Impact Evaluation Report:
A new homicide investigation policy in Bogotá, Colombia[*]

Daniela Collazos[†]   Leopoldo Fergusson[‡]   Miguel La Rota[§]   Daniel Mejía[¶]

Daniel Ortega[‖]

November 30, 2017

Abstract

This document evaluates a new homicide investigation procedure aimed at increasing coordination, accountability, and leadership among teams of functionaries involved. Results from a randomized controlled trial suggest that the new procedure increased the number of actions that are expected in homicide investigation, as well as additional forensic tests and procedures that can improve the quality of the investigation. The magnitude of the effects are economically meaningful, with estimated increases that range from 12% to over 20% of a standard deviation of the control group. Attempts to understand the underlying mechanisms of these results lead to some (more tentative) conclusions suggesting that the treatment produces well-coordinated teams that can communicate directly more fluently. Finally, evidence from surveys of functionaries suggest (again tentatively, since there is no random assignment to teams and there is differential attrition) that work motivation and the extent to which they have feedback on their performance, the pertinence and effectiveness of team member's roles, and most robustly the perceived quality and coordination of the teamwork, are all higher for functionaries working under the new scheme.

Keywords: Crime, Homicides, Team work.
JEL: K42, C93, K14.

---

1   Introduction

Homicide investigation studies focus on identifying and analyzing best practices to solve homicide cases and increase the probability that the offender is prosecuted for the crime. The empirical academic research on this topic is scarce, and to the best of our knowledge there is no rigorous evaluation of the effectiveness of alternative investigation strategies. We contribute filling this gap with a randomized controlled trial evaluating a change in the homicide investigation process in Bogotá, Colombia. The main objective of the new policy is to improve the investigation process and its judicial outcomes.

The context is ideal bearing in mind the poor performance of homicide investigation, evident in very low indictment rates (i.e., the percentage of criminal cases where a person is charged with committing the crime). In 2015, Colombia's homicide rate was 24 per 100,000 people, its indictment rate 21.5%, and its conviction rate (percentage of cases with a conviction, out of those with indictment) was 62%. For Bogotá, the homicide rate was 18 per 100,000 people, its indictment rate 22.4%, and its conviction rate 74.6%. That implies a mere 17% of homicide cases produce a final conviction, even in Bogotá which exhibits better performance than the average.

Another measure used in the literature is the homicide clearance rate, or the percentage of total cases in a year that are solved in that same time period. According to the FBI's uniform crime reports (UCR) 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 by arrest or exceptional means in the United States. When contrasting with the Colombian figures of mere one out of five simply indicted, it is clear that an understanding of how to improve judicial efficiency is key.

Brookman and Innes (2013) identify four definitions of investigative success, of which we highlight two: outcome success, which concerns the identification, prosecution and conviction of the suspects, and procedural success, which is about maintaining the integrity and quality of the investigation. These definitions of success capture best the main objectives of the intervention: improving the initial investigation process to identify the perpetrator and obtaining sufficient evidence to establish beyond reasonable doubt his or her culpability.

Many factors can affect the results of a homicide investigation process. The final outcome is sensitive to the circumstances of the crime, but also to the capacity of the homicide investigation unit and the judicial system. It is therefore possible to identify problems and best practices in order to improve the effectiveness of the investigations. For example, the lack

of resources, a poor working relationship between District Attorneys (DAs) and investigators, and poor procedures for processing and analyzing evidence can negatively affect the murder clearance rate (Cronin, Murphy, Spahr, Toliver, & Weger, 2007). Instead, formal training of homicide detectives and the use of sophisticated analytical tools improve the solvability of homicides (Keel, Jarvis, & Muirhead, 2009). Also, a cooperative relationship between DAs and detectives is related with better clearance rates and with successful prosecutions.

La Rota and Bernal Uribe (2014) conducted a diagnosis of the homicide investigation process in Colombia and identified, among others, two main problems: (1) there are multiple officers acting at different stages under a diffuse leadership and (2) there are delays generated by successive reallocations between different prosecutors. This breaks the link between the different stages of the criminal procedure. For the authors, this disruption is caused by the lack of leadership taking responsibility for the case from the beginning.

Based on this analysis, the Attorney's office concluded that the current homicide investigation structure needed to be restructured into a more efficient one, avoiding the loss of information between public servants and more effectively using time, human and physical resources. The Office for Public Policy Affairs of the Attorney General developed a new structure for the murder investigation process to overcome these problems. This is the policy that we experimentally evaluate, by (randomly) assigning a fraction of homicide cases to follow a new investigative procedure.

The new procedure seeks to overcome the disconnection between the initial investigation, the actions occurring before the trial, and the trial. The main change is structuring teams that investigate murder cases together, from the initial assignment until the arraignment. Each case is led by a DA and is investigated by a team of technical investigators or CSIs (consisting of three to four people) and two detectives. One DA, one CSI team and two detectives conform a "homicide squad", working continuously as a team. The intervention was piloted for two weeks in December 2015, officially started on January 20, 2016, and lasted until December 4 of 2016.

The introduction of these homicide squads changes the existing process of homicide investigation in four main dimensions. First, from the moment the CSIs and the detectives are dispatched to the crime scene they know the DA in charge of the investigation. This allows them to communicate, receive direct orders, and clarify possible questions. Second, the same DA is responsible from the initial urgent acts (*actos urgentes*) until the arraignment. Third, the DA can meet with his squad to discuss the research strategy and the criminal hypotheses. Fourth, after the arraignment the case is transferred to a new DA who continues with the

criminal procedure, so that the DA who investigates the case focuses only on this task and does not spend too much time in court hearings.

In short, the new procedure seeks to improve, from early on, the working relationship between DAs, technical investigators and detectives, fostering both a cooperative relationship between them and creating a sense of responsibility for each case they handle.

This report proceeds as follows. In section 2 we describe the homicide investigation process in Colombia. Section 3 explains the intervention protocol to randomly allocate cases to treatment and control. Sections 4 and 5 describe our data sources and empirical framework for estimating the effects of the homicide squad scheme on relevant outcome variables. Next, Section 6 presents our main quantitative results. Finally, in Section 7 we discuss the preliminary conclusions that can be derived from the findings.

2   Background: the homicide investigation process

In Colombia, two national divisions are responsible of homicide investigation: The Sectional Division of Crime Investigation of the National Police (SIJIN, for its Spanish acronym[1]) and the Technical Division of Crime Investigation of the Attorney General Office (CTI). The homicide investigation policy we evaluate was developed and implemented only within the latter. The Automatic Dispatch Center of the National Police (CAD) responds to an initial complaint–by civilians or police authorities at the crime scene– and informs a Satellite Unit (*Unidad Satélite*). The unit then assigns the case to either the SIJIN or the CTI. For CTI cases, when the initial complaint is made, a DA on duty closest to the Immediate Reaction Unit (URI) is also assigned to the investigation.

The criminal procedure for homicides in Bogotá has three major stages. The first stage is the initial investigation (*Indagación*) of the crime, in which physical evidence is collected in order to identify and arrest the offender. The second is the pre-trial stage (*Investigación*). It encompasses the indictment (*Imputación*) (i.e. the formal accusation against a person for the criminal offense, done in Colombia only before a Judge) and the gathering, if necessary, of corroborating evidence that proves beyond reasonable doubt the culpability of the suspect. The final stage is the trial, starting with the arraignment hearing (*Audiencia de formulación de acusación*) in which the accused pleads guilty or not guilty and the evidence is outlined to determine if it is sufficient to proceed to trial.

---

[1]From hereon we refer to the different units involved in the investigation using their Spanish acronyms.

2.1 Initial investigation (*Indagación*)

Once the CTI is assigned to investigate the case, it dispatches a team of crime scene investigators (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). At the INML the body is processed for external evidence and a forensic pathologist performs the autopsy. The detectives interview witnesses and follow leads, and execute any orders of the URI's DA leading the investigation. After documenting the evidence, they write detailed reports of their activities at the crime scene: the CSI leader writes the record of technical examination of the corpse (*Acta de inspección técnica a cadáver*) and the leading detective drafts the Executive Report (*Informe ejecutivo*).

The URI's DA decides–based on these reports–if further enquiry is needed, and if so, instructs detectives to keep investigating. Then, he or she (or other DA within the URI) prepares the methodological program (*Programa metodológico*), summarizing the facts and proposing the criminal hypotheses, and sends it to a DA of the Life Unit (*Unidad de Vida*), to whom the case is transferred. However, at this point further transfers may occur to other DAs within the Life Unit. Life Unit DAs decide whether to continue or not with the criminal procedure–in the event a suspect was identified during the initial investigation– thus initiating the pre-trial stage, or to classify the investigation as an unsolved case if there are no active leads that detectives can pursue.

2.2 Pre-trial (*Investigación*)

The pre-trial stage begins when a suspect is formally charged. When the offender is arrested, the DA must bring the detainee before a judge of guarantees (*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 the indictment occur simultaneously. Colombia's Penal Code defines *in flagrante* as the situation when a person is (i) caught and apprehended during the commission of the crime, or (ii) the person is caught or identified during the commission of the crime and apprehended immediately after by hot pursuit or calls for help by witnesses, or (iii) when the person is caught and captured with objects indicating probable culpability of having committed the crime. The DA has 30 days after the indictment to continue with the next step of the criminal procedure, the arraignment, or to ask the judge for a dismissal.

During this stage, detectives can and should continue searching for evidence to determine

beyond doubt whether the suspect committed the crime.

2.3 Trial (*Juicio*)

The arraignment (*Formulación de acusación*) is the first step of this stage. It is a procedure whereby the accused is brought before the judge to plea to the criminal charge in the indictment. The indictment bill (*Escrito de acusación*) is read to the defendant so he or she understands what the charges are, what his or her constitutional rights are, and then he or she is asked by the judge to plead guilty or not guilty. When pleading guilty, the case proceeds to sentencing. Otherwise, the judge holds a preparatory hearing (*Audiencia preparatoria*) to evaluate if there is enough evidence that the defendant committed the crime and if the case should continue to trial. In Colombia, trials are conducted before a judge without a jury.

3   Intervention protocol

In our intervention, a (random) fraction of homicide cases under the CTI's responsibility follow a new investigative procedure. This new procedure seeks to overcome the disconnection between the initial investigation, the pre-trial, and the trial. The main change is structuring teams that investigate murder cases together, from the initial assignment or urgent acts until the arraignment. Each case is led by the DA and is investigated by a CSI team (consisting of three to four people) and two detectives. From now on, we will thus use "homicide squad" or "group" to refer to one such team composed by one DA, one CSI team and two detectives. Each unit is simultaneously responsible for several cases.

The introduction of homicide squads changes the existing process of homicide investigation in the following main dimensions.

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

Second, the same DA is responsible from the urgent acts until the arraignment. This eliminates the URI's involvement and possible transfer to a different DA. The only exceptions are when the case is transferred to a DA specialized on homicides with specific characteristics: juvenile offenders, or when the forensic pathologist report is required to determine if the victim died for causes other than natural, or when the case jurisdiction is outside Bogotá or

of a different judicial branch (for example, military justice). When the forensic pathologist report is required, the case is transferred to the 112 DA of the Life Unit who waits for the pathologist report and, if it is indeed a murder, returns the case to the initial DA for further investigation. Otherwise, the 112 DA closes the case since no crime was committed. Even though there is a specialized DA for femicide cases operating under the previous scheme, the treatment group also investigates these cases if they occurred during a treatment squad shift.[2]

Third, the DA can meet with his squad to discuss the research strategy and the criminal hypotheses in order to write the methodological program.

Fourth, after the arraignment the case is transferred to a new DA of the Life Unit, who continues with the criminal procedure. This is done so that the DA who investigates the case focuses only on this task and does not spend too much time in court hearings. The homicide squad thus works on the case up to the moment when the bill of indictment is filed. Afterwards a new DA is assigned in order to bring the case to trial.

4   Data and variables

We rely on administrative data on homicides reported to the Office of the Attorney General of Bogotá. We use actions occurring before December 31st of 2016. The reason is that, though treated and control cases continue their course through 2017, the experimental assignment of new cases and the team divisions was already finalized. Thus, we can no longer so confidently attribute impacts on treatment cases as stemming merely from the intervention. This also implies that we unfortunately cannot look at outcomes at the trial stage or later, since these processes take too long to observe any palpable outcomes within this time frame in which it was feasible to maintain the experimental division of investigation procedures. There are extremely few observations in our entire sample (specifically, 27 out of 1,683, or 1.6%) at the trial stage by the time we had to interrupt the experimental protocol in December of 2016.

4.1   Administrative data and variables

Our main outcomes are actions and decisions taken by the CSIs, detectives, and the DA. 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

---

[2]We conduct the analysis including and excluding femicide cases since these may behave differently.

the crime scene, collecting fingerprints or biological fluids, and interviewing witnesses). The DA's actions are of three types. First, instructing detectives to perform additional investigative actions. Second, making decisions -like filing charges against a person or arraigning him or her- that define the stage the case is in. Third, attending hearings and presenting petitions to the judge, including orders to interview specific witnesses or asking to deny bail to a suspect.[3]

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 homicide investigation process. The set of possible actions or outcomes in a case is extensive and it depends on the characteristics of the homicide.

All actions of cases prosecuted by the Office of the Attorney General must (in theory) be reported by the detectives and the DA in the SPOA (acronym for *Sistema Penal Oral Acusatorio* or Oral Accusatory Penal System, also the name of Colombia's criminal justice system). This is our main data source. It is a mandatory reporting and follow-up automated system, yet officials do not always promptly comply with it. Also, 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 represents some challenges for the evaluation, which entices 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 learned by the Attorney General, and we focus on the section of felonies against persons such as assault, manslaughter, murder, etc.

Another source of information is the Integrated System of Management of the Judicial Police-SIG (acronym for *Sistema Integrado de Gestión de la Policía Judicial*, or Integrated Management System of the Judicial Police), which is used specifically to report the orders issued by the DA to the detectives. This data source has more complete information than the SPOA on the investigation actions.

Finally, we also have access to the reports written by the CSI teams, the detectives and the district attorneys for each case of the experiment. These are the "corpse examination report", the "executive report" and the "methodological program" referred to on the background section. With text analysis on these data, we do additional exercises to explore the nature of the investigative work performed by the teams. Specifically, we identify and count keywords

---

[3]The judges respond to DA petitions and decide whether their actions are legal or not, and assess the evidence and legal arguments presented during the trial to decide on the culpability of the accused. As noted, however, we are unable to examine impacts on judges actions.

and activities that can signal a better investigative process.

There are three challenges in correctly measuring each outcome and estimating 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, data from the SPOA has measurement error since not every action is reported. Third, the same action may be in the system under different labels chosen by officials reporting them.

To deal with 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 under a single label called "search and analysis of documents and databases". During this process we drop actions that do not constitute potential meaningful changes in the quality of the process, for instance merely administrative tasks not likely to be changed by the treatment or to significantly change the course of the investigation. We also drop actions for which we lack a clear hypothesis on whether they should be affected by the intervention, for example, 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 (SD). 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 effects sizes relative to the standard deviation of the control group.

We create the indices only for the first stage of the investigation process because this is where the largest number of actions are undertaken.[4] For the next two stages we focus on the effect of the intervention on individual outcomes. We also complement the actions reported in the SPOA to those reported in the SIG to detect inconsistencies.

After classifying actions and creating summary indices we still have multiple outcomes.

---

[4]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 treatment and control groups and before doing any estimation of the treatment effect. Doing this limits us to work with the part of the data with less measurement error, which reduces the bias but does not solve the problem completely.

Thus, we adjust the p-values to account for multiple inference. We follow Romano and Wolf (2005, 2016) resampling-based stepdown multiple testing method to control the family wise error rate (FWER)-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, effectively invalidating inference (Collazos, Fergusson, La Rota, Mejía, & Ortega, 2017). We highlight the (few) occasions where we had to depart from a choice as planned in the pre-analysis, mainly as a result of practical obstacles with the data or implementation. Appendix Table A-1 describes the list of the outcome variables studied, stage by stage. Outcomes related to the actions performed at the crime scene and to the indictment are of greatest interest to this experiment, thus we will also study them individually, verifying their non-adjusted p-values. The list of outcomes on which we focus on non-adjusted p-values, as stated in our pre-analysis plan, is also listed in Table A-1.

4.2   Survey data and variables

A baseline survey was conducted to the CSI teams, the detectives and the district attorneys to help evaluate the mechanisms that may explain the direct effect of the intervention. This survey explores their roles, their motivations and job satisfaction, the quality of their work and the importance of teamwork. A follow-up was done at the end of the intervention. We also create four standardized indices to aggregate several outcomes, and also evaluate some outcomes individually. The indices cover the following areas:

1. Motivation and feedback;

2. Team members roles, effectiveness, and quality;

3. Teamwork;

4. Coordination and progress.

In each case, several questions (with answers on a scale from 1 to 10) are collected in the index by topic. Thus, 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 Office of the District Attorney to do their job, how motivated and satisfied they are with the work they carry out, and how

9

responsible they feel for the successes and failures achieved in solving a homicide. Each question used in each index is listed in Appendix Table A-1.

We look at the following results individually:

1. Information. This is captured with 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". The respondent answers one of the following: strongly disagree, disagree, neither agree nor disagree, agree or strongly agree.

2. Overall efficiency. This is captured with 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". The respondent answers one of the following: strongly disagree, disagree, neither agree nor disagree, agree or strongly agree.

3. Victims. Captured with the percentage of the 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 a single murder or more, we treat the entire case as one and we focus on its results. Even though the intervention was designed to treat only murder cases, it is actually the investigation process that determines the felony that the defendant is charged with. This means that all deaths suspected to be due to causes other than natural ones –excluding car accidents– and that are investigated by the CTI are eligible to participate in the experiment. These include suicide and abortion.

In order to randomly allocate homicide cases between treatment and control, the Attorney General's office had to place strict rotation rules to the investigative teams on call at each shift. The first step was to form fixed investigative units (9 lab experts, 5 detectives and a DA) that would cover the metropolitan area of Bogotá. There were 6 groups in total and 4 of them were assigned to 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 6 groups is assigned to cover all incidences in each shift. The groups' rotation throughout the month implies that every group covers every one of the 14 possible weekly shifts at

least once every 6 weeks (2.3 shifts per week). Since the experiment lasted 42 weeks, each shift was covered by every team approximately 7 times over the course of the experiment. Controls cover each shift at twice the rate since there were 2 control groups (for whom there was no control over the integrity of the composition of investigative teams). Our Intention to Treat (ITT) estimates therefore 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 investigation stage. Consider a case c for which a specific procedure y was done. We estimate the effects via an Ordinary Least Square (OLS) regression:

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

where: $\text{Team}_c = 1$ if the case is in the treatment group, that is, if it is assigned to an homicide squad working as a team, 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; δ_{month} and $\gamma_{weekend}$ are month and weekend fix effects; and ε_c is the error term. In the baseline, we use robust standard errors to general heteroscedasticity. Bearing in mind that there are different combinations of lab experts, detectives and DAs in each investigative unit that can be assigned to investigate a case during the urgent acts, and that within these teams unobservables may be correlated, for robustness, we also estimate versions of this model where we assume the standard errors are clustered both at the team and at the shift level.

Because actions depend on legal timelines and administrative delays, estimates could be contaminated by these differences in time. To focus on the differences more likely caused by the treatment as a whole, we also estimate the treatment effect on a subset of actions after T days have passed since the case was received by the District Attorney's Office. We define T equal to the 90th percentile of the number of days that pass before action y is observed, and verify robustness for other thresholds.[5]

To increase the precision of the experiment, as robustness we include a number of controls, described in Appendix Table A-2. 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 division within the Office of the Attorney General or to another institution), if the case started as an assault, if the case is initially

---

[5]In the pre-analysis plan we had suggested 75th percentile as a baseline criteria, but in practice this was too lenient a threshold, making the check not that different to the baseline estimation.

assigned to the *Secretaría de Salud* which certifies natural deaths, but it is sent back to be studied by the judicial police[6], if the case occurred in a weekend, if the case occurred at night, and if the suspect was arrested in flagrante.[7]

Table 1 presents descriptive statistics for the control variables, and examines the balance between treatment and control. In general, the differences between treatment and control 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 as 1 for women, 2 for male, and 3 for 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, namely gender, age, location type and location in the city. The figure reveals that not only the averages for treatment are controls are similar, but the distributions of the variables look remarkably similar. All of this suggests that 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 has an effect on some of the outcome variables. We apply the randomization schedule to the data of 2015 in order to get the distribution of the sample between treatment and control groups. From a total of 1667 cases, 568 would have been randomly assigned to the control status and 1099 to treatment. We then create summary indices in order to test the effect of the randomization mechanism. Results indicated that 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

---

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

[7]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. In particular, we dropped: number of victims per case (only 28 cases, or 1.66% had more than 2 victims); dummy variables that indicate if the case is femicide or abortion (only 11 cases, or 0.65%, and 13 cases, or 0.77%, fall in each category respectively); and a dummy variable for whether the victim is unidentified because it is severely mismeasured in the data. In addition to the limited variation in some of these variables, at least those concerning the type of crime are classified by DAs, so they could depend on the treatment and be "bad controls" (Angrist & Pischke, 2008). We also include a dummy variable that indicates if the victim was a newborn, instead of the abortion originally planned. The age and gender of the victim, indicators for the location (type and city area) of the crime, and the dummy for whether the case occurred at night, were also not in the pre-analysis plan, but are arguably exogenous characteristics instead measured with better quality in the data. This motivated their inclusion though, as will be clear below, results are not really sensitive to the inclusion of controls. Thus, none of these departures from the plan are meaningful for the conclusions.

in absolute value and often smaller.[8]

6   Results

6.1   Administrative data on investigation outcomes

We start with the analysis of some of the outcomes from administrative data, from the SPOA and SIG systems. Table 2 describes the key variables. The minimum actions, in both datasets, capture a set of activities that are *always* expected to occur. Therefore, we anticipated treatment no to have any effect on them. However, descriptive statistics suggest an increase in the treatment group relative to control (both in the SPOA and SIG data). The forensic actions, comprising additional investigation actions that can be performed on a case and on which we expected a positive impact in the pre-analysis, also appear higher in the treatment group.

Other noteworthy difference in the descriptive statistics is in the days to indictment, which appears much larger on the treatment group. We expected that this time since the initial complaint to indictment would decrease given the elimination of the reassignments between DAs, as well as the steps involving the URI, which should shorten the time of prosecution. Interestingly, however, the descriptive statistics also show that this does not translate to more days to the *bill* of indictment, which in fact are smaller in treatment.

The indictment and bill of indictment rates are indicators usually used by the Office of the General Attorney to measure performance. All reported cases are divided into those in which a crime was really committed, and the non-effective entries, which are cases not involving a felony and therefore archived. The indictment rate is then typically defined as the number of cases with indictment divided by the total number of confirmed homicide cases. In turn, the bill of indictment rate is the ratio of cases with a bill of indictment to the cases with an indictment. To capture this in our regressions at the case level, we therefore first restrict the sample to the effective entries (1,113), and then examine regressions for a dummy variable for indictment, effectively measuring the determinants of the indictment rate. Similarly, we restrict the sample to cases with indictment (116) and examine regressions for a binary indicator for bill of indictment, thus capturing the bill of indictment rate. Our predictions on the effects on these rates (and indictments themselves) are ambiguous. On the one hand, they

---

[8]Given the long delays to conviction, one exception is the percentage of cases with conviction, where the size is large and imprecise and we have a small sample. As noted, we do not look at conviction rates as an outcome in the evaluation.

could increase because of a better initial investigation process and because the intervention eliminates unnecessary transfers between units, which delays the process. But they could decrease because the homicide squad spends more time doing a better job processing the crime scene, following leads, and collecting physical evidence allowing them to identify a suspect and indicting only with more solid grounds to ensure a conviction. This rationale could explain the conflicting evidence on days to indictment and to bill of indictment referred to above: the treatment group might take longer to produce the indictment when advancing in the investigation (taking more time to do both the higher number of minimal and forensic actions that they appear to do), but is then well prepared to quickly move to the formal bill of indictment.

The number of cases to be established (*casos por establecer*) sent to, and returned by, the specialized 112 DA require some explanation. Recall from the context section that these are cases requiring confirmation from the forensic pathologist's report. If the report confirms the murder, the case is returned to its original DA. Otherwise, the homicide case is closed. We expected a decrease as a result of the treatment, since detectives should pay more attention to the crime scene details and to the evidence, thus determining more precisely than the control group if a given case is a murder or not. The raw descriptive statistics do reflect this though the differences are small. The number of unsolved cases (*archivo por imposibilidad de establecer sujeto activo o pasivo*), are cases without active leads allowing detectives to identify the perpetrator or the victim. We expected this to decrease in the treatment with a better investigative work.

Turning to variables on the SIG database, Panel B of Table 2 presents descriptive statistics[9]. The first two indices are constructed as with the SPOA, and serve to verify the results with an independent data source. In addition, we examine three key variables that are part of the investigation process and are better captured in the SIG: conducting interviews (number of orders sent, where a separate order must be placed for each interview), locating "persons of interest" (typically witnesses or suspects), and orders to verify the identity of suspects. These are typically orders placed in the context of analyzing the crime and crime scene, so in line with our pre-analysis plan we look at individual in addition to Romano-Wolf corrected p-values.

In addition to these variables, we also use the SIG data to look deeper into the times of

---

[9]We have fewer observations in this case simply because some cases of the SPOA system do not appear in the SIG data, and because some cases having actions in the SPOA might not have actions in the SIG which focuses only on orders placed, and registered, by DAs

investigation in the treatment group (which, at least as judged by days to indictment, seems to increase as measured in the SPOA data). In particular, each order placed by the DAs contains several actions that investigators must conduct, and each of these have some time frame to be completed. Extensions, however, can be requested. We therefore measure two variables: the number of extensions requested on the assignment and the average days of extension requested. Also, we can measure the final days of delay to fulfill the assignment, that is, how much longer than the deadline the investigators are taking. The descriptive statistics reveal that both treatment and control have delays on average, but this appear to be shorter for the treatment group.

To evaluate how robust and significant these overall patterns are, we now evaluate the impact of the treatment estimating equation (1) on standardized outcome variables, including the aggregated indices, and correcting inference for multiple hypotheses testing. Table 3 presents the results with the SPOA outcomes. As noted, we evaluate p-values for the difference between treatment and controls with three different assumptions: robust standard errors in columns 2 and 3, clustered at the level of team or "homicide squad" (DA, investigator and CSI) in columns 4 and 5, and clustered at the shift level in columns 6 and 7. These p-values are presented for individual tests and adjusted for multiple hypotheses testing following Romano and Wolf (2005, 2016). Even columns show the corresponding p-value and, in parenthesis, the standard error of the effect under each assumption on errors. Odd columns show the p-value with the Romano-Wolf correction. As also stated in the pre-analysis plan, actions at the criminal scene and related to indictment are deemed central and evaluated in absence of this correction, but we show it for completeness. The six outcomes to be evaluated individually in the SPOA data are: the two indices for minimum actions, the indictment, bill of indictment, and the days it takes to get to the latter two. We thus present these in the top of the table.

This table reveals some interesting effects of the intervention. First, the treatment improves both the minimum actions and the forensic actions index. The magnitude of the effects are economically meaningful, with an increase that equals about 12% of a standard deviation of the control group, in both cases. The estimates are also precise regardless of how we deal with heteroscedasticity, with very small p-values (not so much with the Romano-Wolf correction, though remember that in these cases we focus on the uncorrected measures). In short, there is a significant and sizable increase in the measurable compliance with minimal actions that must be undertaken in every homicide case, and in forensic actions, as a result of the treatment. It also appears to be the case that the treatment group takes longer to get

to the indictment, which could signal a bigger effort to produce a better case, though we do not have conclusive evidence for this interpretation, in particular because once we include control variables this effect disappears (we report these results, with the same format as in Table 3, in Appendix Table A-5).

For the remaining variables, differences are not statistically significant at conventional levels, except for the cases to be established send to the 112 DA. The treatment group sent less of these cases to the DA, in line with expectations.

An important robustness test comes from recognizing that some actions take time to be conducted. Thus, as noted above, we repeat the estimation focusing on a sample of cases occurring sufficiently long before for each action to be typically observed. In practice, we implement this by keeping only cases occurring sufficiently early that the time elapsed from date of events to action exceeds the 90th percentile in this distribution of days from fact to action. For indictment, this implies keeping cases occurring more than 59 days before our records, and for the indictment bill, 75 days. The results are presented in Table 4. This table has three columns with controls and three without, and for simplicity we do not show different types of standard errors and focus only on the basic robust errors adjusted for general heteroscedasticity.[10] Allowing for controls, the only robust finding is a higher bill of indictment rate for the treatment group.

Other robustness exercises are relegated to the Appendix. In particular, we look only at homicide cases (*homicidio doloso*), which excludes femicide, abortion and any other criminal offense that was assigned (possibly in error) to the teams working in the experiment. The motivation is that our focus is homicide strictly, and some crimes like femicide may be investigated differently. Also, we repeat the analysis excluding a few cases that were found to be connected to broader criminal cases. For example, the offender has several cases and will be tried for the entire set, or there is a broader investigation looking into a pattern of crimes conducted by several people. These cases are typically transferred to DAs or special units in charge who may be outside the treatment or the experimental group thus changing the implementation protocol. The DA in the experiment could also get these other cases transferred to him, again changing the experimental protocol. In these cases, the investigation team can also change. These are only 17 cases, and are evenly divided between treatment and control. Finally, we also change the clusters to be based, not on the full "homicide squad", but on the combination of DA and CSI groups. This implies putting

---

[10] Also, there are fewer observations in these specifications, which caused some problems when estimating some of the clustered versions

together in the same cluster three groups of CSIs, serving in the same shift, in the same cluster. The motivation is that clusters by team often have very few observations, implying that clustering on them is a modest correction. Except some increase in standard errors with this latter modification, results remain largely similar and do not change the overall patterns of results under any of these robustness tests.

The results in Table 5 use the SIG data to complement and confirm some of the findings in Table 3 for the SPOA data. The minimum and forensic actions indices indicate a positive, quantitatively even larger (of 27% and 22% of a standard deviation, respectively) and very precise effects of the treatment. The relevant orders are also positive for the treatment group, but the coefficients are small and not significant. Interestingly, the treatment group requests many less and less lengthy extensions (decreases in about 15% of a standard deviation), even though they work longer to fulfill these assignments. How can this be? One interpretation consistent with these results 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 would also be consistent with the one trying to reconcile longer days to indictment, without longer days to the bill of indictment. Results including controls are quite similar, and they are presented in Appendix Table A-9.

6.1.1 Text analysis of written reports

To further try to understand the different implications of the team work imposed by the intervention, we now examine the written reports of the CSI teams, the detectives and the district attorneys. We look at the corpse examination report, the methodological program, and the executive report. The descriptive statistics for all variables in the text analysis are in Table 6. Like every other outcome, these outcome variables are described in Table A-1.

The corpse examination report has a detailed description of the criminal investigation work at the crime scene. The first three variables we look at are simply measures of 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 mechanically appear very often and convey little meaning, implementing two methods. The first method counts exactly how many words are left in each text after removing the "stop words". The second method also removes words that are very rare in the entire corpus of text. We also look at the proportion of key words organized by topics that are important in the process of investigation, namely: first respondent, interviews, DAs or investigators,

evidence, laboratories.[11] The rationale for including first respondent is that much of the success in the investigation can have to do with the quality of the work performed at the crime scene by whoever first arrives. The rest of the topics capture key inputs in the process of investigation.

Several messages emerge from Table 7, which presents the results for these outcomes. First, the length of the text is shorter in the treatment group, and sizable so, close to one half of a standard deviation smaller than the control group. On the frequency of topics, there are mixed results. On the one hand there is more discussion of the first respondent (24% of a standard deviation effect) and more mentions of the DAs and investigators (a large 57% of a standard deviation impact). Both of these suggest paying attention to key inputs in the investigation, and the latter in particular is consistent with better information flow and coordination between the homicide squads, a prime objective of the intervention. However, one surprise is observing less mentions of evidence (30% of a standard deviation fall). The remaining topics have a substantially smaller difference between treatment and control, perhaps with the exception of laboratory, 10% of a standard deviation larger 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 comes out less statistically significant than the process narrative. A similar conclusion applies to the number of evidence material collected, also in line with the finding in the process narrative. A new variable, looking at the number of procedures requested to the National Institute of legal Medicine and Forensic Sciences (INML), also reveals less procedures in the treatment group, in this case with a larger effect of 34% of a standard deviation.[12]

Turning to the second document we analyze, the methodological program (descriptive statistics in Panel B of Table 6), unfortunately there are missing observations lopsided against the treatment group. Thus, findings in this case must be interpreted with some caution. Table 8 shows that there are about 9% less observations in the treatment group (column 1). In these documents, the investigators 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 a better work by the treatment group, when looking deeper this does not appear to be the underlying reason.

---

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

[12]Results with controls are relegated to Appendix Table A-10.

Indeed, the subsequent three rows analyze how similar the research and criminal hypotheses are to each other. We use three methods[13]: the *tokens distance* focuses on comparing the words in each text, considering them closer if they share more words; the *Damarau-Leveshtein (DL)* method estimates distance using the minimal number of changes needed to transform one text to turn it into the other, so fewer transformations means less distance; and the *Qgram* method is similar to the tokens 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. But rather than coordination, what this appears to reflect is the treatment group relying more on hypotheses "templates" that are copied and pasted for different cases, but are not really meaningful. This is what the last two rows show. In these rows, we measure similarity of the hypotheses in each case with those in *other cases*, finding more homogeneity again 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 reflecting these results.

Table 9 (and Appendix Table A-12, including controls) looks at the final written document we examine, the executive report, with descriptive statistics for the relevant variables in panel C of Table 6. The length of the text, as in the corpse examination documents, is shorter in the treatment group (first three rows, with effects close to 15% of a standard deviation). Next, we also look at the frequency of some key topics in the investigation process. In particular, we count keywords having to do with the following topics: witnesses, family member, 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 any other general activities that do not fall under any preestablished category, like visiting and inspecting the landmark cited by a witness after having verified its existence).[14] There are, in this case, no outstanding robust differences between the two groups, except perhaps with less of the last two types of tasks in the treatment group.

Taking stock of the findings for the written reports, text data is inherently noisy, and this creates difficulties in making conclusive statements based on it. However, one thing seems

---

[13]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`.

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

clear: the treatment group spends less words in these written reports. One could venture to hypothesize that this could be precisely because they are in well-coordinated teams, that can communicate directly more fluently, and need less of the paperwork and written footprint. That they seem to even hastily rely on templates in some cases would be consistent with such a view towards the reports. Also, when they do write they seem to make more mentions of team members, and some key aspect of the investigation input like the first respondent actions. But, at the same time, other key elements and processes appear less, including evidence materials and other tasks like forensic tests and others.

6.2   Survey data

We now examine our survey data, examining the four indices and three individual variables described in section 4.2. Descriptive statistics for these variables are in Table 10 (the differences in observations reflect that nonresponse is an option). To protect subjects, we do not have their individual identity, but we do have their treatment condition (recall randomization occurs at the case level, but functionaries may be either in homicide squads or operating as usual). There is, unfortunately, differential attrition between treatment and control (as revealed by the treatment dummy in each panel). In the baseline, the sample is divided. But in the endline, around 30% are functionaries in the control and 70% in the treatment. This implies that results from the survey have to be interpreted with caution, as there may be non-random selection in those functionaries willing to answer in the endline.

Turning to the variables measured in the surveys, notice that indices are computed after standardizing the components in each index (hence the mean zero). The agreement with not having enough information and often facing problems in support for the work are on a 1 to 5 scale, increasing in the level of agreement. Time spent with the victims is the subjective perception of the fraction of the time the functionary devotes to victims[15] Controls in the survey are listed in Table A-14, and include sex, age group, marriage status, experience, indicators for DAs, educational background, number of children, and data on training received. These controls show that functionaries in each group differ on observables, and these differences are often significant (as revealed by the last column of the table).

All indices are larger for functionaries in the treatment group. This must be taken with some caution because, as noted, randomization is at the case level and not at the team level (and also, we have therefore considerably fewer observations). However, this points to an

---

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

overall improvement in the work environment amongst functionaries in the treatment group. And interestingly, the largest effect, with close to one-half of a standard deviation size, is on coordination, which falls very much in line with the nature of the intervention. Since functionaries are not randomized and there are differences between them, regressions with controls are particularly important here, and are shown in Panel B of Table 11. Unfortunately we loose additional observations here given data availability, but at least the coordination effect remains robust. Also, functionaries in the treatment group have more positive views on the information flow between investigators, are more likely to dismiss efficiency problems, and declare spending more time with victims. All of these again suggest better coordination and teamwork, at the very least as perceived by DAs. Of these three variables, the largest and more significant estimates are the information and time with victims variables (sizes of 57% and 47% of a standard deviation, respectively) with the result on inefficiency having smaller size and substantially large p-values that suggest we cannot rule out null effects. These conclusions are also qualitatively similar once including controls, except that the Romano-Wolf corrections produce large p-values for all outcomes, except the coordination index, in those regressions.

These result rely on endline data, reflecting back on the work they did during the intervention. A different approach is to do a difference in difference exercise with the baseline and endline information. While as noted we do not have individual identifiers to do a proper panel, we do have treatment status in the baseline and endline. Therefore, we can look at a regression where the indices or three individual variables are used as outcome variables, and they are regressed on treatment status, a dummy variable, and their interaction. The coefficient of the interaction is the quantity of interest, capturing the change in these variables for treated functionaries relative to controls. And since we focus on the change, assumptions for identification are weaker, allowing for constant differences between the pool in functionaries in the two groups. The results are in Table 12, in Panel A without controls, and in Panel B including controls. The result falls very much in line with the approach looking only at the endline: all indices improve, functionaries spend more time with victims, and are less likely to agree on information and efficiency problems. Of these results, only those for the indices on motivation and feedback, role and effectiveness, and coordination, which also are larger in size, are statistically significant at conventional levels. When including controls, the size and significance for the first of these indices falls, but not for the remaining two.

In short, this section reveals tangible effects on job satisfaction, especially those dimensions involving coordination, for functionaries dealing with cases in the treatment group.

7   Final discussion

This document evaluates a new homicide investigation procedure aimed at increasing coordination, accountability, and leadership in teams of functionaries involved in the process. Results suggest that the new procedure increased the set of minimum investigation actions that are expected from functionaries in homicide cases. Also, that more forensic tests and procedures are conducted relative to the control. The magnitude of the effects are economically meaningful, with estimated increases that range from 12% to over 20% of a standard deviation of the control group, depending on the dataset used. The treatment group appears to take longer to produce an indictment, but this does not delay the issue of an indictment bill. The treatment group requests many less and less lengthy extensions (decreases in 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 expected deadlines. There is also some, though less robust, evidence that the treatment group exhibits higher indictment rates (this is clearest when focusing on cases that are sufficiently mature for an indictment to have typically occurred, thus attempting to remove confounding effects in time to indictment).

Attempts to understand the underlying mechanisms of these results lead to some (more tentative) conclusions, based on the written reports for the cases filed by the functionaries and surveys on their own perceptions about the work environment and conditions. The most robust finding is that treatment cases are described in more succinct terms in the treatment group. We venture to hypothesize that this is precisely because the treatment produces well-coordinated teams, that can communicate directly more fluently, and need less of the paperwork and written footprint. Two complementary findings consistent with this view is that they seem to hastily rely on templates in some cases, yet when they do write they seem to make more mentions of team members, and some key aspect of the investigation input like the first respondent actions. However, we propose this more as a hypothesis for further investigation than as a robust conclusion.[16]

Finally, evidence from surveys of functionaries suggest that in key dimensions like work

---

[16]Also because some key elements and processes in the investigation appear less in the treatment group reports, including evidence materials 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 suggest robustly that the treatment creates a more actively engaged investigation.

motivation and the extent to which they have feedback on their performance, the pertinence and effectiveness of team member's roles, the quality and coordination of the teamwork, are all higher for functionaries working under the new scheme. These functionaries are also more likely to disagree with there being important information or efficiency problems in the homicide investigations, and spend more time with the victims. These findings are quantitatively large, and robust when it comes to perceptions about the level of coordination and progress within functionaries, which falls in line with the nature of the intervention. The remaining findings are more sensitive to the inclusion of functionary characteristics as controls. They should also be interpreted with some caution because there is differential attrition between treatment and control from baseline to endline surveys, and because randomization is at the case level and not at the team level.

References

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.

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

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.

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)

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.

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.

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.

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.

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

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 xxxx xxxxx xxxxx 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

Variables	Full sample Mean	Min	Max	Control group Mean	Treat. group Mean	Mean Difference (p-value)
Suspect arrested in flagrante	0.071	0	1	0.080	0.067	0.014
	(0.257)			(0.272)	(0.250)	(0.304)
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	0.036	0	1	0.037	0.035	0.002
	(0.185)			(0.188)	(0.184)	(0.874)
Case ocurred during a weekend	0.364	0	1	0.373	0.359	0.014
	(0.481)			(0.484)	(0.480)	(0.572)
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)
Newborn	0.036	0	1	0.031	0.038	-0.006
	(0.185)			(0.175)	(0.191)	(0.501)
Lag in days between complaint and death	2.781	0	680	3.977	2.164	1.813
	(25.532)			(29.561)	(23.171)	(0.167)
Sex of the victim	1.836	1	3	1.862	1.823	0.040
	(0.447)			(0.423)	(0.459)	(0.085)
Age of the victim Group	3.663	1	7	3.623	3.684	-0.061
	(1.639)			(1.597)	(1.661)	(0.471)
Type of place of the diligence	2.494	1	4	2.449	2.517	-0.069
	(1.141)			(1.163)	(1.130)	(0.243)
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 parenthesis, for 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								
Minimun actions-Mean effects index (z-score)	0.078	1.108	-2.078	8.249	0	1	0.119	1.159
Forensic actions-Mean effects index (z-score)	0.076	1.184	-0.599	9.607	0	1	0.115	1.267
Indictment	0.069	0.253	0	1	0.070	0.255	0.068	0.253
Bill of indictment	0.059	0.235	0	1	0.054	0.226	0.061	0.240
Days to indictment	15.060	39.018	0	197	3.700	10.644	21.039	46.587
Days to bill of indictment	59.218	19.300	12	191	66.870	32.757	56.469	10.232
Indictment rate	0.104	0.306	0	1	0.101	0.302	0.106	0.308
Bill of indictment rate	0.853	0.355	0	1	0.775	0.423	0.895	0.309
Cases to be established sent to 112 DA	0.371	0.483	0	1	0.469	0.500	0.320	0.467
Cases to be established returned by 112 DA	0.006	0.077	0	1	0.007	0.083	0.005	0.073
Unsolved cases	0.058	0.233	0	1	0.045	0.208	0.064	0.245
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. Total sample in SPOA database is 1,683 cases, 573 in the control group and 1,110 in the treatment group. Total sample in 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)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Minimun actions-Mean effects index	0.119	0.029	0.237	0.094	0.536	0.035	0.247	1,683
		(0.054)		(0.071)		(0.056)		
Forensic actions-Mean effects index	0.115	0.043	0.293	0.065	0.450	0.031	0.247	1,683
		(0.056)		(0.062)		(0.053)		
Indictment	-0.005	0.918	0.974	0.917	0.973	0.924	0.972	1,683
		(0.051)		(0.050)		(0.055)		
Bill of indictment	0.032	0.547	0.959	0.529	0.958	0.572	0.963	1,683
		(0.052)		(0.050)		(0.056)		
Days to indictment	1.629	0.003	0.022	0.001	0.008	0.003	0.031	116
		(0.527)		(0.470)		(0.535)		
Days to bill of indictment	-0.318	0.134	0.531	0.135	0.536	0.135	0.578	87
		(0.210)		(0.210)		(0.210)		
Indictment rate	0.017	0.793	0.974	0.788	0.973	0.804	0.972	1,113
		(0.063)		(0.061)		(0.067)		
Bill of indictment rate	0.283	0.115	0.531	0.107	0.536	0.147	0.578	116
		(0.178)		(0.174)		(0.194)		
Cases to be established sent 112 DA	-0.300	0.000	0.000	0.000	0.000	0.000	0.000	1,683
		(0.050)		(0.060)		(0.056)		
Cases to be established returned by 112 DA	-0.019	0.702	0.974	0.701	0.973	0.701	0.972	1,683
		(0.049)		(0.049)		(0.049)		
Unsolved cases	0.089	0.103	0.531	0.106	0.536	0.100	0.519	1,683
		(0.055)		(0.055)		(0.054)		

Notes: All variables are standardized by the mean and standard deviation of the control group. Even columns show the corresponding p-value and, in parenthesis, the standard error of the effect under each assumption on errors. Columns 2 and 3 have robust standard errors, columns 4 and 5 clustered at the level of team or "homicide squad" (DA, investigator and CSI), and columns 6 and 7 clustered at the shift level. Odd columns show the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016). Indictment rate regressions are for a dummy of indictment on treatment after having restricted the sample to effective cases only, and the bill of indictment regressions are for a dummy of bill of indictment after restricting the sample to cases with indictment only.

Table 4: Effects of the intervention on key outcomes: SPOA database variables
Restricting to sample of sufficiently early events

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	No controls			Controls			
	Treatment Effect	PValue	RW PValue	Treatment Effect	PValue	RW PValue	N
Indictment	-0.015	0.781	0.955	0.049	0.151	0.382	1,530
	(0.054)				(0.034)		
Bill of indictment	0.050	0.395	0.405	0.118	0.005	0.017	1,464
	(0.059)				(0.042)		
Days to indictment	1.790	0.003	0.008	0.134	0.789	0.783	105
	(0.582)				(0.500)		
Days to bill of indictment	-0.311	0.158	0.277	-0.357	0.211	0.192	85
	(0.219)				(0.282)		
Indictment rate	0.004	0.948	0.955	0.059	0.175	0.382	974
	(0.069)				(0.044)		
Bill of indictment rate	0.460	0.004	0.013	0.488	0.021	0.032	103
	(0.158)				(0.207)		

Notes: All variables are standardized by the mean and standard deviation of the control group. Columns 1 to 3 show the results when control variables are not included and columns 4 to 6 when included. Columns 2 and 5 show the corresponding p-value and, in parenthesis, the standard error of the effect under the assumption of simple standard errors. Columns 3 and 6 show the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016)

Table 5: Effects of the intervention on key outcomes: SIG database variables
(No controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Minimun actions SIG-Mean effects index	0.270	0.000	0.000	0.000	0.002	0.000	0.000	1,612
		(0.057)		(0.073)		(0.059)		
Forensic actions SIG-Mean effects index	0.220	0.000	0.002	0.000	0.002	0.001	0.004	1,612
		(0.060)		(0.061)		(0.063)		
Order to interview	0.039	0.459	0.509	0.480	0.493	0.483	0.508	1,612
		(0.053)		(0.056)		(0.056)		
Order to locate person of interest	0.066	0.228	0.509	0.251	0.493	0.253	0.508	1,612
		(0.055)		(0.058)		(0.058)		
Order to individualize suspect	0.071	0.207	0.509	0.203	0.493	0.212	0.508	1,612
		(0.056)		(0.056)		(0.057)		
Number of extensions of the assignment	-0.153	0.001	0.009	0.002	0.016	0.001	0.009	1,612
		(0.048)		(0.050)		(0.047)		
Average days of extension of the assignment	-0.147	0.002	0.011	0.004	0.023	0.003	0.014	1,612
		(0.048)		(0.051)		(0.049)		
Days of delay to fulfill the assignment	-0.104	0.026	0.097	0.029	0.111	0.045	0.174	1,612
		(0.047)		(0.048)		(0.052)		

Notes: All variables are standardized by the mean and standard deviation of the control group. Even columns show the corresponding p-value and, in parenthesis, the standard error of the effect under each assumption on errors. Columns 2 and 3 have robust standard errors, columns 4 and 5 clustered at the level of team or "homicide squad" (DA, investigator and CSI), and columns 6 and 7 clustered at the shift level. Odd columns show the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016)

Table 6: 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

Table 6 – 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 7: Effects of the intervention on key outcomes: text analysis on corpse examination report (No controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Description Length (Raw)	-0.486	0.000	0.000	0.000	0.000	0.000	0.000	1,512
		(0.047)		(0.053)		(0.045)		
Description Length (Method 1)	-0.482	0.000	0.000	0.000	0.000	0.000	0.000	1,512
		(0.047)		(0.054)		(0.045)		
Description Length (Method 2)	-0.476	0.000	0.000	0.000	0.000	0.000	0.000	1,512
		(0.047)		(0.055)		(0.045)		
Keywords by topic (in proportion to total length)								
First Respondent	0.224	0.000	0.000	0.001	0.008	0.000	0.002	1,512
		(0.060)		(0.070)		(0.060)		
Interviews	-0.035	0.535	0.540	0.570	0.554	0.548	0.549	1,512
		(0.056)		(0.061)		(0.058)		
DA or Detectives	0.637	0.000	0.000	0.000	0.000	0.000	0.000	1,512
		(0.098)		(0.147)		(0.101)		
Crime Scene Data	-0.091	0.073	0.145	0.119	0.285	0.097	0.181	1,512
		(0.051)		(0.059)		(0.055)		
Evidence (EMP)	-0.289	0.000	0.000	0.000	0.000	0.000	0.000	1,512
		(0.051)		(0.057)		(0.058)		
Lab	-0.117	0.027	0.074	0.107	0.285	0.026	0.076	1,512
		(0.053)		(0.073)		(0.053)		
Findings Description Length (Raw)	-0.145	0.008	0.039	0.020	0.082	0.011	0.042	1,432
		(0.054)		(0.062)		(0.057)		
Crime Scene Drawing	0.318	0.000	0.000	0.000	0.000	0.000	0.000	1,594
		(0.054)		(0.056)		(0.055)		
Number of Evidence	-0.115	0.009	0.039	0.009	0.052	0.008	0.041	1,643
		(0.044)		(0.044)		(0.043)		
INML Requests Length (Raw)	-0.325	0.000	0.000	0.000	0.000	0.000	0.000	1,487
		(0.051)		(0.060)		(0.057)		

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 methodological program
(No controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Has Methodological Program	-0.497	0.000	0.000	0.000	0.000	0.000	0.000	1,683
	(0.068)			(0.092)		(0.076)		
Number of Criminal Hypothesis	0.301	0.000	0.000	0.000	0.000	0.000	0.000	1,523
	(0.052)			(0.068)		(0.058)		
Number of Research Hypothesis	0.952	0.000	0.000	0.000	0.000	0.000	0.000	1,523
	(0.063)			(0.112)		(0.084)		
Total Number of Hypothesis	0.724	0.000	0.000	0.000	0.000	0.000	0.000	1,523
	(0.058)			(0.097)		(0.073)		
Distance between hypothesis by method								
Method 1 - Tokens	-0.860	0.000	0.000	0.000	0.000	0.000	0.000	515
	(0.123)			(0.124)		(0.135)		
Method 2 - DL	-0.276	0.027	0.030	0.028	0.026	0.036	0.034	515
	(0.124)			(0.125)		(0.131)		
Method 3 - Qgrams	-0.458	0.000	0.001	0.000	0.001	0.001	0.002	515
	(0.124)			(0.123)		(0.132)		
Distance with hypothesis from other cases by type								
Criminal Hypothesis	-1.332	0.000	0.000	0.000	0.000	0.000	0.000	1,076
	(0.068)			(0.098)		(0.077)		
Research Hypothesis	-2.287	0.000	0.000	0.000	0.000	0.000	0.000	515
	(0.134)			(0.164)		(0.155)		

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: Effects of the intervention on key outcomes: text analysis on executive report (No controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Narration Length (Raw)	-0.140	0.008	0.054	0.039	0.231	0.010	0.074	1,557
		(0.053)		(0.068)		(0.055)		
Narration Length (Method 1)	-0.152	0.004	0.036	0.026	0.210	0.005	0.049	1,557
		(0.052)		(0.068)		(0.055)		
Narration Length (Method 2)	-0.147	0.005	0.045	0.031	0.221	0.008	0.058	1,557
		(0.053)		(0.068)		(0.055)		
Keywords by topic (in proportion to total length)								
Witnesses	0.029	0.600	0.825	0.646	0.875	0.620	0.861	1,557
		(0.056)		(0.064)		(0.059)		
Family	-0.073	0.172	0.444	0.238	0.631	0.173	0.436	1,557
		(0.053)		(0.062)		(0.053)		
Video-Cams	0.028	0.627	0.825	0.671	0.875	0.633	0.861	1,557
		(0.057)		(0.066)		(0.058)		
Interviews	0.104	0.073	0.261	0.215	0.631	0.079	0.288	1,557
		(0.058)		(0.084)		(0.059)		
Verification	-0.115	0.035	0.164	0.098	0.409	0.036	0.165	1,557
		(0.055)		(0.069)		(0.055)		
Inspection	0.135	0.012	0.067	0.069	0.348	0.013	0.082	1,557
		(0.054)		(0.074)		(0.054)		

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 10: 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 11: Effects of the intervention on key endline outcomes: functionaries' perceptions

	N	Treat. Effect	PValue	Robust Errors RW PValue

Panel A. No controls

Indices

	N	Treat. Effect	PValue	RW PValue
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

	N	Treat. Effect	PValue	RW PValue
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 12: 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

Outcome	Description	Hypothesis
Minimum actions Index (Non-adjusted p-value)	Mean effects index of set of actions that are usually done in all cases. • For the SPOA database, this set comprises the following actions: search and retrieval of material evidence, interviews, photography of the crime scene, examination to the corpse, documentation of the crime scene, documentation of a location other to the crime scene and verification of things. • For the SIG database, this set comprises the same actions except the verification of things.	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. • For the SPOA database, 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, and other investigative actions. • For the SIG database, this set comprises the same actions and includes an additional physical and chemical analysis component.	We expect an improvement. These actions should be sensitive to the detectives and the CSIs accountability to the DA, 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 DAs 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 DAs to the detectives to locate "persons of interest", typically witnesses or suspects	We expect an increase

Outcome	Description	Hypothesis
Order to individualize suspect (Non-adjusted p-value)	Number of orders issued by the DAs to the detectives to verify the identity of suspects	We expect an increase
Number of extensions of the assignment	Assignments are a 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 less days are 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, and returned by, the specialized 112 DA *(casos por establecer)*	Cases requiring confirmation from the forensic pathologist's report. If the report confirms the murder, the case is returned to its original DA. Otherwise, the homicide case is closed.	We expect a decrease. Detectives will pay more attention to the crime scene details and to the evidence, thus determining more precisely than the control group if it is a murder case or not.
Unsolved cases *(archivo por imposibilidad de establecer sujeto activo o pasivo)*	Cases without active leads allowing detectives to identify the perpetrator or the victim.	We expect a decrease. A 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.	Two scenarios possible: increase from better initial investigation process and less unnecessary transfers between units; decrease because the homicide squad spends more time doing a better job processing the crime scene, following leads, and collecting physical evidence to identify a suspect and indicting only with more solid grounds for conviction.

Outcome	Description	Hypothesis
Indictment rate *(tasa de imputación)*	Dummy variable that indicates if the case has indictment, in regressions for the sample of effective entries (certain homicide cases). This is based on a performance indicator used by the Office of the General Attorney defined as the ratio of cases with indictment to the effective entries.	Same as before
Days to indictment (Non-adjusted p-value)	For those cases with indictment, time since the initial complaint to get to this point.	We expect a decrease. Elimination of the URI step and the reassignments between DAs should shorten the time of prosecution.
Third Stage. Trial		
Bill of indictment *(escrito de acusación)* (Non-adjusted p-value)	A formal written document accusing the suspect(s) of having committed a crime is filed.	May increase or decrease. As a result of a better initial investigation process the DA will have enough high-quality evidence to prosecute a suspect, indict and write the bill. However, since this action depends on the indictment, the second scenario explained above may also arise, with a more careful job leading to a reduction of potentially spurious indictments.
Bill of indictment rate *(tasa de escrito de acusación)*	Dummy variable that indicates if the case has bill of indictment, in regressions for the sample of cases with indictment. This is based on a performance indicator used by the Office of the General Attorney and is defined as the ratio of cases with a bill of indictment to the cases with an indictment.	Same as before
Days to bill of indictment (Non-adjusted p-value)	For those cases with bill of indictment, time since the initial complaint to get to this point.	We expect a decrease. Elimination of the URI step and the reassignments between DAs should shorten the time of prosecution. Although indictments may take longer for the reasons above, we expect it not to dominate the net effect.

Outcome	Description	Hypothesis
	Surveys	
Index 1: Motivation and Feedback	Aggregates four questions about motivation and feedback, in which the respondent answers a number between one and ten, where ten is the highest score. 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 District Attorney 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 role, effectiveness and quality. As the previous index, the possible answers are the same. a) Are the duties that you DA/detective/CSI must develop to solve a murder clear? b) In the development of a murder investigation, do you feel you can exercise all the tasks that are assigned to you? c) How effective do you think is your team on the ultimate goal of the Attorney?s General Office to reduce rates of impunity in the city? d) How satisfied are you with the quality of the homicide investigations that you and your team investigate?	We expect an increase.
Index 3: Teamwork.	Aggregates three questions about teamwork. The respondent is also required to answers on a scale from one to ten, where ten is the highest score. a) When investigating a homicide, how aware are you of the daily tasks that other people of 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.

Outcome	Description	Hypothesis
Index 4: Coordination and progress	Aggregates questions about coordination and the progress of the investigation. The respondent agrees or disagrees with the following nine statements on a five-point scale: strongly disagree, disagree, neither agree nor disagree, agree or strongly agree. 1. The coordination of the team of detectives, CSI and DA 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 on his Executive Report is usually sufficient 4. Meetings with the team are usually carried out as often as necessary 5. There is a person in the homicide investigation team who is responsible for effectively coordinating the progress of the investigation and improving its probability of success 6. The dynamics of the current work scheme promotes the emergence of new ideas and useful innovations for the documentation of evidence and case resolution 7. It is important that the DA of the Life Unit knows the investigation in detail from the start 8. When a case is in some sense more complex than the majority of cases, it is easy to contact a specialist on homicides to ask for help 9. It is useful that the detectives and CSI participate in the definition of the criminal hypothesis and orders of further investigation activities	We expect an increase.
Information	Proportion of respondents that agree detectives and CSI 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." The respondent answers one of the following: strongly disagree, disagree, neither agree nor disagree, agree or strongly agree.	We expect an increase.

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" The respondent answers one of the following: 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 the time the respondent spends on attention to the victims in a typical work week.	We expect an increase.

Text analysis

Record of technical examination to 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 done in the scene in chronological order. For this outcome we have three variables that count the number of words in each description. One where we count the exact number of words in the report (raw), a second one where we eliminate from the description words like pronouns and connectors that mechanically appear very often know as "stop words" (method 1), and a final one where we first remove "stop words" and then remove 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 respond to the quality of the activities done at the crime scene.

Outcome	Description	Hypothesis
Keywords by topic	Proportion of words related to each topic found the the description of the crime scene. To get this proportion, we count keywords related to each topic as detailed in Table A-13 and then divide that number by the raw number of words in the description. The topics are:	We expect an increase in the frequency of these words relative to the control group.

- First respondent, as it is 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 CSI are aware of the interviews done by the detectives.

- DA or detectives, which is a proxy for coordination indicating that the CSI take into account that their work is an input for the other 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.

Outcome	Description	Hypothesis
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 residuals or 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 respond to the quality of the activities done at the crime scene.
Crime scene drawing	Variable that takes the value of 1 if the CSI team reports to have 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).

Outcome	Description	Hypothesis
Number of evidence	Number of evidence elements collected by the CSI teams as if 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).
INML requests length	Number of words written in the corpse examination report to describe the procedures to be conducted on the corpse requested to 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 the value of 1 if the methodological program was filled by the DA.	We expect an increase with treatment (not in the pre-analysis plan).
Number of criminal and research hypotheses	Number of hypothesis that appear on the methodological program of each type. We also include a variable counting the total number of hypothesis by adding the number of hypothesis of both types.	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.

Outcome	Description	Hypothesis
Distance between hypothesis	Inverse measure of how similar are the two hypothesis texts for each case. We calculate this variable only for the cases that have the two types of hypothesis to be able to compare them. We use three methods to perform this calculation. First, the tokens distance method focuses on comparing the words in each text, considering them closer if they share more words. The Damarau-Leveshtein (DL) method estimates distance using the minimal number of changes needed to transform one text into the other, where fewer transformations means less distance. The Qgram method is similar to the tokens 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 have more accurate hypotheses.
Distance with hypothesis from other cases	Inverse measure of how similar is the hypothesis of each case to hypothesis from other cases. One variable calculate this for criminal hypothesis and another one for research hipothesis. We create this measure as the average distance between the hypothesis of each case relative to all other hypothesis 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 have more accurate hypotheses.
Executive report		
Narration length	Number of words written in the executive report to describe 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 where we count the exact number of words in the report (raw), a second one where we eliminate from the narration words like pronouns and connectors that mechanically appear very often know as "stop words" (method 1), and a final one where we first remove "stop words" and then remove 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 respond to the quality of the activities done at the crime scene.

Outcome	Description	Hypothesis
Keywords by topic	Proportion of words related to each topic. To get this proportion, we count keywords related to each topic as detailed in Table A-13 and the divide that number by the raw number of words in the narration. The topics we analyzed were the following.	We expect and increase.

- Witnesses, as it is a proxy for quality indicating interactions with persons relevant for solving the case.

- Family, another proxy for quality, indicating that the detectives have contact with the family of the deceased.

- Interviews, also a proxy for quality, indicating the relevance of interviews done by the detectives for the case.

- Verification and inspection also measure quality as these concepts are related to important tasks done throughout the investigation.

Table A-2: Control variables: description

Variable	Description
Suspect arrested *in flagrante*	Dummy variable that equals one when a suspect is arrested *in fragante* at the scene.
Case started as an assault	Dummy variable that equals one if the case is an assault and turned into murder because the victims dies from injuries.
Transferred case	Dummy variable that equals one if the case is transferred to a special attorney or unit; typically reflecting jurisdictions competencies.
Case returned by Quincy	Dummy variable that equals one if the case is initially assigned to the *Secretaría de Salud* which certifies natural deaths, but it is sent back to be studied by the judicial police (CTI or Sijin)
Time of report	Dummy variable that equals one if the case was reported during the night
Newborn	Dummy variable that equals one if the victim was a newborn, usually abortion cases
Lag between complaint and death	Days from the time of occurrence of the homicide and the date in which it is known by the General Attorney Office.
Sex of the victim	Categorical variable that equals one if the female, two if male and three if unknown.
Age of the victim	Categorical variable that equals one if the victim was between zero and 17 years old, two if between 18-25, three if between 26-35, four if between 36-50, five if between 51-70, six if 71-older and seven 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. Categories are: 1 public road, 2 residence, 3 hospital, and 4 other.
City area	Categorical variable that indicates the geographic city are where the first investigative actions were carried out. Categories are 1 north, 2 west, 3 east, 4 south, 5 other.

Table A-3: Treatment group

District attorney	CSI team	Groups of detectives	
District attorney 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
District attorney 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
District attorney 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
District attorney 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 between Jan 20 and Dec 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: Effects of the intervention on key outcomes: SPOA database variables
(With controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Minimun actions-Mean effects index	0.152	0.003	0.035	0.029	0.131	0.005	0.028	1,683
		(0.052)		(0.070)		(0.055)		
Forensic actions-Mean effects index	0.176	0.001	0.019	0.004	0.031	0.001	0.004	1,683
		(0.055)		(0.062)		(0.051)		
Indictment	0.037	0.266	0.783	0.248	0.431	0.265	0.460	1,683
		(0.033)		(0.032)		(0.033)		
Bill of indictment	0.072	0.061	0.362	0.046	0.136	0.070	0.196	1,683
		(0.039)		(0.036)		(0.040)		
Days to indictment	0.285	0.540	0.795					116
		(0.464)						
Days to bill of indictment	-0.266	0.342	0.795					87
		(0.277)						
Indictment rate	0.039	0.334	0.795					1,113
		(0.041)						
Bill of indictment rate	0.350	0.097	0.439					116
		(0.209)						
Cases to be established sent 112 DA	-0.360	0.000	0.000	0.000	0.000	0.000	0.000	1,683
		(0.042)		(0.050)		(0.047)		
Cases to be established returned by 112 DA	-0.026	0.603	0.795	0.603	0.605	0.594	0.590	1,683
		(0.051)		(0.051)		(0.050)		
Unsolved cases	0.117	0.028	0.208	0.033	0.131	0.026	0.105	1,683
		(0.053)		(0.055)		(0.053)		

Notes: All variables are standardized by the mean and standard deviation of the control group. Even columns show the corresponding p-value and, in parenthesis, the standard error of the effect under each assumption on errors. Columns 2 and 3 have robust standard errors, columns 4 and 5 clustered at the level of team or "homicide squad" (DA, investigator and CSI), and columns 6 and 7 clustered at the shift level. Odd columns show the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016). Clustered standard errors and p-values for the days to indictment and bill of indictment, the indictment rate and the bill of indictment rate cannot be estimated because the Romano-Wolf correction yields a high percentage of invalid iterations.

Table A-6: Effects of the intervention on key outcomes: SPOA database variables.
Robustness exercise: Only homicide cases
(With controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Minimun actions-Mean effects index	0.172	0.002	0.018	0.017	0.085	0.002	0.012	1,571
		(0.054)		(0.072)		(0.057)		
Forensic actions-Mean effects index	0.180	0.002	0.019	0.006	0.032	0.001	0.003	1,571
		(0.058)		(0.065)		(0.052)		
Indictment	0.032	0.359	0.876	0.342	0.560	0.358	0.540	1,571
		(0.035)		(0.034)		(0.035)		
Bill of indictment	0.072	0.078	0.422	0.062	0.190	0.088	0.230	1,571
		(0.041)		(0.038)		(0.042)		
Days to indictment	0.285	0.540	0.876					116
		(0.464)						
Days to bill of indictment	-0.266	0.342	0.876					87
		(0.277)						
Indictment rate	0.032	0.459	0.876					1,016
		(0.044)						
Bill of indictment rate	0.350	0.097	0.437					116
		(0.209)						
Cases to be established sent 112 DA	-0.390	0.000	0.000	0.000	0.000	0.000	0.000	1,571
		(0.043)		(0.049)		(0.048)		
Cases to be established returned by 112 DA	-0.051	0.332	0.876	0.332	0.560	0.319	0.540	1,571
		(0.053)		(0.053)		(0.051)		
Unsolved cases	0.109	0.048	0.323	0.052	0.190	0.044	0.161	1,571
		(0.055)		(0.056)		(0.054)		

Notes: Sample includes only homicide cases, this implies we are leaving out femicide, abortion and any other type of criminal offense that may have been included. The newborn covariable is excluded. All variables are standardized by the mean and standard deviation of the control group. Even columns show the corresponding p-value and, in parenthesis, the standard error of the effect under each assumption on errors. Columns 2 and 3 have robust standard errors, columns 4 and 5 clustered at the level of team or "homicide squad" (DA, investigator and CSI), and columns 6 and 7 clustered at the shift level. Odd columns show the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016). Clustered standard errors and p-values for the days to indictment and bill of indictment, the indictment rate and the bill of indictment rate cannot be estimated because the Romano-Wolf correction yields a high percentage of invalid iterations.

Table A-7: Effects of the intervention on key outcomes: SPOA database variables.
Robustness exercise: Excluding cases connected to broader criminal cases
(With controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Minimun actions-Mean effects index	0.153	0.003	0.038	0.028	0.133	0.005	0.021	1,666
		(0.052)		(0.070)		(0.055)		
Forensic actions-Mean effects index	0.181	0.001	0.013	0.004	0.024	0.000	0.002	1,666
		(0.055)		(0.062)		(0.051)		
Indictment	0.041	0.215	0.696	0.199	0.349	0.215	0.392	1,666
		(0.033)		(0.032)		(0.033)		
Bill of indictment	0.077	0.047	0.275	0.035	0.133	0.055	0.158	1,666
		(0.039)		(0.036)		(0.040)		
Days to indictment	0.285	0.540	0.784					116
		(0.464)						
Days to bill of indictment	-0.266	0.342	0.710					87
		(0.277)						
Indictment rate	0.045	0.267	0.699					1,097
		(0.041)						
Bill of indictment rate	0.350	0.097	0.437					116
		(0.209)						
Cases to be established sent 112 DA	-0.364	0.000	0.000	0.000	0.000	0.000	0.000	1,666
		(0.042)		(0.049)		(0.047)		
Cases to be established returned by 112 DA	-0.004	0.936	0.935	0.936	0.942	0.935	0.936	1,666
		(0.047)		(0.047)		(0.046)		
Unsolved cases	0.121	0.025	0.182	0.029	0.133	0.023	0.081	1,666
		(0.054)		(0.055)		(0.053)		

Notes: Cases that were found to be connected to a broader criminal case (*conexidad*) are excluded. These are 17 cases for which the protocol was not fully implemented. All variables are standardized by the mean and standard deviation of the control group. Even columns show the corresponding p-value and, in parenthesis, the standard error of the effect under each assumption on errors. Columns 2 and 3 have robust standard errors, columns 4 and 5 clustered at the level of team or "homicide squad" (DA, investigator and CSI), and columns 6 and 7 clustered at the shift level. Odd columns show the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016). Clustered standard errors and p-values for the days to indictment and bill of indictment, the indictment rate and the bill of indictment rate cannot be estimated because the Romano-Wolf correction yields a high percentage of invalid iterations.

Table A-8: Effects of the intervention on key outcomes: SPOA database variables.
Robustness exercise: Changing clusters
(With controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Lab Group*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Minimun actions-Mean effects index	0.152	0.003	0.035	0.229	0.542	0.322	0.686	1,683
		(0.052)		(0.127)		(0.154)		
Forensic actions-Mean effects index	0.176	0.001	0.019	0.009	0.052	0.023	0.151	1,683
		(0.055)		(0.068)		(0.077)		
Indictment	0.037	0.266	0.783	0.283	0.542	0.421	0.686	1,683
		(0.033)		(0.034)		(0.046)		
Bill of indictment	0.072	0.061	0.362	0.047	0.186	0.129	0.444	1,683
		(0.039)		(0.037)		(0.048)		
Cases to be established sent 112 DA	-0.360	0.000	0.000	0.000	0.002	0.044	0.227	1,683
		(0.042)		(0.099)		(0.179)		
Cases to be established returned by 112 DA	-0.026	0.603	0.795	0.596	0.595	0.630	0.686	1,683
		(0.051)		(0.050)		(0.055)		
Unsolved cases	0.117	0.028	0.208	0.041	0.186	0.114	0.444	1,683
		(0.053)		(0.057)		(0.074)		

Notes: Cluster team is changed to be based on the combination of DA and CSI groups, instead of the full "homicide squad". Cluster shift is changed for lab group, which implies putting together in the same cluster three groups of CSIs, serving in the same shift. We do not show the rate and days variables in this table because the Romano-Wolf correction yields a high percentage of invalid iterations. All variables are standardized by the mean and standard deviation of the control group. Even columns show the corresponding p-value and, in parenthesis, the standard error of the effect under each assumption on errors. Columns 2 and 3 have robust standard errors, columns 4 and 5 clustered at the level of team or "homicide squad" (DA, investigator and CSI), and columns 6 and 7 clustered at the shift level. Odd columns show the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016)

Table A-9: Effects of the intervention on key outcomes: SIG database variables
(With controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Minimun actions SIG-Mean effects index	0.294	0.000	0.000	0.000	0.000	0.000	0.000	1,612
		(0.055)		(0.071)		(0.056)		
Forensic actions SIG-Mean effects index	0.267	0.000	0.000	0.000	0.000	0.000	0.000	1,612
		(0.058)		(0.060)		(0.059)		
Order to interview	0.092	0.058	0.163	0.067	0.185	0.063	0.181	1,612
		(0.049)		(0.050)		(0.050)		
Order to locate person of interest	0.098	0.062	0.163	0.081	0.185	0.071	0.181	1,612
		(0.053)		(0.056)		(0.054)		
Order to individualize suspect	0.085	0.133	0.163	0.126	0.185	0.126	0.181	1,612
		(0.057)		(0.056)		(0.056)		
Number of extensions of the assignment	-0.129	0.006	0.031	0.009	0.051	0.004	0.025	1,612
		(0.047)		(0.049)		(0.045)		
Average days of extension of the assignment	-0.124	0.007	0.031	0.012	0.061	0.007	0.032	1,612
		(0.046)		(0.050)		(0.046)		
Days of delay to fulfill the assignment	-0.103	0.029	0.116	0.032	0.118	0.041	0.151	1,612
		(0.047)		(0.048)		(0.050)		

Notes: All variables are standardized by the mean and standard deviation of the control group. Even columns show the corresponding p-value and, in parenthesis, the standard error of the effect under each assumption on errors. Columns 2 and 3 have robust standard errors, columns 4 and 5 clustered at the level of team or "homicide squad" (DA, investigator and CSI), and columns 6 and 7 clustered at the shift level. Odd columns show the p-value with the Romano-Wolf correction Romano and Wolf (2005, 2016)

Table A-10: Effects of the intervention on key outcomes: text analysis on corpse examination report (With controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Description Length (Raw)	-0.461	0.000	0.000	0.000	0.000	0.000	0.000	1,512
		(0.043)		(0.051)		(0.042)		
Description Length (Method 1)	-0.452	0.000	0.000	0.000	0.000	0.000	0.000	1,512
		(0.043)		(0.052)		(0.042)		
Description Length (Method 2)	-0.447	0.000	0.000	0.000	0.000	0.000	0.000	1,512
		(0.043)		(0.052)		(0.042)		
Keywords by topic (in proportion to total length)								
First Respondent	0.250	0.000	0.000	0.000	0.000	0.000	0.000	1,512
		(0.056)		(0.066)		(0.055)		
Interviews	-0.033	0.549	0.568	0.578	0.573	0.551	0.540	1,512
		(0.055)		(0.060)		(0.056)		
DA or Detectives	0.560	0.000	0.000	0.000	0.000	0.000	0.000	1,512
		(0.087)		(0.136)		(0.086)		
Crime Scene Data	-0.071	0.148	0.280	0.208	0.358	0.171	0.306	1,512
		(0.049)		(0.057)		(0.052)		
Evidence (EMP)	-0.251	0.000	0.000	0.000	0.000	0.000	0.000	1,512
		(0.050)		(0.055)		(0.057)		
Lab	-0.151	0.004	0.018	0.030	0.120	0.003	0.015	1,512
		(0.052)		(0.069)		(0.051)		
Findings Description Length (Raw)	-0.097	0.055	0.158	0.103	0.280	0.056	0.157	1,432
		(0.051)		(0.060)		(0.051)		
Crime Scene Drawing	0.379	0.000	0.000	0.000	0.000	0.000	0.000	1,594
		(0.052)		(0.054)		(0.052)		
Number of Evidence	-0.093	0.025	0.090	0.024	0.120	0.023	0.096	1,643
		(0.041)		(0.041)		(0.041)		
INML Requests Length (Raw)	-0.279	0.000	0.000	0.000	0.000	0.000	0.000	1,487
		(0.048)		(0.059)		(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 A-11: Effects of the intervention on key outcomes: text analysis on methodological program (With controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Has Metodological Program	-0.457	0.000	0.000	0.000	0.000	0.000	0.000	1,683
		(0.067)		(0.087)		(0.074)		
Number of Criminal Hypothesis	0.314	0.000	0.000	0.000	0.000	0.000	0.000	1,523
		(0.051)		(0.067)		(0.059)		
Number of Research Hypothesis	0.967	0.000	0.000	0.000	0.000	0.000	0.000	1,523
		(0.064)		(0.113)		(0.083)		
Total Number of Hypothesis	0.742	0.000	0.000	0.000	0.000	0.000	0.000	1,523
		(0.058)		(0.097)		(0.073)		
Distance between hypothesis by method								
Method 1 - Tokens	-0.863	0.000	0.000	0.000	0.000	0.000	0.000	515
		(0.122)		(0.122)		(0.132)		
Method 2 - DL	-0.306	0.018	0.021	0.019	0.019	0.024	0.026	515
		(0.129)		(0.130)		(0.135)		
Method 3 - Qgrams	-0.470	0.000	0.000	0.000	0.000	0.000	0.001	515
		(0.122)		(0.123)		(0.129)		
Distance with hypothesis from other cases by type								
Criminal Hypothesis	-1.387	0.000	0.000	0.000	0.000	0.000	0.000	1,076
		(0.067)		(0.095)		(0.075)		
Research Hypothesis	-2.358	0.000	0.000	0.000	0.000	0.000	0.000	515
		(0.137)		(0.164)		(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 A-12: Effects of the intervention on key outcomes: text analysis on executive report
(With controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Robust Errors		*Cluster Team*		*Cluster Shift*		
	Treatment Effect	PValue	RW PValue	PValue	RW PValue	PValue	RW PValue	N
Narration Length (Raw)	-0.098	0.033	0.189	0.121	0.539	0.041	0.232	1,557
		(0.046)		(0.063)		(0.048)		
Narration Length (Method 1)	-0.109	0.016	0.146	0.084	0.539	0.022	0.184	1,557
		(0.045)		(0.063)		(0.048)		
Narration Length (Method 2)	-0.103	0.023	0.180	0.102	0.539	0.030	0.206	1,557
		(0.045)		(0.063)		(0.048)		
Keywords by topic (in proportion to total length)								
Witnesses	0.067	0.232	0.293	0.300	0.562	0.248	0.278	1,557
		(0.056)		(0.064)		(0.058)		
Family	-0.103	0.050	0.225	0.090	0.539	0.049	0.232	1,557
		(0.052)		(0.061)		(0.052)		
Video-Cams	0.079	0.146	0.293	0.211	0.562	0.146	0.278	1,557
		(0.054)		(0.063)		(0.054)		
Interviews	0.100	0.083	0.293	0.225	0.562	0.079	0.278	1,557
		(0.058)		(0.082)		(0.057)		
Verification	-0.090	0.094	0.293	0.186	0.562	0.097	0.278	1,557
		(0.054)		(0.068)		(0.054)		
Inspection	0.122	0.026	0.183	0.101	0.539	0.026	0.196	1,557
		(0.055)		(0.074)		(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 A-13: 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-14: Descriptive statistics: survey controls

Variables	Full sample Mean	Min	Max	Control Mean	Treat. Mean	MeanDiff (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)

Table A-14: Descriptive statistics: survey controls

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

Continued on next page

Table A-14: Descriptive statistics: survey controls

Variables	Full sample Mean	Min	Max	Control Mean	Treat. Mean	MeanDiff (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.