Felipe Hernández and Andrés Felipe Rodríguez provided superb research assistance. Special thanks go to the functionaries at Fiscalía General de la Nación, in particular Carolina Bernal, Gina Cabarcas, Natalia Cardona, Nelson Campos, Blanca Duque-Morales, and María Milena Mendez.

[*] Universidad de los Andes, Facultad de Economía, Cra 1 No 18A - 12, Bogotá, Colombia. E-mail: d.collazos148@uniandes.edu.co

[†] Universidad de los Andes, Facultad de Economía, Cra 1 No 18A - 12, Bogotá, Colombia. E-mail: lfergusson@uniandes.edu.co

[‡] Co-director, Crime and Justice Lab, Email: mlarota@gmail.com

[§] Universidad de los Andes, Facultad de Economía, Cra 1 No 18A - 12, Bogotá, Colombia. E-mail: dmejia@uniandes.edu.co

[**] Latin America Development Bank - CAF, and IESA business school dortega@caf.com

1 Introduction

The provision of prompt and effective justice to its citizens is an essential duty of a capable and democratic state (Cappelletti & Garth, 1978; Ely, 1980). Besides its direct importance as a fundamental right, strengthening the justice system may also contribute to economic well-being (e.g., Chemin, 2009b, 2009a; Jappelli, Pagano, & Bianco, 2005; Lichand & Soares, 2014; Visaria, 2009).

Also, given the imperative to protect the inalienable human right to life, it is hard to think of a more important aspect of justice provision than the prevention, investigation and prosecution of homicide cases. Nevertheless, across the world and in developing regions in particular, deficiencies and challenges are commonplace (United Nations, 2019).

Improving homicide investigation and prosecution is therefore an urgent task, especially for countries with high homicide rates. This paper studies one effort to do so in Bogotá, Colombia. We implemented a randomized controlled trial to evaluate a change in the homicide investigation process. The new policy seeks to improve the quality of the investigation and its judicial outcomes, by overcoming the disconnect between the initial crime scene investigation, the pre-trial actions, and the trial.

The main change was the introduction of "homicide squads" comprising a prosecutor, 3–4 technical crime scene investigators (CSIs), and two detectives. These teams investigate multiple murder cases together, from the initial assignment until the charges hearing. Their introduction fundamentally changed the process of homicide investigations, seeking to improve the working relationship among prosecutors, technical investigators, and detectives, fostering a cooperative relationship between them from the outset and creating a sense of responsibility for each case they handle.

Cases were randomly assigned to the usual investigation protocol or to squads following the new procedure. The experiment was piloted for two weeks in December 2015, officially started on January 20, 2016, and lasted until December 4, 2016. To allow time for criminal procedures to take place, the results of the control and treatment teams were analyzed using data up to May 2019.

The new procedure produced a statistically significant increase of 30% in the conviction rate relative to the control group. This increase is explained by a series of intermediate effects on indicators of the quality of the investigative process (the number of actions performed during the investigation, and forensic tests and procedures that improve the quality of the evidence increased by 22–29% of a standard deviation) and on the rate at which a formal

accusation is presented before a court (increasing by 25% of a standard deviation).

Analysis of the underlying mechanisms tentatively suggests that the treatment produces well-coordinated teams that can communicate more fluently. A survey of investigative team members also suggests that the new scheme has improved work motivation, the extent to which they receive feedback on their performance, the pertinence and effectiveness of their roles, and the perceived quality and coordination of the team.

We make several contributions to the academic literature and policy discussion on homicide investigations. We build on a large literature in economics using field experiments to examine monetary and non-monetary incentives to increase worker productivity. While our treatment does not include differential monetary incentives, a series of non-monetary mechanisms play a role.

The results from empirical studies in this area and the underlying theory help interpret some of our findings. First, team work such as the one encouraged in our experiment creates "social incentives", or changes in behavior induced because of the presence and identity of co-workers (List & Rasul, 2011). Incentivizing teams presents both challenges and opportunities. On the one hand, moral hazard encouraging free-riding (Holmstrom, 1982) may reduce productivity (Corgnet, Hernan-Gonzalez, & Rassenti, 2015). On the other, peer pressure and competition may help motivate agents to work harder (Kandel & Lazear, 1992; Mohnen, Pokorny, & Sliwka, 2008; Mas & Moretti, 2009; Babcock, Bedard, Charness, Hartman, & Royer, 2015; Georganas, Tonin, & Vlassopoulos, 2015). Also, some existing research suggests that measures of cooperation and effort increase when team incentives are strengthened and that teams allow exploiting collaborative skills and create non-pecuniary benefits for workers (Berger, Herbertz, & Sliwka, 2011; Hamilton, Nickerson, & Owan, 2003).

In our setting, no explicit team incentives were provided (that is, the treatment groups were not rewarded as a function of teams' performance through rankings, tournaments, or other schemes). Still, and as confirmed by our survey evidence, a shared sense of responsibility to do well could have activated these mechanisms, including peer pressure within the groups, as well as facilitated exploiting collaborative skills.

A second explanation for increased productivity that emerges from our surveys is feedback provision. In theory, feedback could either increase or decrease effort, depending on factors such as whether past and future performance are complements or substitutes and on the type of information feedback reveals about workers' returns to effort (Lizzeri, Meyer, & Persico, 2002; Ederer, 2010). As Bandiera, Barankay, and Rasul (2013) show, feedback may moreover decrease productivity in teams when these form endogenously. Intuitively, by strengthening

incentives workers seek to form teams based on ability rather than friendship, and this hurts the ability of the group to ameliorate free riding. While we do not directly measure the impact of feedback on performance, our groups were fixed exogenously. This feature, together with the treatment group reporting that feedback improved in the new scheme, is consistent with it playing a positive role.

The preponderance of the literature on financial and non-financial incentives effects on worker performance focuses on the private sector. However, incentivizing public sector functionaries poses some special challenges (Dixit, 2002; Prendergast, 2001; Delfgaauw & Dur, 2008). For instance, since public agencies have to perform multiple tasks, respond to various stakeholders with potentially often-conflicting interests, might be staffed by "motivated agents" who subscribe to their agencies missions (Besley & Ghatak, 2005), and have hard-to-measure or vague goals, naive high-powered incentives like competition or performance-based rewards can generate dysfunctional reactions (Acemoglu, Fergusson, Robinson, Romero, & Vargas, in press) or be relatively ineffective.

For this reason, non-financial incentives of the sort induced in our study can be relevant and cost-effective strategies for improving government worker's performance. Despite this, the emerging and growing "personnel economics of the state" evaluating public worker's incentives has focused largely on financial incentives (Finan, Olken, & Pande, 2017). Our paper thus contributes to this literature by showing the potential benefits of strategies that are both low-cost and avoid the risks of high-powered incentives in a high-stakes and relevant area of public good provision.

Finally, a quantitative and qualitative literature in criminology and law discusses homicide investigation and prosecution, with a focus on identifying and analyzing best practices to solve homicide cases and increase the probability that offenders will be prosecuted (Carter & Carter, 2016; R. Davis et al., 2011; Hough, 2019; McClellan, 2008; Nicol et al., 2003; Wellford et al., 2019). However, to the best of our knowledge ours is the first study to experimentally evaluate a new criminal investigation procedure and its effects on both investigation and prosecution outcomes.[1]

The article proceeds as follows. Section 2 describes Colombia's homicide investigation process. Section 3 explains the intervention protocol used to randomly allocate cases to treatment and control groups. Sections 4 and 5 describe our data sources and the empirical

[1]There are, in contrast, several empirical studies on the impacts of policing on crime using natural or controlled experiments (examples include Blattman, Green, Ortega, & Tobn, 2017; Collazos, García, Mejía, Ortega, & Tobón, 2020; Di Tella & Schargrodsky, 2004; Klick & Tabarrok, 2005).

framework we used to estimate the effects of the homicide squad scheme on relevant outcome variables. Section 6 presents our main quantitative results, and Section 7 discusses the conclusions that can be derived from the findings.

2 Background: the homicide investigation process

2.1 Unsuccessful homicide investigation

Colombia is an ideal case study, given the poor performance of homicide investigations in the country, evident in the very low *indictment and charging rates* (i.e., the percentage of criminal cases for which a person is indicted or charged with committing a crime). In 2015, Colombia's overall homicide rate was 24 per 100,000 people; its indictment rate was 21.5%, and its conviction rate (percentage of cases with a conviction, out of those with an indictment) was 62%. Bogotá's homicide rate was 18 per 100,000 people; its indictment rate was 22.4%, and its conviction rate 74.6%. Therefore, a mere 17% of homicide cases produce a final conviction, even in Bogotá, where the conviction rate is above the country's average.

Another measure used in the literature is the *homicide clearance rate*, or the percentage of total cases solved in a year. According to the US Federal Bureau of Investigation's uniform crime reports[2], an offense is solved or cleared if "at least one person is arrested, charged with the commission of the offense, and turned over to the court for prosecution." In 2015, 61.5% of murder offenses were cleared in the United States. Likewise, the European Homicide Monitor[3] defines cases as cleared when police authorities have arrested a suspect, or when a suspect is known even if for some reason he cannot be arrested. Using this database, Liem et al. (2019) find that between 2009 and 2014 the clearance rate for intentional homicide ranged from an average of 98.2% in Finland to 77.1% in The Netherlands. Osborne, Lau, Britton, and Smith (2012) calculated clearance rates in England and Wales at 85% in 2012.

Given that a suspect is indicted in only one out of five homicides in Colombia, improving investigative effectiveness is crucial for its criminal justice system. Before explaining the changes we implement and evaluate in the study to achieve this goal, we next describe the steps of the investigation process in Colombia.

[2]See `https://ucr.fbi.gov/crime-in-the-u.s/2018/crime-in-the-u.s.-2018/topic-pages/clearances.pdf`, last accessed May 14, 2020.

[3]Developed by the Swedish National Council for Crime Prevention, the Institute of Criminology and Legal Policy at the University of Helsinki, and Leiden University.

2.2 The investigation process in Bogotá

In Colombia, two national agencies are responsible for investigating homicides: the Sectional Division of Crime Investigation of the National Police (hereafter SIJIN, for its Spanish acronym) and the Technical Division of Crime Investigation of the National Prosecution Office (CTI). The homicide investigation policy we evaluate was developed and implemented only within the CTI. The Automatic Dispatch Center of the National Police responds to an initial complaint – by civilians or police authorities at the crime scene – and informs a Satellite Unit (*Unidad Satélite*), which then assigns the case to either the SIJIN or the CTI. For CTI cases, when the initial complaint is made, the on-duty prosecutor from the closest Immediate Reaction Unit (URI) is also assigned to the investigation.

The procedure for investigating homicides in Bogotá has three major stages. The first is the initial investigation (*Indagación*), during which physical evidence is collected in order to conclude whether a crime was committed, and identify and arrest the offender. The second is the pre-trial stage (*Investigación*), which encompasses the indictment (*Imputación*) (i.e. the formal presentation of the alleged offender before a court of law, where the defendant is informed of future charges that will potentially be brought against him), and the gathering of corroborating evidence that proves the suspect's culpability. The final stage is the trial, starting with the charges hearing (*Audiencia de formulación de acusación*) in which the prosecution formally brings charges, the accused pleads guilty or not guilty, and the evidence is outlined to determine whether it is sufficient to proceed to trial.

2.2.1 Initial investigation (*Indagación*)

Once the CTI is assigned to investigate a case, it dispatches a team of CSIs and two detectives. The CSIs document the scene, collect material evidence, and prepare the body to transport it to the National Institute of Legal Medicine and Forensic Sciences (INML), where it is processed for external evidence and a forensic pathologist performs an autopsy. The detectives interview witnesses and follow leads, and execute any orders from the URI's prosecutor leading the investigation. After documenting the evidence, they write detailed reports of their activities at the crime scene: the lead CSI writes the record of the technical examination of the corpse (*Acta de inspección técnica a cadáver*) and the lead detective drafts the executive report (*Informe ejecutivo*).

Based on these reports, the URI's prosecutor decides if further enquiry is needed, and if so, instructs the detectives to keep investigating. Then, he or she (or another URI prosecutor)

prepares the methodological program (*Programa metodológico*) that summarizes the facts and proposes a criminal hypotheses, and sends it to the prosecutor in the "Life Unit" (*Unidad de Vida*) to which the case is transferred. At this point, the case may be further transferred to other prosecutors within the same unit. Life Unit prosecutors decide whether to continue with the criminal procedure – if a suspect was identified during the initial investigation – thus initiating the pre-trial stage, or to classify a case as unsolved if there are no active leads for detectives to pursue. However, many unsolved cases remain open with no progress because they are too serious to close, and there is not enough evidence to advance.

2.2.2 Pre-trial (*Investigación*)

The pre-trial stage begins when suspects are formally brought before a criminal court, so they understand their constitutional rights and what the charges against them are likely to be. When the offender is arrested, the prosecutor must bring the detainee before a guarantees judge (*juez de garantías*) within 36 hours to ensure the legality of the procedure and to determine if the person should be released on bail or held in custody. If the person is arrested *in flagrante*, the initial investigation and indictment occur simultaneously.[4]

The prosecutor has 30 days after the indictment to continue with the next step of the criminal procedure, the charges hearing, or to ask the judge for a dismissal. During this stage, detectives can continue searching for evidence to determine whether the suspect committed the crime.

2.2.3 Trial (*Juicio*)

The charges hearing (*Formulación de acusación*) is the first step of this stage, when charges are brought against the accused. The bill of charges (*Escrito de acusación*) is read and the defendant is asked to plead guilty or not guilty. If they plead guilty, the case proceeds to sentencing. Otherwise, the judge holds a preparatory hearing (*Audiencia preparatoria*) for a first round of evidence discovery, in which the court decides whether to continue to trial. In Colombia, trials are conducted before a judge without a jury.

[4]Colombia's Penal Code defines *in flagrante* as a situation in which a person is: (i) caught and apprehended while committing a crime, (ii) identified during the commission of a crime and apprehended immediately afterwards by hot pursuit, or (iii) apprehended with objects indicating probable culpability of having committed a crime. If a suspect is indicted, the prosecution can ask the court to detain them during trial if there is evidence that he or she is a flight risk, can manipulate evidence, or re-offend during the proceedings.

The next section describes an assessment of some potential problems with this procedure, and the changes introduced in our intervention to attempt an improvement.

3 Intervention protocol

Many factors affect the results of a homicide investigation process. Outcomes are sensitive to the criminological circumstances of specific locations, since organized crime or gang-related homicides are generally more difficult to solve. However, the capacity of the homicide investigation unit and the judicial system is also relevant. Identifying and overcoming problems and following best practices can improve the effectiveness of the investigations. For example, a lack of resources, a poor working relationship between prosecutors and investigators, and poor procedures for processing and analyzing evidence negatively affect the murder clearance rate (Cronin, Murphy, Spahr, Toliver, & Weger, 2007). Conversely, formal training of homicide detectives and the use of sophisticated analytical tools help solve homicides (Keel, Jarvis, & Muirhead, 2009). Also, a cooperative relationship between prosecutors and detectives may increase clearance rates and successful prosecutions.

La Rota and Bernal Uribe (2014) studied the homicide investigation process in Colombia and found that it entailed multiple officers acting at different stages under a diffuse leadership. Detectives and forensic teams were only in charge of specific tasks, and coordination was rarely effective. The authors concluded that these disruptions increased the probability of losing evidence and created a disincentive for investigators to tackle cases from start to end as a unified team. The results of La Rota and Bernal Uribe (2014) motivated the Unit for Strategy and Public Policy of the National Prosecution Office to restructure its murder investigation procedure more efficiently, to avoid the loss of information between public servants and increase coordination and leadership.

To evaluate this initiative, a fraction of homicide cases was assigned to follow the new process. We experimentally manipulate assignments so that a (random) fraction of homicide cases under the CTI's responsibility follow the new investigative procedure that seeks to overcome the disconnect between the three main stages of investigations. The main change, the introduction of homicide squads, changes the homicide investigation process in four main ways.

First, from the moment CSIs and detectives are dispatched to a crime scene, they know which prosecutor is in charge of the investigation. This allows them to communicate, receive direct orders, and clarify possible questions.

Second, the same prosecutor is responsible from the crime scene until the charges hearing. This eliminates the URI's involvement and possible transfers to different prosecutors. The only exceptions are when a case is transferred to a prosecutor who specializes in homicides with specific characteristics: juvenile offenders, when a forensic pathologist report is required to determine whether the victim died of non-natural causes, or when the case jurisdiction is outside Bogotá or of a different judicial branch (for example, military justice).[5]

Third, the prosecutor can meet with her squad to discuss the research strategy and the criminal hypotheses in order to design the methodological program.

Fourth, after the charges hearing, the case is transferred to a new prosecutor in the Life Unit, who brings the case to trial. This eases the burden on the prosecutor in charge of the investigation.

4 Data and variables

We rely on survey sources and administrative data on homicides reported to Bogota's Prosecution Branch Unit. We use actions that occurred between December 2015 and May 2019. Since some criminal procedures take long (specifically those at the trial stage), we analyze medium-term outcomes. This implies we look at some outcomes even after the experimental phase with random assignment of *new* cases ended. However, the route of those already assigned to the control and treatment groups remained intact after such phase ended.[6]

In defining our outcomes, we highlight two of Brookman and Innes (2013)'s four definitions of investigative success: (1) outcome success, which concerns the identification, prosecution, and conviction of suspects and (2) procedural success, which involves maintaining the integrity and quality of investigations. These definitions best capture the main objectives of the policy intervention in Colombia: improving the investigation process to identify the perpetrator and obtaining sufficient evidence to establish his or her culpability in court.

[5]If a forensic pathologist report is required, the case is transferred to a special prosecutor called the 112 prosecutor from the Life Unit, who waits for the pathologist's report; if it is confirmed to be a murder, the pathologist returns the case to the initial prosecutor for further investigation. Otherwise, the 112 prosecutor closes the case since no crime was committed.

[6]The random assignment stopped in December 2016. Analysis of the longer-term outcomes was not included in our pre-analysis plan.

4.1 Administrative data and variables

Our main outcomes are actions and decisions taken by the CSIs, detectives, and the prosecution. Those from detectives and CSIs usually involve activities at the crime scene and additional investigative procedures to find and analyze physical evidence (for example, photographing the crime scene, collecting fingerprints or biological fluids, and interviewing witnesses). The prosecution's actions consist of three types: (1) instructing detectives to perform additional investigative actions, (2) making decisions, such as indicting or filing charges against a person, and (3) attending hearings and presenting petitions to a judge, including orders to interview specific witnesses or asking to detain a suspect.[7]

Some actions are conditional on preceding ones and can only occur at certain stages of the process. Therefore, we estimate the effect of the intervention on actions at each stage of the investigation. The set of possible actions or outcomes in a case is extensive, and depends on the characteristics of the homicide.

The main actions on cases must (in theory) be reported by detectives and prosecutors from the Oral Accusatory Penal System (*Sistema Penal Oral Acusatorio*, SPOA, the name of Colombia's criminal justice regime). This is our main data source. It is an automated mandatory reporting and follow-up system, yet officials do not always promptly comply with it. When reporting in the system, they can either choose an action from a menu or write it in their own words, so the same activity might be reported in different ways. This presents challenges for the evaluation, which motivates us to measure outcomes in more than one way and rely on as much complementary evidence as we can. The database contains information on all crimes investigated by the prosecution, and we focus on the section of felonies against persons such as assault, manslaughter, murder, etc.

We also use information from the Integrated System of Management of the Judicial Police (*Sistema Integrado de Gestión de la Policía Judicial*, SIG), which is used to report prosecutors' orders to detectives. This database contains more complete information than the SPOA on investigative actions.

Finally, we also have access to the reports written by the CSI teams, detectives, and prosecutors for each case in the sample – "corpse examination reports," "executive reports" and "methodological programs," respectively. We analyze these texts to explore the quality of the teams' investigative work.

There are three challenges associated with correctly measuring each outcome and esti-

[7]The judges respond to prosecution petitions and decide whether their actions are legal, and assess the evidence and legal arguments presented during trial to decide on the responsibility of the accused.

mating the effects of the intervention. First, having a large number of outcomes increases the probability of falsely rejecting the null hypothesis (Anderson, 2008; Romano & Wolf, 2010, 2016). Second, SPOA data contains measurement errors, since not every action is reported. Third, the same action may be labeled differently in the system by different officials.

To address these problems, we start by aggregating into a single action all reports of similar activities. For example, actions like finding documents, analyzing databases, studying documents, and all similar actions are classified as "search and analysis of documents and databases." We exclude actions that do not constitute potentially meaningful changes in the quality of the process, such as administrative tasks that are unlikely to change with treatment or to influence the course of the investigation. We also drop actions for which we lack a clear hypothesis regarding whether they should be affected by the intervention, such as indicators of the occurrence of certain control hearings.

We then use the reclassified actions to create summary indices that combine actions to both reduce the number of hypotheses and to produce more precise measures of performance. To create the indices, if necessary, we first switch the signs of variables so that increases indicate a better outcome. We convert all outcomes to standardized "z-scores" by subtracting the control group's mean and dividing by the control group's standard deviation. Finally, we construct the indices as the unweighted average of z-scores for similar actions (Kling, Liebman, & Katz, 2007; Casey, Glennerster, & Miguel, 2012). Effects on the indices can thus be interpreted as mean effect sizes relative to the standard deviation of the control group. We create the indices for only the first stage of the investigation process, because this is where the largest number of actions is undertaken.[8] For the next two stages we focus on the effect of the intervention on individual outcomes.

After classifying actions and creating summary indices, we still have multiple outcomes. Thus, we adjust the p-values to account for multiple inferences. We follow Romano and Wolf (2005, 2016)'s resampling-based stepdown multiple testing method to control for the family wise error rate – the probability of type I error. We also pre-registered our analysis to protect ourselves from choosing estimates from among the large number of indicators and specifications to capture the effects of the intervention, which would effectively invalidate

[8]Because not all additional forensic and investigative actions, on which the second of our indices is based, are properly reported, we focus only on those more commonly reported in the data. To define frequency, in the case of this index (defined in Appendix Table A-1), we use the distribution of the total number of times each action is reported and drop those that are in lowest decile. We do this without comparing the distribution between the treatment and control groups and before estimating the treatment effect. Doing this limits us to working with the part of the data with less measurement error, which reduces the bias.

the inference (Collazos, Fergusson, La Rota, Mejía, & Ortega, 2017). We highlight the (few) occasions in which we had to depart from a choice as planned in the pre-analysis, mainly as a result of practical obstacles related to the data or implementation.

Appendix Table A-1 describes the list of outcome variables studied, stage by stage. Since our main interests are the outcomes related to the actions performed at the crime scene and to the indictment, we also study them individually, verifying their non-adjusted p-values. Table A-1 lists the outcomes for which we focus on non-adjusted p-values, as stated in our pre-analysis plan.

4.2 Survey data and variables

We administered two surveys. First, CSI teams, detectives, and prosecutors were surveyed at baseline to help evaluate the mechanisms that may explain the direct effect of the intervention. They were questioned about their roles, motivations, and job satisfaction, the quality of their work, and the importance of teamwork. A follow-up survey was conducted at the end of the intervention. We also created four standardized indices to aggregate several outcomes, and also evaluate some outcomes individually:

1. Motivation and feedback;
2. Team members' roles, effectiveness, and quality;
3. Teamwork;
4. Coordination and progress.

In each case, several questions (with answers scaled from 1 to 10) are collected in the index by topic. For example, the motivation and feedback index includes respondents' opinions on the extent to which they receive feedback from their superior about their work performance, how satisfied they are with the support from the Prosecution Office to do their job, how motivated by and satisfied they are with their work, and how responsible they feel for the successes and failures achieved in solving a homicide. Appendix Table A-1 lists all questions used by index. We examine three types of results individually:

1. *Information*, captured by responses to the question "How much do you agree or disagree with the following statement? Detectives and CSI are not sufficiently informed about the progress and results of their investigation" (strongly disagree, disagree, neither agree nor disagree, agree or strongly agree).

2. *Overall efficiency*, captured by responses to the question "How much do you agree or disagree with the following statement? There are often efficiency problems (such as loss of information or evidence, duplication of tasks, wasted work) during a murder investigation" (strongly disagree, disagree, neither agree nor disagree, agree or strongly agree).

3. *Victims*, measured as the percentage of time the respondent spends on attention to the victim's family in a typical work week.

5 Empirical framework

Our unit of analysis is the case. Whether it involves one or more murders, we treat the entire case as a single entity and focus on its results. Although the intervention was designed to treat only murder cases, it is the investigation process that determines which felony the defendant is charged with. Thus all deaths suspected to be due to non-natural causes – excluding car accidents – and that are investigated by the CTI are eligible to participate in the experiment. These include suicides and illegal abortion cases.

In order to randomly allocate homicide cases between treatment and control groups, in cooperation with the Prosecution Office we placed strict rotation rules on the investigative teams on call for each shift. The first step was to form six fixed investigative units (nine lab experts, five detectives, and a prosecutor) that would cover the city of Bogotá. Four of the groups were assigned to a treatment status via a simple raffle. Appendix Table A-3 shows the organization of the treated groups. There are two 12-hour shifts per day, and one of the six groups was assigned to cover each part of each shift. The groups' rotation throughout the month implies that every group covers each of the 14 possible weekly shifts at least once every 6 weeks (2.3 shifts per week). Each shift was covered by every team approximately seven times over the course of the 42-week experiment. Controls covered each shift at twice the rate since there were two control groups. Contrary to the treatment groups, in the control groups there are no fixed teams of prosecutor, CSIs and detectives treating each case, and the process instead followed the traditional scheme described in Section 2. Our intention-to-treat estimates compare homicide cases that fall into treatment shifts versus control shifts. Appendix Table A-4 shows the distribution of treatment groups across possible shifts.

The key parameters to estimate are the direct treatment effects of the intervention on actions taken during each investigative stage. Consider a case c for which a specific procedure

y was carried out. We estimate the effects using an ordinary least squares (OLS) regression:

$$y_c = \beta_0 + \beta_1 \text{Team}_c + \beta_2 X_c + \beta_3 + \text{lag}_c + \text{days}_c + \delta_{month} + \gamma_{weekend} + \varepsilon_c \quad (1)$$

where $\text{Team}_c = 1$ if the case is in the treatment group (that is, if it is assigned to a homicide squad), and $\text{Team}_c = 0$ otherwise; X_c is a vector of case controls; lag_c is the difference in hours between the date of the complaint and date of death; days_c is the number of days since the Prosecution Office took on the case; δ_{month} and $\gamma_{weekend}$ are month and weekend fixed effects; and ε_c is the robust standard error clustered at the shift level.

To increase precision, we include a number of controls, described in Appendix Table A-2, as robustness checks. These are the age and gender of the victim, indicators for the location (type and city area) of the crime, and dummy variables that indicate if the case was transferred (to another city, to another unit within the General Prosecution, or to another agency), if the case started as an assault, if the case was initially assigned to the *Secretaría de Salud*, which certifies natural deaths, but was sent back to be investigated by the judicial police[9], if the case occurred on a weekend, if the case occurred at night, and if the suspect was arrested *in flagrante*.[10]

Table 1 presents descriptive statistics for the control variables and examines the balance between the treatment and control groups. In general, the differences between these groups are not statistically significant at conventional levels, as expected given random assignment. Most p-values for these differences are much larger than 10%, except for the victim's gender, but even in this case the magnitude of the difference is very small (1.86 versus 1.82 male in each group, respectively, for a variable that is coded 1 for women, 2 for male, and 3 for

[9]Importantly, this is a process that takes place *before* the cases get to the teams, so it cannot be affected by the treatment. However, it could affect outcomes because then the case takes longer to reach the judicial police laboratory specialists and investigators, which could compromise the quality of the evidence.

[10]We had to exclude some controls that we were planning to include as laid out in the pre-analysis plan because there were many missing observations or reporting issues in the data or they exhibited very little variation. We dropped the number of victims per case (only 28 cases, or 1.66%, had more than two victims); dummy variables that indicate if the case is femicide or abortion (only 11 cases (0.65%) and 13 cases (0.77%) fall into each category, respectively); and a dummy variable indicating whether the victim was unidentified because it is severely mismeasured in the data. In addition to dropping these variables due to the limited variation, we also dropped the variable indicating the type of crime since it is classified by prosecutors and could depend on the treatment and thus represent a "bad control" (Angrist & Pischke, 2008). We also include variables for the victim's age and gender, indicators for the location (type and city area) of the crime, and a dummy for whether the case occurred at night, which were not in the pre-analysis plan, but are arguably exogenous characteristics measured with better-quality data. While this motivated their inclusion, as will be made clear below, the results are not sensitive to the inclusion of controls. Thus, these departures from the plan do not affect our conclusions.

unknown). Beyond the means of the variables for treatment and control, in Figure 1 we plot histograms for the distribution of some of the key control variables – gender, age, location type, and location in the city. The figure reveals that both the averages and distributions of the variables look remarkably similar for the treatment and control groups. All of this suggests that the randomization worked as planned.

In our pre-analysis plan (Collazos et al., 2017), we used baseline data to carry out a placebo analysis in order to test if the randomization schedule alone affects some of the outcome variables. We apply the randomization schedule to the data from 2015 in order to get the distribution of the sample between the treatment and control groups. Out of a total of 1,667 cases, 568 would have been randomly assigned to the control status and 1,099 to treatment. We then create summary indices to test the effect of the randomization mechanism. After adjusting the p-values, we find no significant difference due to the randomization mechanism, and they are rarely bigger than 10% of a standard deviation in absolute value and often smaller.[11]

6 Results

We present the results by source. We first analyze the outcomes from administrative data collected from the SPOA and SIG systems (key variables are presented in Table 2), and follow with our survey data (illustrated in Table 9). All outcome variables are listed in Table A-1.

6.1 Administrative data on investigation and procedural outcomes

We use the SPOA data to measure five outcomes: indictment, bill of charges, sentence and conviction rates, and the number of days it takes to get to an indictment or charge a suspect. Panel A of Table 2 presents the statistics for the sample of cases in the SPOA system for which there is a recorded action (1,501 of 1,683). For the 182 cases without an action recorded, we cannot determine if no action was actually performed or if it was just not recorded. The treatment group has statistically significantly fewer cases with no action (mean=0.084, sd=0.008) than the control group (mean=0.153, sd=0.015). This indicates that the treatment group either acted in more cases, or at least was more disciplined in

[11] Given the long delays until a conviction, one exception is the percentage of cases with a conviction: the size is large and imprecise and we have a small sample.

registering the duties they performed.

The Prosecution Office uses indictment, bill of charges, sentence, and conviction rates to measure performance. All reported cases are divided into (1) those in which a crime was really committed and (2) non-effective entries – cases that do not involve a felony and are therefore closed. The rates are then typically defined as the number of cases with either an indictment or bill of charges divided by the total number of confirmed homicide cases. To capture this in our regressions at the case level, we examine regressions for dummy variables on restricted samples. First, we restrict the sample to cases with reported actions and identify those in which the criminal process ceased (1,280). Then we restrict the sample to the effective entries (230), and examine regressions for dummy variables for indictment, bill of charges, sentencing and conviction, effectively measuring how determinant the treatment is. Recall also that, in order to evaluate how robust and significant overall patterns are, we evaluate the impact of the treatment estimating equation (1) on standardized outcome variables, including the aggregated indices, and correct inferences for multiple hypotheses testing. Table 3 presents the results. P-values are presented for individual tests and adjusted for multiple hypotheses testing following Romano and Wolf (2005, 2016).

We expect indictment, bill of charges, and sentencing and conviction rates to improve for the treatment group. A better initial investigation process and the elimination of unnecessary transfers between units should increase the availability of evidence, raise the proportion of cases cleared, and increase the level of trial success in criminal court. The data shows that the average proportion of indicted cases is higher for the treatment group, but this difference is not statistically significant at conventional levels. Similarly, sentencing rates for the treatment group increase by around 18% of a standard deviation of the control group, but the estimate is not statistically significant. However, the positive differences for the bill of charges and conviction are more precisely estimated. For the bill of charges, we find evidence of a significant increase that equals about 25% of a standard deviation of the control group. This difference is significant under the OLS model but not under the corrected Romano-Wolf p-values. Lastly, the treatment improves the conviction rate by 30% of a standard deviation of the control group; this result is significant under the Romano-Wolf correction.

In summary, all our main performance indicators show the expected sign of the effect of the treatment; two show significant differences under OLS p-values, but only one – the conviction rate – is statistically significant when adjusting for multiple hypotheses testing. The combination of these results allows us to conclude that the treatment groups appear to perform better on average, at least at the pre-trial and trial stages. An important implication

of these results is that the effects of the treatment on a better investigation not only impact the immediate decision to charge taken by the treatment group prosecution. They also have a future positive effect on trial, which is controlled by prosecutors who are not directly involved with our treatment.

In addition, our results show that the treatment group sent fewer cases to the special 112 prosecutor, which is in line with expectations. Recall that these are cases requiring confirmation from the forensic pathologist's report. If the report finds grounds to suspect murder, the case is returned to its original prosecutor; otherwise, the case is closed. We expected these internal transfers to diminish as a result of treatment, as a better investigation should produce a higher degree of certainty about the existence (or not) of homicide offenses. The statistics reflect this hypothesis, as treatment groups are, on average, around 30% of a standard deviation less likely to send cases for further review. This difference is significant under the Romano-Wolf correction.

We expected the treatment to have two contradictory effects on the duration of investigative procedures. First, the time it takes to advance in the investigation and conviction process could decrease, given the elimination of reassignments between prosecutors, as well as the steps involving the URI. Conversely, a better ability to investigate complex cases should result in success for more prolonged investigations, extending the typical time it takes to obtain an indictment. Our results show that the treatment group takes 32% of a standard deviation longer, on average, to get an indictment, which could signal a bigger effort to produce better and more complex cases. But we lack conclusive evidence for this interpretation, in particular because the effect disappears and the magnitude of the estimate decreases to 7% of a standard deviation once control variables are included.

Interestingly, this does not translate into more delays for the bill of charges, which is on average faster in the treatment group – a non-statistically significant decrease of 18% of a standard deviation. One possible explanation reflects the conflicting expected effects: the treatment group might take longer to indict when advancing in the investigation because the homicide squads are taking more time to perform more detective and forensic actions, as well as tackling more complex cases. But treatment prosecutors would then be better prepared to quickly move to charge.

Turning to variables in the SIG database, Panel B of Table 2 presents descriptive statistics.[12] The *minimum actions* index captures a set of activities that are *always* expected

[12]We have fewer observations for the SIG database simply because some cases in the SPOA system do not appear in the SIG data, and because some cases that have actions reported in the SPOA might not have

to occur. We anticipated that the treatment would not have any effect on them, as these actions were almost always compulsory. In turn, the forensic actions index comprises several additional forensic tasks that can be performed in some cases. We expected a positive impact in the pre-analysis plan. According to the results reported in Table 4, the treatment improves both the minimum actions and the forensic actions indexes. The magnitude of the effect is economically meaningful, with increases of 27% and 22% of a standard deviation in the control group, respectively. The estimates are also precise with very small p-values. In short, our treatment results in a significant and sizable increase in measurable compliance with minimal investigative and forensic actions that must be undertaken in every case.

In addition, we examine three key variables captured in the SIG about the number of investigative orders that prosecutors issue to police detectives to: conduct interviews (a separate order must be placed for each interview), locate "persons of interest" (typically witnesses or suspects), and verify the identity of suspects. In line with our pre-analysis plan, we look at individual as well as Romano-Wolf corrected p-values. As expected, we find that the number of investigative actions increases for all types of orders in the treatment group, but the coefficients are small and insignificant.

We also use the SIG data to look deeper into the length of investigations in the treatment group, which seem to increase, at least as judged by the number of days until an indictment as measured in the SPOA data. Each order placed by prosecutors contains several actions that investigators must perform within a designated period. However, extensions can be requested. We therefore measure (i) the number of extensions requested in relation to the case, (ii) the average number of days of extension requested, and (iii) the number of additional days necessary to fulfill assignments; that is, how much longer than the deadline the investigators take.

From Table 4 we conclude that the treatment group requests far fewer – and shorter – extensions than the control group (decreases by about 15% of a standard deviation), even though they take longer to fulfill their assignments. One possible explanation for this apparent contradiction is that the treatment group is conducting a deeper analysis that requires more time, but it is being done within the expected deadlines. This interpretation is consistent with the one from the SPOA database, which reconciles the increase in the number of days until an indictment is handed down with a commensurate increase in the number of days until the bill of charges is read. The results when we include controls are quite similar.

actions in the SIG, which focuses only on orders placed, and registered, by prosecutors.

Finally, a potential concern may be that the results are biased because the treatment group prosecutors were not randomly selected. This could happen in two ways. First, the selected prosecutors could be the highest performing of all their peers; thus it would be natural to obtain positive results. Second, treatment prosecutors started with no initial workload, while control group prosecutors began with the workload of their ongoing cases.

To test these hypotheses, we compare the results obtained by the treatment group to those of two new prosecutors hired to work under the same policy but after the experiment ended. Table A-5 presents the results. In Panel A, we first test the effect of zero initial workload by comparing the cases known to the treatment teams and the new prosecutors immediately after the latter were hired. At that time, treatment prosecutors already had an 18-month backlog and new prosecutors had none. We find no significant differences between the two groups, indicating that the initial workload is not biasing our main results. To test whether the treatment prosecutors were "superstars," we compare the outcomes obtained by each team after they started working under the new scheme and up to six months later. During that period both teams started with zero workload and under the same investigative scheme, although in different months and years. Panel B of Table A-5 shows the results. We find statistically significant results under OLS for cease of criminal process, indictment, and bill of charges, although none survives the Romano-Wolf correction. Therefore, the new prosecutors hired after the experiment performed better than the treatment prosecutors, indicating that the latter were not necessarily the best.

6.1.1 Text analysis of written reports

To try to further understand the different implications of the team work imposed by the intervention, we now examine the written reports of the CSI teams, detectives, and prosecutors – corpse examination reports, methodological programs, and executive reports, respectively. The descriptive statistics for all variables in the text analysis are in Table 5. Like every other outcome, these variables are described in Table A-1.

The corpse examination report contains a detailed description of the criminal investigation work at the crime scene. The first three variables we look at simply measure the length of this description. The first is the simple raw count of words. The second and third correct the raw count by removing "stop words" like pronouns and connectors that appear very often but convey little meaning. We analyze these variables using two main methods. First, we count exactly how many words are left in each text after removing the stop words. Second, we remove words that are very rare in the text. We then look at the proportion

of key words organized by topics that are important in the process of investigation, such as first responder, interviews, prosecutors, investigators, evidence, and laboratories.[13] The rationale for including first responders is that much of the success in the investigation may relate to the quality of the work performed at the crime scene by the first responder. The rest of the topics capture key inputs in the investigation process.

The results, presented in Table 6, contain three main findings. First, the reports are nearly 0.5 of a standard deviation shorter in the treatment group compared to the control group. Second, there is more discussion of first responders (25% of a standard deviation effect) and more mentions of prosecutors and investigators (a large 56% of a standard deviation impact). Both of these suggest that treatment groups were paying more attention to key inputs in the investigation, and were taking more of each other into account, which is consistent with a better information flow and coordination between the homicide squads, a prime objective of the intervention. But surprisingly, we observe fewer mentions of evidence (25% of a standard deviation decrease). Third, the remaining topics reveal a substantially smaller difference between the treatment and control groups, with the possible exception of laboratory reports, which are 15% of a standard deviation longer in the treatment group, but with larger p-values.

We also analyze a second description in the examination report, focused on findings rather than the process. The treatment group also has more concise texts (close to 13% of a standard deviation shorter), which is less statistically significant than the process narrative. A similar conclusion applies to the amount of material evidence collected, which is also in line with the finding in the process narrative. A new variable, the number of procedures requested by the INML, also reveals fewer procedures in the treatment group, in this case with a larger effect of 32% of a standard deviation.

There are missing observations in our analysis of the methodological program (descriptive statistics in Panel B of Table 5). Table 7 shows that there are about 9% fewer observations in the treatment group. Thus, the results for the analysis of this type of document must be interpreted with caution.

In the methodological programs, investigators and prosecutors are supposed to write their investigative and criminal hypotheses. We count these hypotheses and find that they are more common for each type and as a whole in the treatment group. While this seems to suggest better work by the treatment group, a more in-depth examination raises questions about this explanation. We use three methods to analyze how similar the research and

[13]Keywords for each topic are listed in Appendix Table A-7.

criminal hypotheses are to each other (Table 7):[14] (1) the *token distance* focuses on comparing the words in each text, considering them closer if they share more words; (2) the *Damarau-Leveshtein (DL)* method estimates distance using the minimal number of changes needed to transform one text into another, so fewer transformations means less distance; and (3) the *Qgram* method is similar to the token distance, but uses all possible sequences of Q characters instead of words to assess similarity between texts, considering them closer if they share more of these sequences (our estimates use $Q = 4$).

In all cases, the treatment group produces more similar hypotheses than the control group, with sizable estimated effects. Thus it appears that treatment groups generally rely more on hypothesis "templates" that are copied and pasted for different cases, but are not really meaningful, than on coordination. We confirm this by measuring the similarity of the hypotheses in each case with those in *other cases* (last two rows of Table 7), and again find more homogeneity in the treatment group, with very large effects of more than two standard deviations. Figure 2 shows some particularly revealing examples of the type of vacuous writing that is likely producing these results.

Table 8 reports the results for executive reports, with descriptive statistics for the relevant variables in Panel C of Table 5. As we found for the corpse examination documents, executive reports are shorter in the treatment group (with effects close to 15% of a standard deviation). We also assess the frequency of key topics in the investigation process. We count keywords related to the following topics: witnesses, family members, videos, interviews, verification tasks (for example, verify information and testimonies against hard information, such as making sure an address or person or landmark cited by a witness exists), inspection tasks (for example, inspect places of interest, review documents, and other general activities that do not fall under a pre-established category, like visiting and inspecting the landmark cited by a witness after having verified its existence).[15] We find no outstanding robust differences between the two groups.

The text data for the written reports is inherently noisy, which complicates the analysis. However, it is clear that treatment group reports are shorter. One possible explanation is that improved coordination within treatment teams facilitates direct communication and decreases the need for paperwork. The reliance on templates could also be consistent with such an interpretation, as is the fact that when they do write they seem to mention team

[14]See Cohen, Ravikumar, and Fienberg (2003) for more details. All methods are implemented in R using the 'stringdist' package of Van der Loo (2014), see `https://cran.r-project.org/web/packages/stringdist/stringdist.pdf`.

[15]Keywords for each topic are in Appendix Table 2.

members more frequently, as well as some key aspect of the investigation input such as first responder actions. But at the same time, other key elements and processes appear less frequently, including evidence materials and other tasks such as forensic tests.

In any case, even if improved coordination results in better performance in the short term, a decline in the reliability of written reports may pose a medium-term challenge: the transparency of investigations may decrease, and investigative information may be less available for future purposes; for example, for trial-sourced examinations, audits, or inquiries about past cases.

While this is only a tentative conclusion, it relates to a relevant discussion about the potential tradeoff between performance and oversight under increased coordination. Coordination, by bringing different groups together, may increase performance. However, it could decrease oversight. Indeed, the separation between the prosecution and criminal investigation is common (though not universal) practice.[16] One interpretation is that choices over separation reflect different evaluations of this tradeoff.

The traditional Colombian model, in theory, has no separation: the prosecution leads the investigation with units that are an integral part of prosecution agencies. However, as we have seen, in practice functionaries from the CTI are part of the prosecution authority but work separately from prosecutors.[17] Thus, criminal investigation activities could be both ineffective *and* weakly supervised.[18] Our findings suggest that coordination costs do matter. Evaluating all relevant dimensions for oversight goes beyond our study's scope. The evidence on decreased quality of written reports might, as noted, compromise oversight. On the other hand, a number of additional results point at greater accountability, including some findings from our survey evidence which we review next.

[16]For instance, independent bodies perform these duties in the U.S and the U.K: the investigation depends entirely on the Police, and public attorneys or the Crown Prosecution Service evaluate the evidence to decide if they can prosecute. Some Continental-European or Latin-American jurisdictions such as France or Chile have hybrid models: prosecutors (or "Juges d'Instruction") start and have some leading role in the investigation, but independent Judicial Police bodies gather the evidence.

[17]Something similar occurs in Mexico, where some Judicial Police squads work separately from prosecutors, even if they are both parts of the same Mexican State Prosecution Office

[18]Interestingly, the trade-off between oversight and effectiveness does not seem to motivate the architecture of law enforcement agencies according to our interviews with heads of these investigative bodies. Instead, they see the division of labor as positive for effectiveness by encouraging specialization by expertise. Bogotá's system with three separate components (prosecutors, detectives, and forensic investigators) might reflect these beliefs. Another possible origin of these divisions might be an influential police culture inside agencies, inheriting a hierarchical, quasi-military archetype of various Latin American police bodies that are suspicious of, and impermeable to, civilian interaction (D. E. Davis, 2008; Hinton, 2005).

6.2 Survey data

We use our survey data to examine the perceptions of judicial operators along four indices: (i) motivation and feedback, (ii) team members' roles, effectiveness and quality, (iii) team work, and (iv) coordination and progress, as well as three individual variables (sufficient information, overall efficiency, and time spent with victims). These variables are described in Section 4.2 and Table A-1, while Table 9 reports the descriptive statistics (the differences in observations reflect that non-response is an option). To protect subjects, we do not register their individual identities. But we do have their treatment condition (recall that randomization occurs at the case level, but functionaries may be either in homicide squads or operating as usual). There is, unfortunately, a differential survey attrition between the treatment and control groups (as revealed by the treatment dummy in each panel). Although the sample is equally divided at the baseline, it shifts at the endline to around 30% from control and 70% from treatment. Thus the survey results must be interpreted with caution, as there may be non-random selection at the endline.

Our four indices are computed after standardizing the components in each index (hence the mean zero). Agreement with not having enough information and having efficiency problems is reported on a 1 to 5 scale, which increases as the respondent agrees with the statement. Time spent with the victims is the subjective perception of the fraction of time operators devote to victims.[19]

Controls in the survey are listed in Table A-8, and include sex, age group, marriage status, experience, and indicators for prosecutors, educational background, number of children, and data on training received. These controls show that functionaries in each group differ on some variables, and that these differences are sometimes significant (as revealed in the last column). This calls for more caution when interpreting these outcomes.

The results on the perception of operators at the end of the experiment are shown in Table 10. Outcomes without controls are illustrated in Panel A: compared to controls, functionaries in the treatment group have more positive views of the information flow with investigators, are less worried about efficiency problems, and report spending more time with victims. These findings all suggest better coordination and teamwork, as perceived by prosecutors. The regressions with controls (shown in Panel B) are essential in this exercise, as operators in the sample may be different between them. For this exercise, the index on coordination shows a large positive and significant effect, with close to 83% of a standard deviation in

[19] When the respondent's time adds up to more than 100%, the measure is adjusted in proportion to the total – larger than 100% – available time.

favor of the treatment groups, falling very much in line with the nature of the intervention. This effect remains robust in the model with controls and under the Romano-Wolf p-values. The effects on information flow and time with victims are preserved, although they are not significant under the Romano-Wolf correction.

These results rely on endline data from operators reflecting back on their work during the intervention. A different and supplemental approach is to compare before and after the experiment by performing a difference-in-differences exercise between baseline and endline information. While we do not have individual identifiers to construct a proper panel, as noted above, we do have treatment status at baseline and endline. We look at a model in which the four indices and three individual variables are used as outcome variables, which are regressed on treatment status, a dummy variable, and their interaction. The coefficient of the interaction is our quantity of interest, capturing the change in these variables for treated functionaries relative to controls.[20] Outcomes for this exercise are described in Table 11. The results are consistent with the expectations of the treatment. In the model without controls (Panel A), the motivation, qualification of team members, and coordination indices improve with treatment to a statistically significant degree at conventional levels. When including controls (Panel B), the size and significance are preserved for the quality of fellow team members and coordination indices.

In short, this section reveals tangible effects on job satisfaction, especially dimensions involving coordination, for functionaries dealing with cases in the treatment group. As explained, these results are more illustrative than conclusive, given the limitations of the survey's sample. However, they corroborate the findings from the administrative data, which show that treatment squads are more active and coordinated, and have a better performance than the previous approach.

7 Final discussion

Developing state capabilities is key to protect basic human rights, foster economic development, and achieve political stability (Acemoglu, 2005; Acemoglu, Moscona, & Robinson, 2016; Besley & Persson, 2010). Prime among these capabilities is delivering fair and effective criminal justice (Binder, 2011; Zedner, 2004). However, improving public sector outcomes faces formidable challenges. This article demonstrates the potential benefits of a strategy

[20]Since we focus on the change from the beginning to the end of the experiment, our identification assumptions are weaker, allowing for constant differences between the pool in functionaries in the two groups.

that is both low-cost and avoids the risks of high-powered incentives in a high-stakes and relevant area of public good provision: homicide investigation.

We study a new tactical arrangement designed to increase coordination and leadership in homicide investigation teams. The new scheme focuses on teams of forensic investigators, detectives, and prosecutors that work on individual cases from crime scene to trial, instead of specializing these operators in specific tasks on different cases.

Likely reflecting the impact of social incentives within a team, leadership, communication, and feedback, the results reveal that treatment groups perform better: they are able to charge and convict in a higher proportion of cases. The treatment appears to improve investigations not only during the investigative and pre-trial stages; the impact subsists until trial, even when homicide squads are no longer responsible for handling cases. We found positive and statistically significant results for two main indicators: an increase of 25% of a standard deviation for the bill of charges and of 30% for the conviction rate.

The evaluation also suggests that the new procedure increased the set of minimum investigative actions that is expected from operators on homicide cases. It also implies that more forensic tests and procedures were conducted in the treatment group relative to the control. The magnitude of the effects is economically meaningful, with estimated increases of 22–29% of a standard deviation of the control group. The treatment group appears to take longer to produce an indictment, but this does not delay the charging of suspects. The treatment group requests fewer, and less lengthy, extensions (decreases of about 15% of a standard deviation), even though they work longer to fulfill these assignments. One interpretation consistent with these results is that the treatment group is conducting a deeper analysis that requires more time, but it completes it within the expected deadlines.

Attempts to understand the underlying mechanisms of these results lead to some (more tentative) conclusions, based on the written reports for cases filed by the operators and surveys of their own perceptions of the work environment and conditions. The most robust finding is that cases are described in more succinct terms in the treatment group. We hypothesize that this is because the treatment produces well-coordinated teams that can communicate directly, and require less paperwork. Two complementary findings consistent with this view are that the treatment reports seem to rely more on templates, yet when the investigators do write they seem to mention team members more, as well as key aspects of the investigation like first responder actions. We leave full investigation of this hypothesis for future research.[21] However, even if better face-to-face coordination makes documenting less

[21] In addition, some key elements and processes in the investigation appear less frequently in the treatment

important in operators' everyday tasks, a decrease in reporting might affect accountability and undermine future criminal investigative efforts.

Finally, evidence from surveys of operators suggests that some key dimensions, like the extent to which they receive feedback on their performance, the pertinence and effectiveness of team members' roles, and the quality and coordination of teamwork, are higher under the new scheme. More tentatively, members of the homicide squads are better motivated, less likely to report having information deficiencies, and appear to spend more time with victims. Crucially, our findings are quantitatively large, and robust when it comes to operators' perceptions of the level of coordination and progress, which is in line with the nature of the intervention. However, these survey findings should be interpreted with some caution because there is differential attrition between the treatment and control groups from baseline to endline surveys, and because randomization is at the case level rather than the team level.

group reports, including material evidence and other tasks like forensic tests. Nevertheless, for this last category we have arguably better data than the written reports in the administrative data used to measure minimum and forensic actions, which instead robustly suggests that the treatment creates a more actively engaged investigation.

References

Acemoglu, D. (2005). Politics and economics in weak and strong states. *Journal of Monetary Economics*, *52*, 1199-1226.

Acemoglu, D., Fergusson, L., Robinson, J. A., Romero, D., & Vargas, J. F. (in press). The perils of high-powered incentives: Evidence from Colombia's false positives. *American Economic Journal: Economic Policy*.

Acemoglu, D., Moscona, J., & Robinson, J. A. (2016). State capacity and American technology: Evidence from the nineteenth century. *American Economic Review: Papers & Proceedings*, *106*(61–67), 601–619.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, *103*(484), 1481–1495.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton university press.

Babcock, P., Bedard, K., Charness, G., Hartman, J., & Royer, H. (2015, October). Letting down the team? Social effects of team incentives. *Journal of the European Economic Association*, *13*(5), 841–870.

Bandiera, O., Barankay, I., & Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, *11*(5), 1079–1114.

Berger, J., Herbertz, C., & Sliwka, D. (2011, April). *Incentives and cooperation in firms: Field evidence* (IZA Discussion Papers No. 5618). Institute of Labor Economics (IZA).

Besley, T., & Ghatak, M. (2005). Competition and incentives with motivated agents. *American economic review*, *95*(3), 616–636.

Besley, T., & Persson, T. (2010). State capacity, conflict, and development. *Econometrica*, *78(1)*, 1-34.

Binder, A. M. (2011). *Análisis político criminal: bases metodológicas para una política criminal minimalista y democrática*. Buenos Aires: Editorial Astrea De Alfredo y Ricardo Depalma.

Blattman, C., Green, D., Ortega, D., & Tobn, S. (2017). *Place-based interventions at scale: The direct and spillover effects of policing and city services on crime* (Working Paper No. 23941). National Bureau of Economic Research.

Brookman, F., & Innes, M. (2013). The problem of success: What is a goodhomicide investigation? *Policing and society*, *23*(3), 292–310.

Cappelletti, M., & Garth, B. G. (Eds.). (1978). *Access to justice vol. i: A world survey* (Vol. 4). Giuffrè Editore/Alphen aan den Rijn, Sijthoff/Noordhoff, European University Institute, The Florence Access-to-Justice Project.

Carter, D. L., & Carter, J. G. (2016). Effective police homicide investigations: Evidence from seven cities with high clearance rates. *Homicide Studies*, *20*(2), 150–176.

Casey, K., Glennerster, R., & Miguel, E. (2012). Reshaping institutions: Evidence on aid impacts using a preanalysis plan. *The Quarterly Journal of Economics*, *127*(4), 1755–1812.

Chemin, M. (2009a). Do judiciaries matter for development? Evidence from India. *Journal of Comparative Economics*, *37*(2), 230–250.

Chemin, M. (2009b). The impact of the judiciary on entrepreneurship: Evaluation of Pakistan's "access to justice programme". *Journal of Public Economics*, *93*(1-2), 114–125.

Cohen, W., Ravikumar, P., & Fienberg, S. (2003). A comparison of string metrics for matching names and records. In *Kdd workshop on data cleaning and object consolidation* (Vol. 3, pp. 73–78).

Collazos, D., Fergusson, L., La Rota, M., Mejía, D., & Ortega, D. (2017). Pre-analysis plan for "impact evaluation of a new homicide investigation policy in Bogot D.C., Colombia". Retrieved from `http://egap.org/registration/2526` (Registered at the Evidence in Governance in Politics repository)

Collazos, D., García, E., Mejía, D., Ortega, D., & Tobón, S. (2020). Hot spots policing in a high-crime environment: An experimental evaluation in Medellin. *Journal of Experimental Criminology*, 1–34.

Corgnet, B., Hernan-Gonzalez, R., & Rassenti, S. (2015, December). Peer pressure and moral hazard in teams: Experimental evidence. *Review of Behavioral Economics*, *2*(4), 379–403.

Cronin, J. M., Murphy, G. R., Spahr, L. L., Toliver, J. I., & Weger, R. E. (2007). *Promoting effective homicide investigations.* Community Oriented Policing Services, US Department of Justice.

Davis, D. E. (2008). Who polices the police. , 188–212.

Davis, R., Jensen, C., & Kitchens, K. (2011). *Cold-case investigations: An analysis of current practices and factors associated with successful outcomes.* RAND Corporation.

Delfgaauw, J., & Dur, R. (2008). Incentives and workers' motivation in the public sector. *The Economic Journal*, *118*(525), 171–191.

Di Tella, R., & Schargrodsky, E. (2004, March). Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *American Economic Review*, *94*(1), 115-133.

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of human resources*, 696–727.

Ederer, F. (2010). Feedback and motivation in dynamic tournaments. *Journal of Economics & Management Strategy*, *19*(3), 733–769.

Ely, J. H. (1980). *Democracy and distrust: A theory of judicial review*. Cambridge: Harvard University Press.

Finan, F., Olken, B. A., & Pande, R. (2017). The personnel economics of the developing state. In *Handbook of economic field experiments* (Vol. 2, pp. 467–514). Elsevier.

Georganas, S., Tonin, M., & Vlassopoulos, M. (2015). Peer pressure and productivity: The role of observing and being observed. *Journal of Economic Behavior & Organization*, *117*(C), 223-232.

Hamilton, B. H., Nickerson, J. A., & Owan, H. (2003, June). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, *111*(3), 465–497.

Hinton, M. S. (2005). A distant reality: Democratic policing in Argentina and Brazil. *Criminal Justice*, *5*(1), 75–100.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 324–340.

Hough, R. M. (2019). The investigation of homicide. *Homicide Studies*, *23*(2), 87-92.

Jappelli, T., Pagano, M., & Bianco, M. (2005). Courts and banks: Effects of judicial enforcement on credit markets. *Journal of Money, Credit and Banking*, 223–244.

Kandel, E., & Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy*, *100*(4), 801–817.

Keel, T. G., Jarvis, J. P., & Muirhead, Y. E. (2009). An exploratory analysis of factors affecting homicide investigations: Examining the dynamics of murder clearance rates. *Homicide Studies*, *13*(1), 50–68.

Klick, J., & Tabarrok, A. (2005). Using terror alert levels to estimate the effect of police on crime. *The Journal of Law and Economics*, *48*(1), 267–279.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, *75*(1), 83–119.

La Rota, M. E., & Bernal Uribe, C. (2014). *Esfuerzos irracionales: investigación penal del homicidio y otros delitos complejos*. Bogotá: Dejusticia.

Lichand, G., & Soares, R. R. (2014). Access to justice and entrepreneurship: Evidence from brazils special civil tribunals. *The Journal of Law and Economics*, *57*(2), 459–499.

Liem, M., Suonpää, K., Lehti, M., Kivivuori, J., Granath, S., Walser, S., & Killias, M. (2019). Homicide clearance in Western Europe. *European journal of criminology*, *16*(1), 81–101.

List, J. A., & Rasul, I. (2011). Field experiments in labor economics. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 4, p. 103-228). Elsevier.

Lizzeri, A., Meyer, M., & Persico, N. (2002). *The incentive effects of interim performance evaluations* (Penn CARESS Working Papers). Penn Economics Department.

Mas, A., & Moretti, E. (2009). Peers at work. *The American Economic Review*, *99*(1), 112-45.

McClellan, J. (2008). Evidence-based practice in criminal investigations. *Journal of Applied Security Research*, *3*(1), 37–56.

Mohnen, A., Pokorny, K., & Sliwka, D. (2008, January). *Transparency, inequity aversion, and the dynamics of peer pressure in teams: Theory and evidence* (IZA Discussion Papers No. 3281). Institute of Labor Economics (IZA).

Nicol, C., Innes, M., Gee, D., & Feist, A. (2003). *Reviewing murder investigations: an analysis of progress reviews from six police forces*. Home Office. Research, Development and Statistics Directorate.

Osborne, S., Lau, I., Britton, A., & Smith, K. (2012). Homicides, firearm offences and intimate violence 2010/11: Supplementary volume 2 to crime in England and Wales 2010/11. *London: Home Office*.

Prendergast, C. (2001, December). *Selection and oversight in the public sector, with the Los Angeles Police Department as an example* (Working Paper No. 8664). National Bureau of Economic Research.

Romano, J. P., & Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, *73*(4), 1237–1282.

Romano, J. P., & Wolf, M. (2010). Balanced control of generalized error rates. *The Annals of Statistics*, 598–633.

Romano, J. P., & Wolf, M. (2016). Efficient computation of adjusted p-values for resampling-based stepdown multiple testing. *Statistics & Probability Letters*, *113*, 38–40.

United Nations. (2019). Global study on homicide 2019. *Office on Drugs and Crime*. Retrieved from `https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide.html`

Van der Loo, M. P. (2014). The stringdist package for approximate string matching. *The R*, 2.

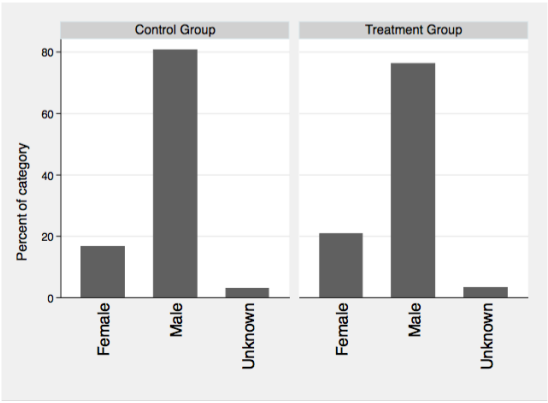
Visaria, S. (2009). Legal reform and loan repayment: The microeconomic impact of debt recovery tribunals in India. *American Economic Journal: Applied Economics*, *1*(3), 59–81.

Wellford, C. F., Lum, C., Scott, T., Vovak, H., & Scherer, J. A. (2019). Clearing homicides: Role of organizational, case, and investigative dimensions. *Criminology & Public Policy*, *18*(3), 553–600.

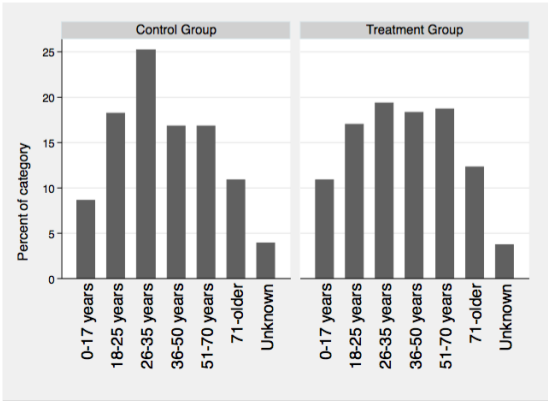
Zedner, L. (2004). *Criminal justice*. Oxford: Oxford University Press.

Figure 1: Distribution of some key control variables by treatment and control

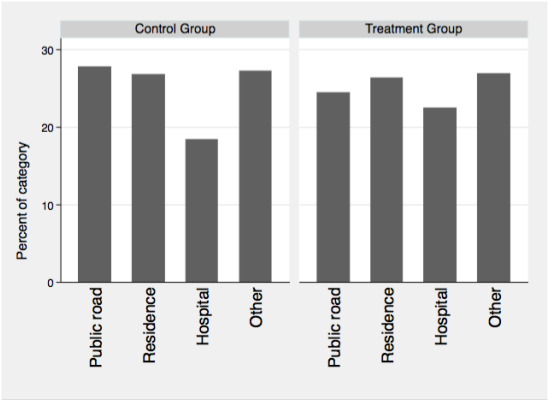
Victim's gender



Victim's age



Location (type)



Location (city area)

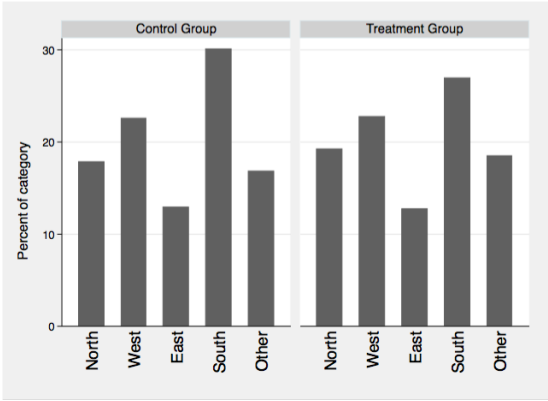


Figure 2: Examples of hypothesis found in the methodological program (Spanish)

Template Hypothesis	
Criminal Hypothesis	Research Hypothesis
De conformidad con los elementos materiales probatorios, evidencia física e información legalmente allegada a las diligencias, es factible inferir, de manera razonable, que nos encontramos frente a la ocurrencia de la presunta conducta punible de homicidio que consagra el código penal en su artículo 103, en las circunstancias temporo modales y espaciales de las que dan cuenta los emp, evidencia física e información legalmente allegada.	*Se llevaran a cabo diferentes labores investigativas para, de los indicios, testimonios, evidencia física e información legalmente allegada a las diligencias, intentar obtener todo el material probatorio para llevar ante el juez de conocimiento plena convicción de la materialidad del delito; sus circunstancias temporo modales y espaciales y la probable responsabilidad de los investigados.*
Not Real Hypothesis	
Criminal Hypothesis	Research Hypothesis
La señora xxx xxxx xxxx es encontrada por su hija en la tina del baño de su casa , sin vida. Se establecerá las causas de su fallecimiento.	*Se buscará información con sus allegados, del estado de salud de la hoy occisa, igualmente se establecerá ,a través de su eps , las enfermedades que la aquejaban. Se recepcionará entrevistas a las personas cercanas particularmente a la familia, que convivían con ella.*

Table 1: Control variables: descriptive statistics and balance

	Full sample			Control group	Treat. group	Mean Difference
Variables	Mean	Min	Max	Mean	Mean	(p-value)
Lag in days between complaint and death	2.835	0	680	4.063	2.201	1.862
	(25.550)			(29.607)	(23.168)	(0.157)
Days since start date	1,068.3	894	1,292	1,069.5	1,067.7	1.718
	(98.193)			(98.728)	(97.955)	(0.734)
Suspect arrested in flagrante	0.071	0	1	0.080	0.066	0.015
	(0.256)			(0.272)	(0.248)	(0.271)
Case started as an assault	0.056	0	1	0.058	0.056	0.002
	(0.231)			(0.233)	(0.230)	(0.884)
Case was transferred	0.037	0	1	0.031	0.040	-0.008
	(0.188)			(0.175)	(0.195)	(0.396)
Case ocurred during a weekend	0.367	0	1	0.380	0.360	0.020
	(0.482)			(0.486)	(0.480)	(0.418)
Case returned by Quincy	0.397	0	1	0.353	0.420	-0.067
	(0.489)			(0.478)	(0.494)	(0.007)
Time of report Night=1	0.474	0	1	0.471	0.476	-0.004
	(0.499)			(0.500)	(0.500)	(0.862)
Sex of the victim	1.839	1	3	1.864	1.826	0.038
	(0.445)			(0.421)	(0.457)	(0.099)
Age of the victim Group	3.666	1	7	3.628	3.686	-0.057
	(1.640)			(1.600)	(1.660)	(0.497)
Type of place of the diligence	2.494	1	4	2.450	2.517	-0.067
	(1.141)			(1.162)	(1.130)	(0.255)
City area of diligence	3.037	1	5	3.054	3.028	0.026
	(1.404)			(1.381)	(1.416)	(0.717)

Notes: The first group of columns shows the mean, the standard deviation in parentheses, and the minimum and maximum values for the full sample. This includes 1,683 cases: 573 in the control group and 1,110 in the treatment group. The second and third groups show the mean, and standard deviation in parentheses, for the treatment and control groups. The last column presents the balance between treatment and control with the corresponding p-value in parentheses.

Table 2: Descriptive statistics: Outcome variables in administrative data systems

Outcome	Full sample				Control group		Treatment group	
	Mean	SD	Min	Max	Mean	SD	Mean	SD
Panel A. SPOA database								
Cease of criminal process	0.892	0.310	0	1	0.876	0.330	0.900	0.301
Indictment	0.748	0.435	0	1	0.704	0.460	0.767	0.424
Days to indictment	176.75	280.95	0	1,145	117.42	259.31	201.06	286.83
Bill of charges	0.730	0.445	0	1	0.662	0.476	0.761	0.428
Days to bill of indictment	60.95	46.79	1	486	70.31	48.918	58.12	45.97
Sentencing	0.604	0.490	0	1	0.563	0.499	0.623	0.486
Conviction	0.578	0.495	0	1	0.493	0.504	0.616	0.488
Cases to be established sent to 112 prosecutor	0.380	0.486	0	1	0.471	0.500	0.333	0.472
Panel B. SIG database								
Minimun actions SIG-Mean effects index (z-score)	0.179	1.159	-0.724	15.718	0	1	0.270	1.221
Forensic actions SIG-Mean effects index (z-score)	0.146	1.258	-0.492	17.051	0	1	0.220	1.364
Order to interview	0.487	0.750	0	9	0.468	0.739	0.497	0.756
Order to locate person of interest	0.096	0.317	0	3	0.083	0.296	0.103	0.327
Order to individualize suspect	0.048	0.229	0	2	0.039	0.203	0.053	0.241
Number of extensions of the assignment	0.274	1.020	0	12	0.403	1.277	0.208	0.855
Average days of extension of the assignment	1.567	5.929	0	60	2.285	7.369	1.204	5.012
Days of delay to fulfill the assignment	11.101	32.655	-152	375	14.078	43.135	9.597	25.686

Notes: The first group of columns shows the descriptive statistics for the full sample, the second for the control group, and the third for the treatment group. The total sample in SPOA database is 1,683 cases: 573 in the control group, and 1,110 in the treatment group. The total sample in the SIG database is 1,612 cases: 541 in the control group, and 1,071 in the treatment group.

Table 3: Effects of the intervention on key outcomes: SPOA database variables

	No controls			*With Controls*			
	Treatment Effect	PValue	RW PValue	Treatment Effect	PValue	RW PValue	N
Cease of criminal process	0.071	0.187	0.518	0.060	0.171	0.524	1,501
		(0.054)			(0.044)		
Indictment	0.137	0.317	0.518	0.169	0.177	0.524	230
		(0.137)			(0.125)		
Days to indictment	0.323	0.071	0.300	0.072	0.678	0.672	172
		(0.177)			(0.173)		
Bill of charges	0.208	0.140	0.432	0.249	0.064	0.275	230
		(0.140)			(0.134)		
Days to bill of indictment	-0.249	0.201	0.518	-0.184	0.389	0.623	151
		(0.194)			(0.212)		
Sentencing	0.119	0.380	0.518	0.179	0.116	0.420	230
		(0.135)			(0.113)		
Conviction	0.245	0.069	0.300	0.309	0.008	0.050	230
		(0.134)			(0.116)		
Cases to be established sent to 112 prosecutor	-0.276	0.000	0.000	-0.330	0.000	0.000	1,683
		(0.056)			(0.048)		

Notes: The first group of columns shows the results without control variables: the treatment effect with the corresponding p-value and, in parentheses, the clustered standard error, and last the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016). The second group of columns shows the results including control variables in the regression. All variables are standardized by the mean and standard deviation of the control group. We exclude the newborn dummy variable since it applied to very few observations.

Table 4: Effects of the intervention on key outcomes: SIG database variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	No Controls			*Controls*			
	Treatment Effect	PValue	RW PValue	Treatment Effect	PValue	RW PValue	N
Minimun actions SIG-Mean effects index	0.270	0.000	0.000	0.294	0.000	0.000	1,612
		(0.059)			(0.056)		
Forensic actions SIG-Mean effects index	0.220	0.001	0.004	0.267	0.000	0.000	1,612
		(0.063)			(0.059)		
Order to interview	0.039	0.483	0.508	0.092	0.063	0.181	1,612
		(0.056)			(0.050)		
Order to locate person of interest	0.066	0.253	0.508	0.098	0.071	0.181	1,612
		(0.058)			(0.054)		
Order to individualize suspect	0.071	0.212	0.508	0.085	0.126	0.181	1,612
		(0.057)			(0.056)		
Number of extensions of the assignment	-0.153	0.001	0.009	-0.129	0.004	0.025	1,612
		(0.047)			(0.045)		
Average days of extension of the assignment	-0.147	0.003	0.014	-0.124	0.007	0.032	1,612
		(0.049)			(0.046)		
Days of delay to fulfill the assignment	-0.104	0.045	0.174	-0.103	0.041	0.151	1,612
		(0.052)			(0.050)		

Notes: The first group of columns shows the results without control variables: the treatment effect with the corresponding p-value and, in parentheses, the clustered standard error, and last the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016). The second group of columns shows the results including control variables in the regression. All variables are standardized by the mean and standard deviation of the control group.

Table 5: Descriptive statistics: text analysis outcomes

	N	Mean	Sd	Min	Max	P25	P75
Panel A. Corpse examination report							
Description Length (Raw)	1,513	977.970	638.107	11	8,675	596	1,153
Description Length (Method 1)	1,513	514.413	331.865	6	4,741	314	613
Description Length (Method 2)	1,513	501.169	322.442	6	4,620	306	594
Keywords by topic (in proportion to total length)							
First Respondent	1,513	0.280	0.288	0	1.987	0.000	0.417
Interviews	1,513	0.027	0.080	0	0.762	0.000	0.000
DA or Detectives	1,513	0.340	0.553	0	16.667	0.000	0.491
Crime Scene Data	1,513	0.492	0.357	0	3.407	0.260	0.676
Evidence (EMP)	1,513	0.397	0.441	0	4.056	0.000	0.593
Lab	1,513	0.387	0.398	0	3.315	0.131	0.509
Findings Description Length (Raw)	1,432	26.049	27.006	2	250	6	37
Crime Scene Drawing	1,595	0.391	0.488	0	1	0	1
Number of Evidence	1,644	2.715	3.840	1	81	1	3
INML Requests Length (Raw)	1,488	33.054	30.908	3	314	16	38
Panel B. Methodological program							
Has Methodological Program	1,683	0.905	0.293	0	1	1	1
Number of Criminal Hypothesis	1,523	0.735	0.481	0	2	0	1
Number of Research Hypothesis	1,523	0.349	0.481	0	2	0	1
Total Number of Hypothesis	1,523	1.084	0.804	0	3	0	2
Distance between hypothesis by method							
Method 1 - Tokens	515	0.489	0.128	0	1.000	0.421	0.521
Method 2 - DL	515	0.894	0.195	0	1.766	0.790	0.948
Method 3 - Qgrams	515	0.699	0.111	0	1.000	0.662	0.736

Continued on next page

Table 5 – Descriptive statistics, continued from previous page

	N	Mean	Sd	Min	Max	P25	P75
Distance with hypothesis from other cases by type							
Criminal Hypothesis	515	0.578	0.143	0.422	0.934	0.435	0.656
Research Hypothesis	515	0.602	0.136	0.449	0.983	0.460	0.693
Panel C. Executive report							
Narration Length (Raw)	1,557	558.766	432.319	37	5,804	302	687
Narration Length (Method 1)	1,557	271.546	209.705	18	2,746	147	330
Narration Length (Method 2)	1,557	279.322	215.517	19	2,835	152	339
Keywords by topic (in proportion to total length)							
Witnesses	1,557	0.228	0.466	0	5.000	0.000	0.340
Family	1,557	0.390	0.538	0	4.878	0.000	0.627
Video-Cams	1,557	0.391	0.715	0	8.696	0.000	0.567
Interviews	1,557	0.947	0.901	0	10.526	0.373	1.312
Verification	1,557	0.295	0.494	0	4.762	0.000	0.476
Inspection	1,557	0.378	0.509	0	3.333	0.000	0.602

Notes: See all variable definitions in Appendix Table A-1..

Table 6: Effects of the intervention on key outcomes: text analysis of corpse examination reports

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	No Controls			*Controls*			
	Treatment Effect	PValue	RW PValue	Treatment Effect	PValue	RW PValue	N
Description Length (Raw)	-0.486	0.000	0.000	-0.461	0.000	0.000	1,512
		(0.045)			(0.042)		
Description Length (Method 1)	-0.482	0.000	0.000	-0.452	0.000	0.000	1,512
		(0.045)			(0.042)		
Description Length (Method 2)	-0.476	0.000	0.000	-0.447	0.000	0.000	1,512
		(0.045)			(0.042)		
Keywords by topic (in proportion to total length)							
First Respondent	0.224	0.000	0.002	0.250	0.000	0.000	1,512
		(0.060)			(0.055)		
Interviews	-0.035	0.548	0.549	-0.033	0.551	0.540	1,512
		(0.058)			(0.056)		
DA or Detectives	0.637	0.000	0.000	0.560	0.000	0.000	1,512
		(0.101)			(0.086)		
Crime Scene Data	-0.091	0.097	0.181	-0.071	0.171	0.306	1,512
		(0.055)			(0.052)		
Evidence (EMP)	-0.289	0.000	0.000	-0.251	0.000	0.000	1,512
		(0.058)			(0.057)		
Lab	-0.117	0.026	0.076	-0.151	0.003	0.015	1,512
		(0.053)			(0.051)		
Findings Description Length (Raw)	-0.145	0.011	0.042	-0.097	0.056	0.157	1,432
		(0.057)			(0.051)		
Crime Scene Drawing	0.318	0.000	0.000	0.379	0.000	0.000	1,594
		(0.055)			(0.052)		
Number of Evidence	-0.115	0.008	0.041	-0.093	0.023	0.096	1,643
		(0.043)			(0.041)		
INML Requests Length (Raw)	-0.325	0.000	0.000	-0.279	0.000	0.000	1,487
		(0.057)			(0.053)		

Notes: Adjusted p-value calculated following Romano and Wolf (2005, 2016). Standard errors in parentheses. All variables are standardized by the mean and standard deviation of the controls. See Appendix Table A-1 for detailed variable definitions.

Table 7: Effects of the intervention on key outcomes: text analysis on methodological program

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	No Controls			*Controls*			
	Treatment Effect	PValue	RW PValue	Treatment Effect	PValue	RW PValue	N
Has Methodological Program	-0.497	0.000	0.000	-0.457	0.000	0.000	1,683
		(0.076)			(0.074)		
Number of Criminal Hypothesis	0.301	0.000	0.000	0.314	0.000	0.000	1,523
		(0.058)			(0.059)		
Number of Research Hypothesis	0.952	0.000	0.000	0.967	0.000	0.000	1,523
		(0.084)			(0.083)		
Total Number of Hypothesis	0.724	0.000	0.000	0.742	0.000	0.000	1,523
		(0.073)			(0.073)		
Distance between hypothesis by method							
Method 1 - Tokens	-0.860	0.000	0.000	-0.863	0.000	0.000	515
		(0.135)			(0.132)		
Method 2 - DL	-0.276	0.036	0.034	-0.306	0.024	0.026	515
		(0.131)			(0.135)		
Method 3 - Qgrams	-0.458	0.001	0.002	-0.470	0.000	0.001	515
		(0.132)			(0.129)		
Distance with hypothesis from other cases by type							
Criminal Hypothesis	-1.332	0.000	0.000	-1.387	0.000	0.000	1,076
		(0.077)			(0.075)		
Research Hypothesis	-2.287	0.000	0.000	-2.358	0.000	0.000	515
		(0.155)			(0.156)		

Notes: Adjusted p-value calculated following Romano and Wolf (2005, 2016). Standard errors in parentheses. All variables are standardized by the mean and standard deviation of the controls. See Appendix Table A-1 for detailed variable definitions.

Table 8: Effects of the intervention on key outcomes: text analysis on executive report

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	No Controls			*Controls*			
	Treatment Effect	PValue	RW PValue	Treatment Effect	PValue	RW PValue	N
Narration Length (Raw)	-0.140	0.010	0.074	-0.098	0.041	0.232	1,557
		(0.055)			(0.048)		
Narration Length (Method 1)	-0.152	0.005	0.049	-0.109	0.022	0.184	1,557
		(0.055)			(0.048)		
Narration Length (Method 2)	-0.147	0.008	0.058	-0.103	0.030	0.206	1,557
		(0.055)			(0.048)		
Keywords by topic (in proportion to total length)							
Witnesses	0.029	0.620	0.861	0.067	0.248	0.278	1,557
		(0.059)			(0.058)		
Family	-0.073	0.173	0.436	-0.103	0.049	0.232	1,557
		(0.053)			(0.052)		
Video-Cams	0.028	0.633	0.861	0.079	0.146	0.278	1,557
		(0.058)			(0.054)		
Interviews	0.104	0.079	0.288	0.100	0.079	0.278	1,557
		(0.059)			(0.057)		
Verification	-0.115	0.036	0.165	-0.090	0.097	0.278	1,557
		(0.055)			(0.054)		
Inspection	0.135	0.013	0.082	0.122	0.026	0.196	1,557
		(0.054)			(0.055)		

Notes: Adjusted p-value calculated following Romano and Wolf (2005, 2016). Standard errors in parentheses. All variables are standardized by the mean and standard deviation of the controls. See Appendix Table A-1 for detailed variable definitions.

Table 9: Descriptive statistics: survey outcomes

	N	Mean	Sd	Min	Max	P25	P75
Panel A. Baseline							
Indices							
Motivation and Feedback	200	-0.307	1.104	-2.679	4.164	-1.031	0.499
Role, Effectiveness and Quality	200	-0.107	1.122	-4.707	4.585	-0.690	0.623
Teamwork	200	-0.115	1.043	-2.847	1.444	-0.677	0.670
Coordination and Progress	200	0.002	1.499	-3.682	16.552	-0.697	0.588
Not Enough Information	197	3.102	1.381	1	5	2	4
Often Efficiency Problems	197	3.640	1.292	1	5	3	5
Time Dedicated to Victims	168	0.222	0.208	0	1	0.100	0.300
Treatment	172	0.430	0.497	0	1	0	1
Panel B. Endline							
Indices							
Motivation and Feedback	89	0.057	1.001	-2.977	2.083	-0.391	0.695
Role, Effectiveness and Quality	89	0.183	1.108	-2.781	1.862	-0.107	0.907
Teamwork	89	0.208	0.909	-2.740	1.587	-0.362	0.926
Coordination and Progress	89	0.342	1.059	-3.331	2.239	-0.238	1.178
Not Enough Information	85	3.059	1.339	1	5	2	4
Often Efficiency Problems	86	3.430	1.315	1	5	2	4
Time Dedicated to Victims	56	0.193	0.121	0	0.500	0.100	0.250
Treatment	82	0.695	0.463	0	1	0	1

Notes: See Appendix Table A-1 for detailed variable definitions.

Table 10: Effects of the intervention on key endline outcomes: functionaries' perceptions

	N	Treat. Effect	*Robust Errors* PValue	RW PValue
Panel A. No controls				
Indices				
Motivation and Feedback	82	0.172	0.460	0.712
			(0.231)	
Role, Effectiveness and Quality	82	0.289	0.268	0.603
			(0.259)	
Teamwork	82	0.370	0.095	0.317
			(0.219)	
Coordination and Progress	82	0.524	0.025	0.108
			(0.230)	
Not Enough Information	80	-0.576	0.017	0.091
			(0.237)	
Often Efficiency Problems	81	-0.045	0.838	0.832
			(0.219)	
Time Dedicated to Victims	52	0.471	0.007	0.035
			(0.166)	
Panel B. Controls				
Indices				
Motivation and Feedback	59	-0.078	0.782	0.778
			(0.280)	
Role, Effectiveness and Quality	59	0.384	0.265	0.700
			(0.341)	
Teamwork	59	0.324	0.271	0.700
			(0.290)	
Coordination and Progress	59	0.833	0.009	0.064
			(0.303)	
Not Enough Information	58	-0.718	0.058	0.284
			(0.368)	
Often Efficiency Problems	58	-0.271	0.319	0.700
			(0.269)	
Time Dedicated to Victims	38	0.547	0.065	0.284
			(0.282)	

Notes: See Appendix Table A-1 for detailed variable definitions.

Table 11: Effects of the intervention on key outcomes: functionaries perceptions Difference in difference estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Indices						
	Motivation	Role	Teamwork	Coordination	Information	Efficiency	Victims
Panel A. No controls							
Treatment group	-0.411*	-0.428**	0.169	-0.0177	-0.249	0.197	0.240
	(0.220)	(0.207)	(0.215)	(0.211)	(0.214)	(0.215)	(0.274)
Post intervention	-0.346	-0.317	0.0704	0.0157	0.423	0.0646	-0.783***
	(0.259)	(0.262)	(0.261)	(0.260)	(0.260)	(0.256)	(0.228)
Treatment × Post intervention	0.583*	0.718**	0.202	0.542*	-0.327	-0.242	0.231
	(0.319)	(0.332)	(0.307)	(0.312)	(0.319)	(0.307)	(0.320)
Constant	0.147	0.134	-0.0298	-0.00665	-0.178	-0.0267	0.249
	(0.171)	(0.167)	(0.177)	(0.180)	(0.176)	(0.188)	(0.200)
Observations	186	186	186	186	182	184	131
R-squared	0.022	0.032	0.030	0.064	0.045	0.008	0.091
Panel B.Controls							
Treatment group	-0.424	-0.491*	0.127	-0.0984	-0.186	0.146	0.0622
	(0.289)	(0.283)	(0.244)	(0.284)	(0.233)	(0.264)	(0.345)
Post intervention	-0.255	-0.393	-0.00115	-0.0213	0.462	0.118	-0.627
	(0.373)	(0.364)	(0.326)	(0.384)	(0.340)	(0.368)	(0.474)
Treatment × Post intervention	0.250	0.824*	0.167	0.793*	-0.504	-0.335	0.636
	(0.430)	(0.434)	(0.357)	(0.416)	(0.392)	(0.396)	(0.500)
Constant	2.021***	0.101	-0.367	-0.206	1.273**	-1.458**	0.459
	(0.738)	(0.636)	(0.611)	(0.714)	(0.590)	(0.619)	(0.788)
Observations	138	138	138	138	135	137	100
R-squared	0.284	0.191	0.251	0.253	0.321	0.206	0.382

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. See Appendix Table A-1 for detailed variable definitions.

A Additional tables and results

Table A-1: Outcome variables: definition and hypotheses

Outcome	Description	Hypothesis
	Administrative data	
First stage: Initial investigation		
Minimum actions index (Non-adjusted p-value)[22]	Mean effects index of set of actions that are usually performed in all cases. • This set comprises the following actions: search and retrieval of material evidence, interviews, photography of the crime scene, examination of the corpse, documentation of the crime scene and documentation of a location other than the crime scene.	We expect no statistically significant difference because minimum actions should be followed in all cases.
Forensic actions index (Non-adjusted p-value)	Mean effects index of set of additional investigative actions that can be performed in a case. Variables included depend on threshold frequency as noted in the main text. • This set comprises the following actions: ballistic analysis, search and analysis of documents, topography of the crime scene, documentation of fingerprints, digital storage and computer analysis, location of persons of interest and identification through photos, physical and chemical analysis components and other investigative actions.	We expect an improvement. These actions should be sensitive to the detectives' and CSIs' accountability to prosecutors, which we expect to be affected from the beginning of the investigation in the treatment status, as well as on their cooperation and coordination.
Order to interview (Non-adjusted p-value)	Number of orders issued by the prosecutors to the detectives to conduct interviews	We expect an increase
Order to locate person of interest (Non-adjusted p-value)	Number of orders issued by the prosecutors to the detectives to locate "persons of interest," typically witnesses or suspects	We expect an increase
Order to individualize suspect (Non-adjusted p-value)	Number of orders issued by the prosecutors to the detectives to verify the identity of suspects	We expect an increase

Continued on next page

[22]Deviation from the PAP: We decided to use only the SIG database to create the minimum actions and forensic actions indices as it is the system used by the CTI judicial police.

Table A-1 – Outcome variables, continued from previous page

Outcome	Description	Hypothesis
Number of extensions of the assignment	Assignments are sets of orders that have a deadline and the possibility of requesting extensions. Number of extensions requested.	We expect a decrease
Average days of extension of the assignment	Average days of extension requested	We expect fewer days to be requested
Days of delay to fulfill the assignment	Final days of delay to fulfill the assignment, that is, how much longer than the deadline the investigators are taking.	We expect a decrease
Cases to be established sent to the specialized 112 prosecutor *(casos por establecer)*	Cases requiring confirmation from the forensic pathologist's report. If the report confirms a murder has occurred, the case is returned to its original prosecutor. Otherwise, the homicide case is closed.	We expect a decrease. Detectives will pay more attention to the crime scene details and the evidence, thus determining more precisely than the control group if it is a murder case or not.
Cease of criminal process	Cases in which a criminal offense did not occur or when the requirements for further investigation were not met.	We expect a decrease. Better investigative work will lead to the identification of the suspect and the victim.
Second stage: Pre-trial		
Indictment *(imputación)* (Non-adjusted p-value)	The suspect(s) was (were) formally charged with the commission of the crime.	We expect an increase from a better initial investigation process, clearer leadership, and fewer unnecessary transfers between prosecutors and units.
Days to indictment (Non-adjusted p-value)	For cases with an indictment, time since the initial complaint to get to this point.	Two scenarios possible. We expect a decrease, due to the elimination of the URI step and the reassignments between prosecutors. But time may increase if new teams are able to tackle more complex cases that the control group may discard as too complicated.
Third stage: Trial		

Continued on next page

Table A-1 – Outcome variables, continued from previous page

Outcome	Description	Hypothesis
Bill of charges *(escrito de acusación)* (Non-adjusted p-value)	A formal written document accusing the suspect(s) of having committed a crime is filed.	We expect an increase, as the result of a better initial investigation process that allows the prosecutor to charge suspects.
Days to bill of charges (Non-adjusted p-value)	For cases with a bill of charges, time since the initial complaint to get to this point.	Two scenarios possible, for the same reasons as for the number of days to indictment.
Sentencing (Non-adjusted p-value)	Cases that had a sentence, although the defendant was not necessarily found guilty.	We expect an increase.
Conviction (Non-adjusted p-value)	Cases in which the defendant was found guilty.	We expect an increase. Cases will have better evidence to prove the suspect's guilt.
	Surveys	
Index 1: Motivation and Feedback	Aggregates four questions about motivation and feedback, in which the respondent answers with a number from 1 (lowest) to 10 (highest). a) How much feedback do you receive from your superior about your work performance? b) In general, how satisfied are you with the support you get from the Office of the Prosecutions to do your job? c) How motivated and satisfied are you with the work you carry out? d) How responsible do you feel for the successes and failures achieved in solving a homicide?	We expect an increase.
Index 2: Team members roles, effectiveness, and quality	Aggregates questions about roles, effectiveness, and quality. Answers from 1 (lowest) to 10 (highest). a) Are the duties that you prosecutor/detective/CSI must develop to solve a murder clear? b) In the development of a murder investigation, do you feel you can carry out all the tasks that are assigned to you? c) How effective do you think your team is on the ultimate goal of the prosecution – to reduce rates of impunity in the city? d) How satisfied are you with the quality of the homicide cases that you and your team investigate?	We expect an increase.

Continued on next page

Table A-1 – Outcome variables, continued from previous page

Outcome	Description	Hypothesis
Index 3: Teamwork.	Aggregates three questions about teamwork. Answers from 1 (lowest) to 10 (highest). a) When investigating a homicide, how aware are you of the daily tasks that other people on your team are doing? b) To what extent do you feel that your opinions are valued when making decisions to solve a homicide? c) Do you feel part of a team?	We expect an increase.
Index 4: Coordination and progress	Aggregates questions about coordination and the progress of the investigation. Answers: strongly disagree, disagree, neither agree nor disagree, agree or strongly agree. 1. The coordination of the team of detectives, CSIs, and prosecutors in investigating a homicide is good 2. The investigative actions taken to solve homicide cases are often extensive and sufficient 3. The evidence presented by detectives as the basis of the facts of a homicide case presented in the executive report is usually sufficient 4. Meetings with the team are usually carried out as often as necessary 5. There is a person on the homicide investigation team who is responsible for effectively coordinating the progress of the investigation and increasing its probability of success 6. The dynamics of the current work scheme promote the emergence of new ideas and useful innovations for documenting evidence and resolving cases 7. It is important that the prosecutor from the Life Unit is familiar with the details of the investigation from the start 8. When a case is more complex, it is easy to contact a specialist on homicides to ask for help 9. It is useful that the detectives and CSIs participate in defining the criminal hypothesis and ordering further investigation activities	We expect an increase.
Information	Proportion of respondents that agree detectives and CSIs are not sufficiently informed about the progress and results of their investigation. "How much do you agree or disagree with the following statement? Detectives and CSI are not sufficiently informed about the progress and results of their investigation." Answers: strongly disagree, disagree, neither agree nor disagree, agree or strongly agree.	We expect an increase.

Continued on next page

Table A-1 – Outcome variables, continued from previous page

Outcome	Description	Hypothesis
Overall efficiency	Proportion of respondents that agree: there are often efficiency problems during a murder investigation. "How much do you agree or disagree with the following statement? There are often efficiency problems (such as loss of information or evidence, duplication of tasks, wasted work) during a murder investigation" Answers: strongly disagree, disagree, neither agree nor disagree, agree or strongly agree.	We expect an increase.
Victims	Time spent assisting the victim's family. Percentage of time the respondent spends on attention to the victims in a typical work week.	We expect an increase.
	Text analysis	
Record of technical examination of the corpse		
Description length	Number of words written in the corpse examination report to describe the crime scene. This description includes all the findings and procedures performed at the scene in chronological order. For this outcome we have three variables that count the number of words in each description: one in which we count the exact number of words in the report (raw), a second that eliminates from the description words like pronouns and connectors that mechanically appear very often know as "stop words" (method 1), and a third variable that first removes stop words and then removes other words that are very rare in the entire set of texts analyzed (method 2).	We expect more details in the treatment group's description, which accounts for a larger number of words, as this should reflect the quality of the activities performed at the crime scene.

Continued on next page

Table A-1 – Outcome variables, continued from previous page

Outcome	Description	Hypothesis
Keywords by topic	Proportion of words related to each topic found in the description of the crime scene. We count keywords related to each topic as detailed in Table A-7 and then divide that number by the raw number of words in the description. The topics are: • First responder, which serves as a proxy for communication, indicating an interaction with this person about the circumstances in which he or she found the crime scene. • Interviews, another proxy for communication, indicating that the CSIs are aware of the interviews conducted by the detectives. • Prosecutors or detectives, which is a proxy for coordination indicating that the CSIs consider their work to be an input for their fellow team members. • Crime scene data, including words about photography, video, DNA, ballistics, and topography. These concepts are a proxy for quality in the crime scene documentation.	We expect an increase in the frequency of these words relative to the control group.
Findings description length	Number of words written in the corpse examination report to describe all the elements found in the crime scene that could be useful for the investigation, specifically in the process of examining the corpse (clothing, textures, gun powder residue, body fluids). This text was not processed in any way, hence we count the raw number of words in this section of the report.	We expect more details in the treatment group's description, which accounts for a larger number of words, as this should reflect the quality of the activities performed at the crime scene.
Crime scene drawing	Variable that takes a value of 1 if the CSI team reports having done a topographic drawing of the crime scene. This variable was manually coded by looking at each file.	We expect an increase with treatment (not in the pre-analysis plan).

Continued on next page

Table A-1 – Outcome variables, continued from previous page

Outcome	Description	Hypothesis
Number of evidence	Number of evidence elements collected by the CSI teams as it appears in the report. This variable was manually coded by looking at each file.	We expect an increase with treatment (not in the pre-analysis plan).
The INML requests length	Number of words written in the corpse examination report to describe the procedures to be conducted on the corpse requested from the National Institute of Forensic Medicine by the CSI team. This text was not processed in any way, hence we count the raw number of words in this section of the report.	We expect an increase with treatment (not in the pre-analysis plan).
Methodological program		
Has methodological program	Variable that takes a value of 1 if the methodological program was filed by the prosecutor.	We expect an increase with treatment (not in the pre-analysis plan).
Number of criminal and research hypotheses	Number of hypotheses that appear on the methodological program of each type. We also include a variable counting the total number of both types of hypotheses.	The quality of the investigation may lead to a more careful consideration of various hypotheses, but may also help to easily rule out unlikely hypotheses at the outset. We therefore have no clear prediction of the sign of a potential effect, if any.

Continued on next page

Table A-1 – Outcome variables, continued from previous page

Outcome	Description	Hypothesis
Distance between hypothesis	Inverse measure of how similar the two hypothesis texts are for each case. We calculate this variable only for cases that have both types of hypotheses. We use three methods to perform this calculation: (1) the token distance method focuses on comparing the words in each text, considering them closer if they share more words; (2) the Damarau-Leveshtein (DL) method estimates distance using the minimal number of changes needed to transform one text into another, where fewer transformations indicate less distance; and (3) the Qgram method is similar to the token method, but uses all possible sequences of Q characters instead of words to assess similarity between texts, considering them closer if they share more of these sequences. Our estimates use $Q = 4$.	If both hypotheses reported are exactly the same, this might mean that the fields are being filled only to meet a requirement. We expect the treatment squads to investigate the cases in greater detail and therefore to have more accurate hypotheses.
Distance with hypothesis from other cases	Inverse measure of how similar the hypothesis of each case is to hypotheses from other cases. One variable calculates this for criminal hypotheses and another one for research hypotheses. We create this measure as the average distance between the hypothesis of each case relative to all other hypotheses of the same type using the Qgram method with $Q = 4$ (explained in the row above).	If both hypotheses reported are exactly the same, this might mean that the fields are being filled only to meet a requirement. We expect the treatment squads to investigate the cases in greater detail and therefore to have more accurate hypotheses.
Executive report		
Narration length	Number of words written in the executive report describes what the research team did to solve the case. For this outcome we have three variables that count the number of words in the narration of each case. One counts the exact number of words in the report (raw), a second eliminates words like pronouns and connectors that appear very often ("stop words") (method 1), and the first removes stop words and then excludes other words that are very rare in the whole set of texts analyzed (method 2).	We expect more details in the treatment group's description, which accounts for a larger number of words, as this should reflect the quality of the activities performed at the crime scene.

Continued on next page

Table A-1 – Outcome variables, continued from previous page

Outcome	Description	Hypothesis
Keywords by topic	Proportion of words related to each topic. To calculate this proportion, we count the number of keywords related to each topic as detailed in Table A-7 and divide that number by the raw number of words in the narration. We analyzed the following topics: • Witnesses, to proxy for quality, indicating interactions with persons relevant to solving the case. • Family, another proxy for quality, indicating that the detectives had contact with the family of the deceased. • Interviews, also a proxy for quality, indicating the relevance of interviews conducted by the detectives for the case. • Verification and inspection also measure quality, as these concepts are related to important tasks performed throughout the investigation.	We expect an increase.

Table A-2: Control variables: description

Variable	Description
Suspect arrested *in flagrante*	Dummy variable that equals 1 when a suspect is arrested *in flagrante* at the scene.
Case started as an assault	Dummy variable that equals 1 if the case is an assault and turned into murder because the victims died from their injuries.
Transferred case	Dummy variable that equals 1 if the case was transferred to a special prosecutor or unit; typically reflecting a jurisdiction's competencies.
Case returned by Quincy	Dummy variable that equals 1 if the case was initially assigned to the *Secretaría de Salud*, which certifies natural deaths, but was sent back to be studied by the judicial police (CTI or SIJIN)
Time of report	Dummy variable that equals 1 if the case was reported during the night
Lag between complaint and death	Days from the time of the occurrence of the homicide and the date it is known by the Prosecution Office.
Sex of the victim	Categorical variable that equals 1 if female, 2 if male, and 3 if unknown.
Age of the victim	Categorical variable that equals 1 if the victim was 0–17 years old, 2 if 18–25, 3 if 26–35, 4 if 36–50, 5 if 51–70, 6 if 71 or older, and 7 if unknown.
Type of place	Categorical variable that indicates the type of place where the first investigative actions were carried out, which does not always correspond to the place where the victim was killed: 1 public road, 2 residence, 3 hospital, and 4 other.
City area	Categorical variable that indicates the geographic city where the first investigative actions were carried out: 1 north, 2 west, 3 east, 4 south, and 5 other.

Table A-3: Treatment group

Prosecutor	**CSI team**	**Groups of detectives**	
Prosecutor 1	13	**Group 1**	**Group 2**
		Detective 1	Detective 1
	14	Detective 2	Detective 2
		Detective 3	Detective 3
	15	Detective 4	Detective 4
		Detective 5	Detective 5
Prosecutor 2	4	**Group** 1	**Group 2**
		Detective 1	Detective 1
	5	Detective 2	Detective 2
		Detective 3	Detective 3
	6	Detective 4	Detective 4
		Detective 5	Detective 5
Prosecutor 3	1	**Group 1**	**Group 2**
		Detective 1	Detective 1
	11	Detective 2	Detective 2
		Detective 3	Detective 3
	12	Detective 4	Detective 4
		Detective 5	Detective 5
Prosecutor 4	16	**Group 1**	**Group 2**
		Detective 1	Detective 1
	17	Detective 2	Detective 2
		Detective 3	Detective 3
	18	Detective 4	Detective 4
		Detective 5	Detective 5

Table A-4: Number of times each treatment group covered each type of shift, January 20 to December 4, 2016

Shift	Group 1	Group 2	Group 3	Group 4	Control
Monday AM	7	8	7	7	16
Monday PM	8	7	8	8	14
Tuesday AM	7	7	8	8	15
Tuesday PM	8	8	7	7	16
Wednesday AM	8	8	7	8	15
Wednesday PM	7	8	8	8	15
Thursday AM	7	7	8	8	16
Thursday PM	8	7	8	7	16
Friday AM	8	8	7	7	16
Friday PM	8	7	7	8	16
Saturday AM	7	8	8	8	15
Saturday PM	8	8	7	8	15
Sunday AM	8	7	8	8	15
Sunday PM	7	8	8	7	16
Total	**106**	**106**	**106**	**107**	**216**

Table A-5: Robustness exercise: are the results driven by the treatment prosecutors?

	No controls			*With Controls*			
	Treatment Effect	PValue	RW PValue	Treatment Effect	PValue	RW PValue	N
Panel A: No initial workload							
Cease of criminal process	0.003	0.843	0.984	-0.006	0.616	0.941	2,421
		(0.014)			(0.012)		
Indictment	0.006	0.896	0.984	0.021	0.642	0.941	340
		(0.050)			(0.045)		
Days to indictment	16.446	0.382	0.883	-0.479	0.973	0.976	257
		(18.762)			(14.181)		
Bill of charges	0.014	0.789	0.984	0.030	0.519	0.920	340
		(0.054)			(0.047)		
Days to bill of indictment	-3.956	0.503	0.916	-4.972	0.312	0.862	222
		(5.902)			(4.909)		
Sentencing	0.064	0.245	0.769	0.049	0.345	0.862	340
		(0.055)			(0.052)		
Conviction	0.060	0.271	0.806	0.047	0.362	0.862	340
		(0.054)			(0.052)		
Panel B: Superstar prosecutors							
Cease of criminal process	0.012	0.657	0.887	0.472	0.066	0.241	626
		(0.028)			(0.256)		
Indictment	0.151	0.063	0.223	3.936	0.040	0.188	114
		(0.080)			(1.886)		
Days to indictment	-149.651	0.012	0.050	-10.863	0.993	0.992	87
		(58.102)			(1,189.079)		
Bill of charges	0.164	0.042	0.160	4.214	0.025	0.142	114
		(0.080)			(1.854)		
Days to bill of indictment	-3.939	0.802	0.887	664.628	0.128	0.301	80
		(15.693)			(430.885)		
Sentencing	0.115	0.241	0.580	0.255	0.891	0.985	114
		(0.098)			(1.867)		
Conviction	0.114	0.253	0.580	-0.791	0.651	0.906	114
		(0.099)			(1.743)		

Notes: The first group of columns shows the results without control variables: the treatment effect with the corresponding p-value and, in parentheses, the clustered standard error, and last the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016). The second group of columns shows the results including control variables in the regression. Variables are in levels, not standardized.

Table A-6: Robustness exercise: effects of the intervention on key outcomes, includes the same control variables as Table A-5 for comparison

	No controls			*With Controls*			
	Treatment Effect	PValue	RW PValue	Treatment Effect	PValue	RW PValue	N
Cease of criminal process	0.023	0.187	0.518	0.019	0.192	0.464	1,501
		(0.018)			(0.015)		
Indictment	0.063	0.317	0.518	0.071	0.187	0.464	230
		(0.063)			(0.054)		
Days to indictment	83.646	0.071	0.300	15.992	0.691	0.687	172
		(45.973)			(40.217)		
Bill of charges	0.099	0.140	0.432	0.105	0.074	0.312	230
		(0.067)			(0.059)		
Days to bill of charges	-12.185	0.201	0.518	-12.265	0.115	0.427	151
		(9.490)			(7.735)		
Sentencing	0.059	0.380	0.518	0.085	0.155	0.463	230
		(0.067)			(0.059)		
Conviction	0.123	0.069	0.300	0.140	0.022	0.111	230
		(0.067)			(0.060)		
Cases to be established sent to 112 DA	-0.138	0.000	0.000	-0.141	0.000	0.000	1,683
		(0.028)			(0.027)		

Notes: The first group of columns shows the results without control variables: the treatment effect with the corresponding p-value and, in parentheses, the clustered standard error, and last the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016). The second group of columns shows the results including control variables in the regression. Variables are in levels, not standardized.

Table A-7: Keyword Counting: Methodological Description

Keyword Outcome	Associated Words
	Panel A: Corpse Examination Report
First respondent	First respondent (*Primer respondiente*), Police (*Policía*), Patrolman (*Patrullero*), Nurse (*Enfermero*), Strecher-bearer (*Camillero*), Pathology assistant (*Auxiliar de patología*)
Interviews	Interview (*Entrevista y entrevistar*)
DA or detectives	DA (*Fiscal*), DA's office (*Fiscalia y despacho*), Investigator (*Investigador y saturno*)
Crime scene data	Photos (*Fotos y fotografía*), Video (*Video*), DNA (*ADN*), Ballistics (*Balística*), Topography (*Topografía*)
Evidence (EMP)	EMP (*Elemento de material probatorio*)
Lab	Laboratory (*Laboratorio y coral*)
	Panel B: Executive Report
Witnesses	Witness (*Testigo*)
Family	Family (*Family*), Relative (*Familiar*)
Video-Cams	Video (*Video y videográfico*), Cameras (*Cámaras y videocámaras*)
Interviews	Interview (*Entrevista y entrevistar*)
Verification	Verify (*Verificar y verificación*)
Inspection	Inspect (*Inspeccionar e inspección*)

Notes: This table shows the word associations used to count keywords related to some outcome variables. The second column shows these words in English and their Spanish original in parenthesis and italics. In the text analysis we accounted for simple variations of each word used, such as gender or plural variations for nouns and different tenses for verbs.

Table A-8: Descriptive statistics: survey controls

	Full sample			Control	Treat.	MeanDiff
Variables	Mean	Min	Max	Mean	Mean	(p-value)
	Panel A. Baseline					
Male	0.645	0	1	0.571	0.757	-0.185
	(0.480)			(0.497)	(0.432)	(0.011)
Age group						
18-25	0.005	0	1	0.000	0.014	-0.014
	(0.071)			(0.000)	(0.118)	(0.244)
26-30	0.045	0	1	0.010	0.097	-0.087
	(0.209)			(0.101)	(0.298)	(0.008)
31-35	0.116	0	1	0.061	0.181	-0.119
	(0.321)			(0.241)	(0.387)	(0.015)
36-40	0.177	0	1	0.153	0.153	0.000
	(0.382)			(0.362)	(0.362)	(0.996)
41-50	0.323	0	1	0.357	0.292	0.065
	(0.469)			(0.482)	(0.458)	(0.372)
50+	0.333	0	1	0.418	0.264	0.154
	(0.473)			(0.496)	(0.444)	(0.037)
Civil status						
Single	0.175	0	1	0.163	0.162	0.001
	(0.381)			(0.372)	(0.371)	(0.985)
Married	0.425	0	1	0.480	0.405	0.074
	(0.496)			(0.502)	(0.494)	(0.336)
Other	0.400	0	1	0.357	0.432	-0.075
	(0.491)			(0.482)	(0.499)	(0.319)
DA	0.355	0	1	0.653	0.054	0.599
	(0.480)			(0.478)	(0.228)	(0.000)
Years at DA's Office	14.678	1	35	16.776	12.274	4.502
	(8.075)			(7.118)	(8.394)	(0.000)
Months in Same Unit	71.589	3	299	69.286	69.750	-0.464
	(62.376)			(60.351)	(64.099)	(0.964)
Education level						
Incomplete primary	0.005	0	1	0.000	0.014	-0.014
	(0.071)			(0.000)	(0.116)	(0.251)

Continued on next page

Table A-8: Descriptive statistics: survey controls

	Full sample			Control	Treat.	MeanDiff
Variables	Mean	Min	Max	Mean	Mean	(p-value)
Complete primary	0.005	0	1	0.000	0.014	-0.014
	(0.071)			(0.000)	(0.116)	(0.251)
Complete secondary	0.165	0	1	0.122	0.243	-0.121
	(0.372)			(0.329)	(0.432)	(0.039)
Incomplete undergraduate	0.260	0	1	0.153	0.392	-0.239
	(0.440)			(0.362)	(0.492)	(0.000)
Complete undergraduate	0.160	0	1	0.163	0.149	0.015
	(0.368)			(0.372)	(0.358)	(0.796)
Incomplete graduate	0.015	0	1	0.000	0.041	-0.041
	(0.122)			(0.000)	(0.199)	(0.045)
Complete graduate	0.330	0	1	0.531	0.122	0.409
	(0.471)			(0.502)	(0.329)	(0.000)
Number of Children	1.621	0	8	1.643	1.658	-0.015
	(1.141)			(0.955)	(1.216)	(0.930)
Received Training	0.689	0	1	0.735	0.620	0.115
	(0.464)			(0.444)	(0.489)	(0.113)
Panel B. Endline						
Male	0.778	0	1	0.609	0.830	-0.221
	(0.418)			(0.499)	(0.379)	(0.037)
Age group						
18-25	0.011	0	1	0.000	0.000	0.000
	(0.107)			(0.000)	(0.000)	
26-30	0.069	0	1	0.040	0.088	-0.048
	(0.255)			(0.200)	(0.285)	(0.451)
31-35	0.080	0	1	0.040	0.070	-0.030
	(0.274)			(0.200)	(0.258)	(0.604)
36-40	0.276	0	1	0.200	0.316	-0.116
	(0.450)			(0.408)	(0.469)	(0.288)
41-50	0.310	0	1	0.400	0.281	0.119
	(0.465)			(0.500)	(0.453)	(0.291)
50+	0.253	0	1	0.320	0.246	0.074
	(0.437)			(0.476)	(0.434)	(0.490)
Civil status						

Continued on next page

Table A-8: Descriptive statistics: survey controls

	Full sample			Control	Treat.	MeanDiff
Variables	Mean	Min	Max	Mean	Mean	(p-value)
Single	0.169	0	1	0.080	0.193	-0.113
	(0.376)			(0.277)	(0.398)	(0.202)
Married	0.416	0	1	0.480	0.421	0.059
	(0.496)			(0.510)	(0.498)	(0.626)
Other	0.416	0	1	0.440	0.386	0.054
	(0.496)			(0.507)	(0.491)	(0.651)
DA	0.146	0	1	0.360	0.070	0.290
	(0.355)			(0.490)	(0.258)	(0.001)
Years at DA's Office	14.640	1	50	16.583	14.544	2.039
	(9.070)			(7.120)	(9.763)	(0.358)
Months in Same Unit	69.645	7	240	68.111	70.547	-2.436
	(52.006)			(56.338)	(52.598)	(0.868)
Education level						
Complete primary	0.012	0	1	0.040	0.000	0.040
	(0.110)			(0.200)	(0.000)	(0.147)
Complete secondary	0.193	0	1	0.120	0.245	-0.125
	(0.397)			(0.332)	(0.434)	(0.206)
Incomplete undergraduate	0.313	0	1	0.280	0.283	-0.003
	(0.467)			(0.458)	(0.455)	(0.978)
Comeplete undergraduate	0.229	0	1	0.200	0.264	-0.064
	(0.423)			(0.408)	(0.445)	(0.544)
Incomplete graduate	0.253	0	1	0.360	0.208	0.152
	(0.437)			(0.490)	(0.409)	(0.154)
Number of Children	1.607	0	5	1.640	1.648	-0.008
	(1.172)			(0.995)	(1.261)	(0.977)
Received Training	0.690	0	1	0.680	0.702	-0.022
	(0.465)			(0.476)	(0.462)	(0.846)

Notes: The first group of columns shows the mean, the standard deviation in parentheses, and the minimum and maximum values for the full sample. The second and third group of columns show the mean and standard deviation for each experimental group. The last column presents the balance between control and treatment with the corresponding p-value in parentheses.