Randomization Inference for Before-and-After Studies with Multiple Units: An Application to a Criminal Procedure Reform in Uruguay*

Matias D. Cattaneo† Carlos Díaz‡ Rocío Titiunik§

October 19, 2024

Abstract

We study the immediate impact of a new code of criminal procedure on crime. In November 2017, Uruguay switched from an inquisitorial system (where a single judge leads the investigation and decides the appropriate punishment for a particular crime) to an adversarial system (where the investigation is now led by prosecutors and the judge plays an overseeing role). To analyze the short-term effects of this reform, we develop a randomization-based approach for before-and-after studies with multiple units. Our framework avoids parametric time series assumptions and eliminates extrapolation by basing statistical inferences on finite-sample methods that rely only on the time periods closest to the time of the policy intervention. A key identification assumption underlying our method is that there would have been no time trends in the absence of the intervention, which is most plausible in a small window around the time of the reform. We also discuss several falsification methods to assess the plausibility of this assumption. Using our proposed inferential approach, we find statistically significant short-term causal effects of the crime reform. Our unbiased estimate shows an average increase of approximately 25 police reports per day in the week following the implementation of the new adversarial system in Montevideo, representing an 8 percent increase compared to the previous week under the old system.

Keywords: crime, criminal law, criminal procedure, before and after studies, event studies, interrupted time series studies, randomization inference.

---

*A preliminary version of this paper circulated under the title "Breaking the code: Can a new penal procedure affect public safety". We thank Giorgio Chiovelli, Scott Cunningham, Jennifer Doleac, Libor Dušek, Jeff Grogger, Dorothy Kronick, Emily Owens, Zachary Peskowitz, Yotam Shem-Tov, Gonzalo Vazquez-Bare and participants at the 2021 Chicago/LSE Conference on the Economics of Crime and Justice, the 2021 Annual Meeting of the Latin American and Caribbean Economic Association, the 2021 LACEA/RIDGE Workshop on the Economics of Crime, the 2021 APSA Annual Meeting, the 2020 Annual Conference of the Latin American Society for Political Methodology, the Summer 2020 Online Seminar on the Economics of Crime, and Syracuse University, Universidad de Chile, Universidad ORT, and Universidad de Montevideo Economics Department Seminar Series for thoughtful comments and suggestions. Cattaneo and Titiunik gratefully acknowledge financial support from the National Science Foundation (SES-2241575).

†Department of Operations Research and Financial Engineering, Princeton University.

‡Department of Economics, Universidad Alberto Hurtado.

§Department of Politics, Princeton University.

1 Introduction

On November 1st, 2017, a new code of criminal procedure (CCP or *Código del Proceso Penal* in Spanish) became effective in Uruguay. The reform ended the old inquisitive and written tradition and established an accusatory, adversarial, oral, and public criminal system in its place. Under the old CCP, the inquisitorial judge led the investigation and decided the appropriate punishment for a particular crime. Once the new criminal procedure came into effect, judges became neutral referees focused on ensuring the correct procedure, while prosecutors became responsible for leading the investigation. Under the new regime, the investigation is the exclusive responsibility of prosecutors who, representing society, must present evidence to judges. The judges then decide what evidence to admit into the record and what evidence to exclude. This clear separation between the roles of prosecutors and judges is a key reason why adversarial systems are generally considered fairer and less susceptible to abuse compared to inquisitorial systems.

Despite its advantages, this type of reform can lead to unintended changes in the costs associated with offending. The new CPP may have influenced both the severity and the certainty of punishments through several channels, thus affecting the propensity of individuals to commit crimes. For instance, the adversarial system introduced substantial changes to the adjudication process of criminal law (e.g., plea bargain, alternatives to oral trials, and exceptional use of preventive detention) that have been widely used by prosecutors and might result in lighter sentences. In addition, prosecutors faced both a totally new role under the new legal system and a significant increase in their workload. Meanwhile, the new CCP made police officers conduct investigations under new supervision and new rules and introduced changes to the implementation process, all significant adjustments in the practice of policing that can alter officer behavior. These changes resulted in several coordination problems between prosecutors and police officers during the first months of the adversarial system, potentially affecting the probability of detection and conviction.

In this paper, we propose a methodology to evaluate the immediate impact of this procedural reform on the number of offenses reported to the police in Montevideo, the capital and largest city of Uruguay. The reform was introduced nationwide simultaneously but, in contrast to a standard interrupted time series design (e.g. Cook et al., 2002), we observe multiple units in each time period because we collect crime reports (*denuncias*) at the neighborhood level for all neighborhoods in Montevideo. This leads to a before-and-after observational study with multiple units: a setting where several cross-sectional units are first observed for several periods, an intervention is then introduced at the same time for all units, and finally, the same units are observed after the intervention for several more periods.

Analyses of this type of observational study are often based on linear panel data models where the outcome for unit i in period t is regressed on a unit's fixed effect, a time fixed effect, time trends, unit-level covariates, and time-indexed treatment indicators. See Freyaldenhoven et al. (2019), Miller (2023), and Freyaldenhoven et al. (Forthcoming) for reviews of the general setup. Although this strategy is generally flexible, it requires many units and many periods and crucially relies on time homogeneity over a typically long time span. The number of parameters to be estimated is often large, and multicollinearity issues are common. Moreover, a relatively small number of units can make estimation of the time fixed effects unstable, while a similar problem occurs for the estimation of unit fixed effects when there are few time periods. More generally, the linear panel event study approach relies on a parametric specification to model the global trend of the outcome to separate it from the effect of the treatment. As a consequence, this observational method is most useful when many units are observed over many time periods and the time homogeneity of the flexible parametric model is plausible, allowing researchers to model the time series globally.

Event study methods are not particularly suitable for our study, because we have a small number of cross-sectional units. Moreover, even if our before-and-after study could be analyzed using panel event study methods, the novelty of our approach lies in the attempt to identify a possible immediate effect of the new CPP on the number of police reports, which requires focusing on a small number of time periods and makes global estimation less reliable. Finally, a crucial feature of before-and-after studies like ours is the lack of a control group, which requires special attention to the possibility of time confounders and suggests focusing on small windows around the intervention to isolate the effect of the treatment. In fact, Miller (2023) goes as far as to exclude before-and-after designs from the data structures that allow for event study estimation because “we cannot separate the effects of the event from other confounders that occur in calendar time, and so cannot identify treatment effects” (p. 207). In sum, the lack of a control group, the small number of observations, and our interest in immediate effects call for an alternative, complementary approach to event study analyses.

Building on the causal inference literature (Rosenbaum, 2002b, 2010), we develop a randomization-based or Fisherian approach for multiple-unit before-and-after studies. Our approach is based on localizing around the time of the event rather than modeling the time series globally, and thus complements existing panel event study observational methods. In our framework, the units' potential outcomes are non-stochastic, and we assume that the assignment to receiving the intervention has a distribution over the pre-intervention and post-intervention periods. Although, in reality, all units are treated in the post-intervention period and untreated in the pre-intervention period, we consider an assignment mechanism where each unit could have been treated either in the pre-intervention or in the post-intervention period with equal probability. Because the potential outcomes are non-stochastic, this assignment mechanism provides the distribution of any test statistic under the sharp null hypothesis that the intervention has no effect on any unit. Since the same units appear before and after the intervention and units can only be treated on one side, the treatment indicator is perfectly negatively correlated within a unit; our randomization-based approach allows us to provide p-values that correctly incorporate this correlation. Moreover, our inferences are robust to spillovers (Rosenbaum, 2007), a common concern in our crime application, where our cross-sectional units are geographic regions, and in many other event study applications, where units can interfere with each other.

Our proposed inference approach for before-and-after observational studies with multiple units contributes to a rich literature on randomization inference methods for program evaluation and causal inference. For example, randomization inference methods have been successfully deployed to analyze natural experiments (Ho and Imai, 2006), instrumental variables (Imbens and Rosenbaum, 2005; Kang, Peck and Keele, 2018), and regression discontinuity designs (Cattaneo et al., 2015, 2017). Our approach is also connected to the so-called regression discontinuity design in time (Hausman and Rapson, 2018), where there are multiple cross-sectional units and the running variable is defined as time to the event. However, as noted in Cattaneo and Titiunik (2022), this setting is not a standard regression discontinuity design, and indeed Hausman and Rapson (2018) identify important limitations in deploying standard regression discontinuity methodology to analyze before-and-after observational studies: a lack of cross-sectional variation in policy implementation, which leads to excessive extrapolation and insufficient number of units around the time cutoff. Our proposed methodology circumvents these limitations by localizing near the time of treatment adoption, thereby avoiding the extrapolation inherent in standard parametric regression methods.

Our ability to recover the causal effect of the intervention depends on the assumption that, in a small window around the time of the intervention, there are no trends in the potential outcomes—that is, the average potential outcomes under control in every period in a window around the time of the intervention are the same as they would have been in the periods after the intervention in the absence of the treatment. This assumption is implausible when the window is large but becomes more plausible when the window is sufficiently small. Because we observe multiple units for every time period, we can analyze the smallest possible windows around the time of the intervention, thereby reducing extrapolation to a minimum. This is especially important in our setup, where there is no comparison group with which to estimate trends in the absence of the treatment.

To assess the plausibility of the no-trend assumption, we employ a falsification analysis that replicates the analysis in the past in a window of the same length and respects the time structure of the actual policy. The new CCP in Uruguay became effective on Wednesday, November 1st, 2017, and our smallest window compares crime reports on this day to crime reports the day before, Tuesday, October 31st, 2017. Our day-of-the-week falsification test compares average outcomes on Tuesday versus Wednesday using prior weeks, while our turn-of-the-month effects compare outcomes on October 31 against November 1st in prior years.

Using our framework, we find that the implementation of Uruguay's new criminal procedure resulted in an increase in the number of crimes reported to the police in Montevideo. The point estimates suggest a local increase of about 8.2 percent in the week after the reform, compared to the week immediately before the change from an inquisitorial code to an accusatory code of criminal procedure. These estimates are consistent with the view that legal codes are far from innocuous (Acemoglu and Johnson, 2005). Our Fisherian inference approach indicates that the effect of the policy is distinguishable from zero as soon as seven days after the intervention (p-value ¡ 0.05). Our inferences also reject the null hypothesis of no effect when multiple windows between 1 and 14 days around the time of the intervention are considered jointly.

1.1 Criminal Procedure and Crime

Our application contributes to the literature on the economics of crime, documenting the impact of procedural law on delinquent behavior. Most of these empirical studies have put the spotlight on the probability of detection and conviction, reporting results consistent with our main findings (Atkins and Rubin, 2003; Dalla Pellegrina, 2008; Soares and Sviatschi, 2010; Duŝek, 2015; Zorro Medina et al., 2020). For example, Atkins and Rubin (2003) report an increase in the number of crimes in the United States after the Supreme Court imposed the "exclusionary rule" in the 1960s, which prevents evidence obtained in violation of the Fourth Amendment from being used in a court of law, thus increasing the cost of police investigation. Dalla Pellegrina (2008) finds a positive effect of trial duration on crimes in Italy, supporting the hypothesis that increasing the probability of prescription increases the willingness to offend. Similarly, Soares and Sviatschi (2010) show that more efficient courts in Costa Rica are associated with lower crime rates. Duŝek (2015) exploits a criminal procedure reform in the Czech Republic and shows that simplified procedures incentivize police to reallocate enforcement efforts toward crimes that have reduced enforcement costs.

Regarding the impact of procedural reforms on crime in Latin America, mixed results have been documented on the probability of detection and conviction when an inquisitorial CCP is replaced by an adversarial CCP (Langer, 2007). For example, Zorro Medina et al. (2020) exploited the quasi-experimental implementation of the new CCP in Colombia and documented an increase in both violent and property crimes, suggesting that the increase in property crimes could be due to a decrease in the probability of apprehension or detection under the new regime. These findings align with Becker's (1968) canonical model, which predicts that rational offenders respond to the probability of conviction. In contrast, Kronick (2019) documented significant increases in arrest rates in Colombia and Venezuela following the implementation of new CCPs, 40% and 80%, respectively. However, neither country experienced a corresponding change in crime rates.

New criminal procedures might also influence crime by altering the severity of associated punishments. These reforms often introduce plea bargaining mechanisms, in which defendants enter a guilty plea in exchange for potential sentence reductions or changes (Langer, 2007). Despite some exceptions (Abrams, 2011; Frazier et al., 2018), empirical evidence supports the hypothesis that sentences for defendants who plead guilty are significantly shorter than those convicted at trial (Bushway and Redlich, 2012; Bushway et al., 2014; Ulmer and Bradley, 2006; Ulmer et al., 2010). In addition to suggesting that these reforms may affect crime by creating an expectation of less severe punishment, the literature also shows that they enhance deterrence by speeding up the criminal procedure (Listokin, 2007; Zorro Medina et al., 2020).

Criminal procedure reforms in Latin America have also reduced the pre-trial detention population, likely due to increased efficiency rather than improved human rights protections (Zorro Medina et al., 2020). This reduction, evident in Uruguay where pre-trial detention dropped from 70% in 2015 to 20% in 2020 (Institute for Crime & Justice Policy Research, 2021), suggests a decrease in incapacitation that could also potentially increase crime rates.

2 Criminal Procedure Reform in Uruguay

Uruguay was one of the last countries in Latin America to reform its code of criminal procedure. A wave of reforms in Latin America started in the early 1990s and included most of the countries in the region, with the exception of Brazil and Cuba, which still preserve their old systems (Fandiño and González Postigo, 2020). This transition towards an adversarial system is considered one of the most significant transformations in Latin American criminal procedures in the past two centuries (Langer, 2007).

Latin America's criminal procedures used to be mostly governed by inquisitorial and written systems originally adopted in either the nineteenth century or the beginning of the twentieth century. Langer (2007) describes these systems as generally sharing two key characteristics, which were also present in the Uruguayan case. First, criminal procedures were divided into two written phases: the pre-trial investigation phase and the verdict and sentencing phase. Given its written nature, the process was based primarily on a dossier compiled by the judge and the police, including the evidence that would be evaluated during the verdict phase. Second, the pretrial investigation was led by the judge. This meant that the judge had both an investigatory and adjudicatory role: the same judge who collected the evidence during the pretrial phase evaluated it during the verdict phase. Furthermore, the investigation was kept secret from the defendant, and pretrial detention was the norm rather than the exception. The increasing recognition of human rights (1970s) and the transition to democracy (1980s and 1990s) fueled the perception among local actors in Latin America that these inquisitorial systems did not meet minimum due process standards (Maier and

Struensee, 2000; Langer, 2007).

With the advent of the accusatory, adversarial, oral, and public system in Uruguay and the rest of Latin America, criminal procedures now rest on three well-defined stages: the formalization of the investigation, a preliminary hearing, and a trial phase (Fandiño and González Postigo, 2020). During the first stage, the General Prosecution Office decides whether to undertake criminal prosecution against the defendant or not (i.e., the judicialization of the case). If the investigation is formalized, the second stage consists of a preliminary hearing aimed at discussing the outcomes of the investigation conducted by the police and led by a prosecutor. Finally, there is a trial phase in which an oral and public discussion takes place for those cases that could not be solved by an alternative outcome. Uruguay has a classic adversarial procedure where defense and prosecution arguments are presented at a hearing.

The implementation of Uruguay's adversarial CCP was a complex, multi-year process. Although the law was passed on December 19, 2014, the new adversarial code went into effect on November 1, 2017. The relevance and magnitude of such a reform caused local authorities to take almost three years to bring the new code into force. The adversarial system represented major changes for several of the agents involved in the criminal law adjudication process, in particular for prosecutors and the police. As explained above, prosecutors have new role in the adversarial system relative to what they used to do under the inquisitorial CCP. Under the new system, they are now in charge of the criminal investigation, representing both a change in their tasks and a significant increase in their workload. The impact was so severe that the prosecutors' union opposed the new CCP, citing overwhelming workloads. Some prosecutors handled several hundred cases at once, leading many to take mental health days during the first year of the code's implementation (Solomita, 2019). Regarding the police, officers must investigate under new rules and supervision, with a prosecutor overseeing the case instead of a judge.

In this context, as we discuss in the next section, Uruguay experienced an unprecedented increase in crime following the implementation of the new CPP. Although there was widespread agreement on the benefits of transitioning from an inquisitorial to an adversarial system, the new legal framework was nonetheless blamed for the rise in insecurity by the Uruguayan government. The President and the Interior Minister referred to the sharp increase in the number of police reports as the "November Effect" (*Efecto Noviembre* in Spanish), alluding to the month in which the new CCP was enacted (Palumbo, 2018).

2.1 Data and Overall Patterns

We collected data on the offenses reported to the police in Montevideo from the Ministry of Interior of Uruguay. Montevideo is the capital and largest city of Uruguay, home to 40% of Uruguay's total population. We first present a descriptive analysis of the years before and after the new criminal procedure was entered into force. Since the new CCP came into effect on November 1st, 2017, we start by building a two-year symmetric window from November 1st, 2016 trough October 31st, 2018. Table 1 reports the average number of crimes reported daily to the police during the defined time frame, as well as the corresponding figures for the years before and after the switch to the new CCP. An average of 331 crimes were reported to police every day, while just three categories account for more than 7 of every 10 reports processed in Montevideo: theft ($\approx 46\%$ of all offenses reported to police), robbery ($\approx 16\%$) and domestic violence ($\approx 10\%$).

Table 1 shows a significant increase in the number of police reports in Montevideo during the first year of the new CCP. When we compare the last year of the old inquisitorial system to the first year of the new adversarial system, the total number of offenses reported daily to the police increases by approximately 95 incidents (from roughly 283 to 378 per day), a statistically significant difference that represents a 33% increase in daily offenses. This upward trend is also reflected in the three most frequent crimes reported in Montevideo. Regarding property crimes, the number of thefts and robberies reported to police every day

Table 1: Crime in Montevideo

	Two-Year Window 11/01/16 to 10/31/18	Old CCP (a)	New CCP (b)	Δ (b)-(a)
Theft	152.19	130.05	174.33	44.28***
	46%	(0.942)	(1.221)	(1.543)
Robbery	51.81	41.60	62.02	20.42***
	16%	(0.426)	(0.614)	(0.747)
Domestic Violence	33.66	32.79	34.53	1.75***
	10%	(0.379)	(0.382)	(0.539)
Other Crimes	92.98	78.98	106.99	28.02***
	28%	(0.662)	(0.869)	(1.092)
Total Police Reports	330.63	283.41	377.87	94.47***
	100%	(1.524)	(2.013)	(2.525)
Days	730	365	365	

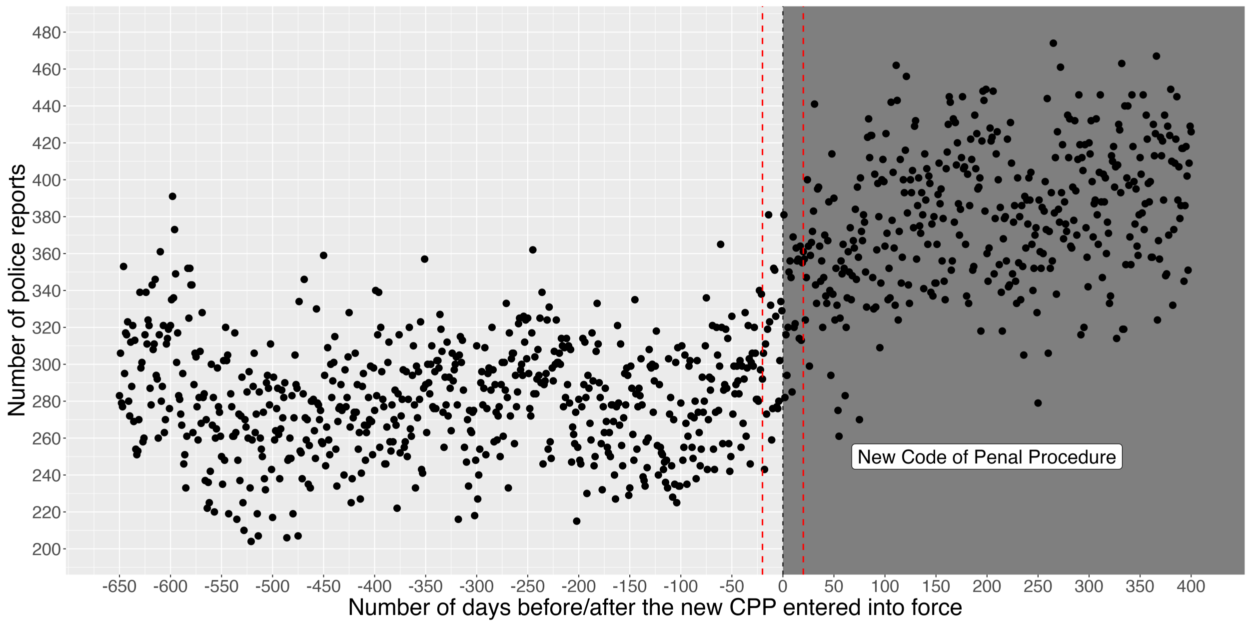
Offenses reported to police, average number per day.
Percentage of the total number of reports in *italic*; standard errors in parentheses.
*** Difference between (b) and (a) is significant at the 0.01 level.
Old CCP: inquisitorial system (November 1st, 2016 to October 31st, 2017).
New CCP: adversarial system (November 1st, 2017 to October 31st, 2018).
Source: Ministry of Interior of Uruguay.

increased by 34% and 49%, respectively. Reports of domestic violence, the most frequent crime against the person, exhibited a smaller but statistically significant increase of 5%.

An initial visual inspection of the data suggests a strong association between the increase in police reports and the timing of the criminal procedure reform. Figure 1 plots the daily number of offenses reported to the police in Montevideo from 600 days before the new CPP came into effect (early 2016) until 400 days after (late 2019), highlighting the period under the adversarial system (grey area).

Figure 1 suggests a sudden increase in the number of police reports shortly after the new penal procedure came into effect, consistent with the previously mentioned "November Effect." In the following section, we develop a randomization inference framework to determine whether this visual discontinuity in the plot can be attributed to the implementation of the adversarial system. To identify a potential immediate causal effect, our design will focus on time windows that span 20 days before and 20 days after the date of implementation of the new CPP, as marked by the dashed red lines in Figure 1.

Figure 1: Crime in Montevideo
(total police reports, 2016-2019)



3 Randomization Inference for Multiple-Unit Before-and-After Studies

Our goal is to evaluate the effect of Uruguay's new CCP on the number of crimes reported to the police in Montevideo. Given the possibility of time trends and other confounders, the patterns reported in the prior section cannot be taken as conclusive evidence that Montevideo's crime wave was caused by the modifications suffered by Uruguay's criminal procedures.

We develop a randomization inference framework to provide further empirical evidence that may address these concerns. Our framework is designed to analyze a before-and-after study with multiple units, that is, a setting where multiple cross-sectional units are first observed for some periods, a reform affecting all units is introduced at the same for all units, and then the same units are observed after the reform for some more periods.

3.1 Setup

We adopt a Fisherian framework with non-random potential outcomes. Every cross-sectional unit $i = 1, 2, \ldots, n$, is observed in $T > 0$ time periods before and after an intervention or treatment goes into effect at a precise and known moment in time, which we normalize to zero. We index the number of periods before and after the treatment goes into effect by $t = -T, -T+1, \ldots, -1, 0, 1, \ldots, T-1, T$. The fixed potential outcomes under treatment and control t periods after the intervention are denoted by $y_{i,t}(1)$ and $y_{i,t}(0)$, respectively. In our study, the new CCP began precisely at 12:00 AM on November 1st, 2017. An important feature of this assignment is that the treatment is active at the same specific moment for all units so that all periods before $t = 0$ are untreated and all periods after $t = 0$ (including it) are treated.

We define a window including τ periods before and τ periods after the intervention as $\mathcal{W}_\tau = \{-\tau, -\tau+1, \ldots, 0, 1, \ldots, \tau-2, \tau-1\}$; for simplicity, we consider only symmetric windows. The length of $\mathcal{W}_\tau$ will depend on the definition of the unit of time. For example, if time is measured in days, $\mathcal{W}_1 = \{-1, 0\}$ includes the day immediately before implementation and the day immediately after the implementation. Since each unit appears in every period, $\mathcal{W}_\tau$ has $n \times 2\tau$ observations.

In the realized, actual treatment assignment, all units are simultaneously treated for all $t \geq 0$ and untreated for all $t < 0$. In order to deploy Fisherian randomization-based methods for inference, we conceive a mechanism that assigns the treatment to either all pre-intervention periods or all post-intervention periods in $\mathcal{W}_\tau$, independently for every unit. For example, if the window is $\mathcal{W}_2 = \{-2, -1, 0, 1\}$, the assignment mechanism assigns all the post-intervention periods ($t \in \{0, 1\}$) to treatment—and therefore assigns all the pre-intervention periods ($t \in \{-2, -1\}$) to control—with some known probability. This assignment of all periods before and after the intervention to the same treatment condition respects the simultaneity of treatment assignment for all units that is characteristic of all before-and-after settings.

For example, although in the realized assignment all neighborhoods in Montevideo were under the new CCP on November 1, 2017, and under the old CCP on October 31, 2017, we can imagine that the assignment of neighborhoods to treatment could have been different, with some neighborhoods instead being assigned to treatment on October 31, 2017, and control on November 1, 2017. Our main assumption is a probability distribution over the assignment of treatment to the sets $\{t : t < 0, t \in \mathcal{W}_\tau\}$ and $\{t : t \geq 0, t \in \mathcal{W}_\tau\}$. Thus, for every unit $i = 1, \ldots, n$, we imagine that the treatment could have been given either before or after the date of the intervention for a small window around the cutoff. Because every unit has two possible assignments ($\{t : t < 0, t \in \mathcal{W}_\tau\}$ treated and $\{t : t \geq 0, t \in \mathcal{W}_\tau\}$ control, or vice-versa), the realized treatment assignment vector is one of the 2^n possible values that the vector could have taken.

To formalize, we define a binary variable Z_i that is equal to one when unit i receives the treatment in all post-intervention periods and is equal to zero when i receives the treatment in all pre-intervention periods. The treatment indicator $D_{i,t}$, which is equal to one if unit i is treated at time t and zero otherwise, is thus defined as

$$D_{i,t} = Z_i \mathbb{1}(t \geq 0) + (1 - Z_i)\mathbb{1}(t < 0).$$

Note that $D_{i,t} = D_{it'}$ for all $t \geq 0$ and $t' \geq 0$, and $D_{i,t} = D_{i,t'}$ for all $t < 0$ and $t' < 0$; therefore $D_{i,0} = 1$ implies $D_{i,t} = 1$ for all $t \geq 0$ and $D_{i,t} = 0$ for all $t < 0$ and, vice-versa, $D_{i,0} = 0$ implies $D_{i,t} = 0$ for all $t \geq 0$ and $D_{i,t} = 1$ for all $t < 0$.

In a before-and-after setting, it follows that $D_{i,-1} + D_{i,0} = 1$, that is, exactly one of two the periods immediately before and after the treatment is introduced is assigned to treatment, and the other is assigned to control. Thus, to establish the distribution of $D_{i,t}$, it suffices to specify the distribution of the random variable $D_{i,0}$. We thus define the treatment assignment vector $\mathbf{D}_0 = (D_{1,0}, D_{2,0}, \ldots, D_{n,0})'$, whose distribution is given by

$$\mathbb{P}(\mathbf{D}_0 = \mathbf{d}) = \prod_{i=1}^{n} \mathbb{P}(D_{i0} = 1)^{d_{i,0}}[1 - \mathbb{P}(D_{i0} = 1)]^{1-d_{i,0}}, \qquad \mathbf{d} \in \{0,1\}^n.$$

We assume that for each unit, we flip an independent coin to determine whether the pre- or post-intervention period is assigned to treatment. We understand the assignment mechanism as a uniform stratified experiment where, for every unit, either all time periods such that $t \geq 0$ are treated and all time periods such that $t < 0$ are control, or vice versa.

When $D_{i0} = 1$, treatment is given to all periods such that $t \geq 0$, so that all outcomes observed before $t = 0$ are control and all outcomes observed starting at $t = 0$ are treated. The situation is reversed when $D_{i0} = 0$. Therefore, the observed outcomes are

$$\begin{aligned} y_{i,t} &= y_{i,t}(0)D_{i0} + y_{i,t}(1)(1 - D_{i0}) \qquad \text{for } t < 0 \\ y_{i,t} &= y_{i,t}(1)D_{i0} + y_{i,t}(0)(1 - D_{i0}) \qquad \text{for } t \geq 0 \end{aligned}$$

for $i = 1, \ldots, n$.

For each unit, we define the average observed outcomes for the pre-intervention and the post-intervention period, in the window $\mathcal{W}_\tau$ around the time of the intervention:

$$\bar{y}_{i,-\tau} = \frac{1}{\tau} \sum_{t \in \mathcal{W}_\tau} \mathbb{1}(t < 0) y_{i,t}, \qquad \text{and} \qquad \bar{y}_{i,\tau} = \frac{1}{\tau} \sum_{t \in \mathcal{W}_\tau} \mathbb{1}(t \geq 0) y_{i,t}.$$

The corresponding average potential outcomes are

$$\begin{aligned} \bar{y}_{i,-\tau}(0) &= \frac{1}{\tau} \sum_{t \in \mathcal{W}_\tau} \mathbb{1}(t < 0) y_{i,t}(0) \qquad \text{and} \qquad \bar{y}_{i,-\tau}(1) = \frac{1}{\tau} \sum_{t \in \mathcal{W}_\tau} \mathbb{1}(t \geq 0) y_{i,t}(1), \\ \bar{y}_{i,\tau}(0) &= \frac{1}{\tau} \sum_{t \in \mathcal{W}_\tau} \mathbb{1}(t < 0) y_{i,t}(0) \qquad \text{and} \qquad \bar{y}_{i,\tau}(1) = \frac{1}{\tau} \sum_{t \in \mathcal{W}_\tau} \mathbb{1}(t \geq 0) y_{i,t}(1), \end{aligned}$$

where each unit has four potential outcomes, but we only observe two of them.

The conditions for using a randomization-based approach in a multiple-unit before-and-after setting are summarized in the assumption below.

Assumption 1 *There exists a window $\mathcal{W}_\tau$ of τ periods before and after the time of the intervention such that the following conditions hold.*

(i) $(y_{i,t}(0), y_{i,t}(1) : i = 1, 2, \cdots, n; t \in \mathcal{W}_\tau)$ *are non-stochastic.*

(ii) $\bar{y}_{i,-\tau}(0) = \bar{y}_{i,\tau}(0)$ *and* $\bar{y}_{i,-\tau}(1) = \bar{y}_{i,\tau}(1)$ *for all* $i = 1, 2, \cdots, n$.

(iii) $\mathbb{P}(\mathbf{D}_0 = \mathbf{d}) = 2^{-n}$ *for all* $\mathbf{d} \in \{0, 1\}^n$.

Assumption 1(i) allows for the deployment of Fisherian inference methods because the potential outcomes are fixed. Assumption 1(ii) rules out time trends in the potential outcomes, and it will be implausible when $\mathcal{W}_\tau$ includes an interval of time that is large relative to the scale of the outcome. In our application, a few days before or after the intervention is a very short time for crime to adjust, but one year before or after may be too wide because crime can change dramatically during long periods. Finally, Assumption 1(iii) implies that each unit has the same probability of being treated during all periods before $t = 0$ as it has of being treated during all periods after (and including) $t = 0$, and this probability is equal to 1/2. Although the real intervention was implemented at $t = 0$ for all units and was active only in the second half of the window $\mathcal{W}_\tau$, our assumption is that it could have been active in the first half of $\mathcal{W}_\tau$.

3.2 Methods

Given our assumptions, our parameter of interest is the average treatment effect in the post-intervention period

$$\theta_\tau = \frac{1}{n} \sum_{i=1}^{n} \big(\bar{y}_{i,\tau}(1) - \bar{y}_{i,\tau}(0)\big),$$

which we estimate with the average of the difference between the (average) outcome in the post-intervention period and the (average) outcome in the pre-intervention period,

$$\widehat{\theta}_\tau = \frac{1}{n}\sum_{i=1}^{n} \bar{y}_{i,\tau} - \frac{1}{n}\sum_{i=1}^{n} \bar{y}_{i,-\tau}$$

For $j = 1, 2, \ldots, 2n$, we define the quantities

$$D_j = \begin{cases} D_{j,-1} & \text{if } j = 1, 2, \ldots, n \\ D_{j-n,0} & \text{if } j = n+1, n+2, \ldots, 2n \end{cases},$$

$$\bar{Y}_{j,\tau} = \begin{cases} \bar{y}_{j,-\tau} & \text{if } j = 1, 2, \ldots, n \\ \bar{y}_{j-n,\tau} & \text{if } j = n+1, n+2, \ldots, 2n \end{cases},$$

and

$$\bar{Y}_{j,\tau}(k) = \begin{cases} \bar{y}_{j,-\tau}(k) & \text{if } j = 1, 2, \ldots, n \\ \bar{y}_{j-n,\tau}(k) & \text{if } j = n+1, n+2, \ldots, 2n \end{cases}$$

for $k = 0, 1$. Using these definitions, we can re-write the estimator as

$$\begin{aligned}
\widehat{\theta}_\tau &= \frac{1}{n}\sum_{i=1}^{n}\sum_{g\in\{-\tau,\tau\}} \big(2D_{i,g} - 1\big)\bar{y}_{i,j} = \frac{1}{n}\sum_{j=1}^{2n}\big(2D_j - 1\big)\bar{Y}_{j,\tau} \\
&= \frac{1}{n}\sum_{j=1}^{2n}\big(2D_j - 1\big)\big(D_j\bar{Y}_{j,\tau}(1) + (1 - D_j)\bar{Y}_{j,\tau}(0)\big) \\
&= \frac{1}{n}\sum_{j=1}^{2n}\big(D_j\bar{Y}_{j,\tau}(1) - (1 - D_j)\bar{Y}_{j,\tau}(0)\big),
\end{aligned}$$

and therefore under Assumption 1 we have

$$
\begin{aligned}
\mathbb{E}[\widehat{\theta}_\tau] &= \frac{1}{n}\sum_{j=1}^{2n}\big(\mathbb{E}[D_j]\bar{Y}_{j,\tau}(1) - (1-\mathbb{E}[D_i])\bar{Y}_{j,\tau}(0)\big)\\
&= \frac{1}{2n}\sum_{j=1}^{2n}\big(\bar{Y}_{j,\tau}(1) - \bar{Y}_{j,\tau}(0)\big)\\
&= \frac{1}{2n}\sum_{i=1}^{n}\big(2\bar{y}_{i,\tau}(1) - 2\bar{y}_{i,\tau}(0)\big)\\
&= \theta_\tau
\end{aligned}
$$

where the first line uses the assumption that the potential outcomes are non-random, the second line uses $\mathbb{E}[D_i] = 1/2$, and the last line relies on the assumption that there are no time effects during time captured by $\mathcal{W}_\tau$. Thus, unbiased estimation of the average effect in the window $\mathcal{W}_\tau$ can be implemented by taking the average across units of the (average) outcome in the post-intervention period and subtracting from it the average across units of the (average) outcomes in the pre-intervention period.

For statistical inference, we deploy Fisherian methods. Because the same units are observed in every period, any window $\mathcal{W}_\tau$ contains the same units above and below the cutoff. This induces a paired data structure where, for a given τ, each unit i has two average outcome observations, $(\bar{y}_{i,-\tau}, \bar{y}_{i,\tau})$ for $i = 1, 2, \ldots, n$, corresponding to the pre-intervention and post-intervention periods, respectively. In each of these outcome pairs, exactly one outcome is treated, and exactly one is a control, inducing a negative correlation between the two treatments of every unit in the sample.

We consider the sharp null hypothesis that the (average) potential outcome under treatment in the τ-length post-cutoff period is equal to the (average) potential outcome under control in the τ-length post-cutoff period for every unit,

$$
\mathsf{H}_{0,\tau} : \bar{y}_{i,\tau}(1) = \bar{y}_{i,\tau}(0) \qquad \text{for all } i = 1, 2, \ldots, n.
$$

Under the sharp null hypothesis $\mathsf{H}_{0,\tau}$ and Assumption 1, the only source of randomness is the randomness in $\mathbf{D}_0$, which is fully characterized.

We define a test statistic $S = S(\mathbf{D}_0, \mathbf{Y}_\tau)$, where $\mathbf{Y}_\tau = \left(\bar{Y}_{j,\tau} : j = 1, 2, \cdots, 2n\right)$. The vectors of (average) potential outcomes under control and treatment, respectively, are defined analogously: $\mathbf{Y}_\tau(0) = \left(\bar{Y}_{j,\tau}(0) : j = 1, 2, \cdots, 2n\right)$ and $\mathbf{Y}_\tau(1) = \left(\bar{Y}_{j,\tau}(1) : j = 1, 2, \cdots, 2n\right)$. Under the sharp null hypothesis, $S(\mathbf{D}_0, \mathbf{Y}_\tau) = S(\mathbf{D}_0, \mathbf{Y}_\tau(0))$, and its distribution is completely determined by the randomization distribution of $\mathbf{D}_0$ given by Assumption 1(iii).

We build the set $\mathcal{D}_0 = \{0, 1\}^n$ of all possible treatment assignments $\mathbf{D}_0$ by considering all the treatment assignment vectors of size $2n$ where in each of the n pairs, exactly one period is assigned to treatment and exactly one period is assigned to control. There are 2^n such pairs, and each one is equally likely. We denote the observed, realized value of the treatment assignment vector by $\mathbf{d}_{\mathtt{obs}}$, and the observed value of test statistic as $s_{\mathtt{obs}} = S(\mathbf{d}_{\mathtt{obs}}, \mathbf{Y}_\tau)$. The exact two-sided p-value is thus given by

$$p = \mathbb{P}\left[|S(\mathbf{D}_0, \mathbf{Y}_\tau)| \geq |s_{\mathtt{obs}}|\right] = \sum_{\mathbf{d} \in \mathcal{D}_0} \mathbb{1}(|S(\mathbf{d}, \mathbf{Y}_\tau)| \geq |s_{\mathtt{obs}}|) \frac{1}{2^n}.$$

For even moderate values of n, 2^n is too large a number, so complete enumeration is typically unfeasible. In our application, $n = 62$. For implementation, we compute the p-value by simulation. In each simulation, we construct one of the 2^n possible treatment assignment vectors by, for each unit, assigning the post-intervention period to treatment and the pre-intervention period to control with probability 1/2, and computing the value of the test statistic. We repeat this simulation 10,000 times, and then calculate the share of the 10,000 test statistics that have absolute value equal to or higher than $|s_{\mathtt{obs}}|$.

This approach can be generalized to a joint test of multiple hypotheses, which in our context is helpful to test hypotheses in many windows around the event time simultaneously, rather than performing one test per window and face the possibility of false discoveries. The sharp null hypothesis can be made for L different windows, $\mathsf{H}_{0,\tau} : \bar{y}_{i,\tau}(1) =$

$\bar{y}_{i,\tau}(0)$ for all $i = 1, 2, \ldots, n$, for $\tau = 1, 2, \ldots L$. Then, a test statistic can be calculated for each hypothesis in a specific window following the steps above. Letting $s_\tau = S(\mathbf{D}_0, \mathbf{Y}_\tau)$ be the test statistic associated with window τ, we can combine the statistics for L windows, $s_1, s_2, \ldots, s_L$ into a single joint statistic such as the mean, maximum, or Hotelling's T^2, and calculate a Fisherian p-value by simulation by repeating the procedure for different realizations of the treatment vector. This procedure can be used to test the joint hypothesis that the effect was zero in all windows $\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_L$, as we illustrate in the following section.

3.3 Discussion

One key advantage of our approach is that it makes it possible to reduce extrapolation to a minimum by considering the smallest discrete time interval before and after the intervention. By choosing θ_τ and $\mathsf{H}_{0,\tau}$ with $\tau = 1$, researchers can compare the average outcome in the period immediately before the intervention to the average outcome in the exact period when the intervention goes into effect. For example, if time is measured in days, θ_1 and $\mathsf{H}_{0,1}$ capture the effect of the intervention on the day of the reform relative to the day immediately before. These parameters minimize extrapolation as much as it is possible given the data and, therefore, are least susceptible to time trends.

Whether θ_1 is expected to be non-zero and $\mathsf{H}_{0,1}$ is expected to be rejected depends on the application. In some cases, treatment effects are expected immediately after reform, while in others, they require some time to mature after the policy is implemented. By considering θ_1 and $\mathsf{H}_{0,1}$, and comparing them to θ_τ and $\mathsf{H}_{0,\tau}$ for $\tau > 1$, researchers can learn about how long it takes for the intervention to affect the outcome.

Assumption 1(ii) requires that there be no time effects, a strong condition. If another variable other than the intervention changes discontinuously between $t = -1$ and $t = 0$, this assumption will not hold. In our crime application, this could happen, for example, if $t = -1$ fell on a Sunday and $t = 0$ on a Monday, and crime were higher during the weekend (see Prieto Curiel, 2021, for evidence and discussion regarding the temporal concentration of crime). In fact, this day-of-the-week effect is very common in event studies. While application-specific, this discussion indicates that $\tau = 1$ could be too small in applications with high-frequency data, such as crime or weather data at a highly disaggregated levels (e.g., daily data). For this reason, we also consider one ($\tau = 7$) and two weeks ($\tau = 14$) in our empirical application, and all windows between days 1 and 14 in joint tests.

Although a direct test of Assumption 1(ii) is fundamentally unfeasible, our framework allows us to study its plausibility by designing falsification tests based on prior periods. In our application, the intervention occurred on November 1st, 2017, but we have data on our outcome variable from several years prior. Our falsification analyses study the effect of the intervention in the years before and after the true intervention date, setting the cutoff to the artificial values $t_1, t_2, \ldots, t_K$. These values can be chosen in several ways, depending on the application. For example, researchers can set the artificial cutoff equal to the same calendar month and day as the true cutoff, but in a different year; equal to the same calendar day but on a different month or week from $t = 0$; or equal to the same day of the week but on a different calendar month or year, etc. We illustrate this procedure in the following section.

4 Results

The new adversarial system came into effect nationwide at 12 a.m. on November 1st, 2017. The cornerstone of our research design is the ability to compare daily police reports for each of Montevideo's 62 neighborhoods before and after the intervention. By focusing on very small time windows around the intervention, we ensure that the assumption of no-time trends remains plausible. Our smallest window, $\mathcal{W}_1$, includes just two days and compares the offenses that occurred on October 31st, 2017, to the offenses that occurred on November 1st, 2017. Since our main outcome of interest is the total number of police reports, interpreting our results as effects on crime requires assuming that the propensity to report crimes did not change due to the new policy. This is a standard assumption when working with this type of data, and we believe it is plausible in our case.

Table 2 reports the results from our randomization-based approach for our main outcome of interest, total crime reports. The table reports the estimated average effect, $\widehat{\theta}_\tau$, and the randomization-based p-value corresponding to the two-sided test of the sharp null hypothesis $\mathsf{H}_{0,\tau} : \bar{y}_{i,\tau}(1) = \bar{y}_{i,\tau}(0)$, for three windows of different length around the time of the intervention, given by different values of τ—1 day, 7 days, and 14 days before/after the adoption of the new code.

Table 2: Effects of CCP Reform for Different Windows Around Event Time
(crime by neighborhood; event time: 12 a.m. on November 1st, 2017)

	Estimation of average effect		Randomization-based inference
	$\widehat{\theta}_\tau$	Pre-intervention mean	P-value for $\mathsf{H}_{0,\tau} : \bar{y}_{i,\tau}(1) = \bar{y}_{i,\tau}(0)$
1 day before/after event ($\tau = 1$)	0.839	5.306	0.127
7 days before/after event ($\tau = 7$)	0.403	4.892	0.016
14 days before/after event ($\tau = 14$)	0.298	5.065	0.022

Sample size is 62 neighborhoods in Montevideo, Uruguay, each observed before and after the event time.
Outcome is the daily average number of crimes reported to police in each neighborhood.
$\widehat{\theta}_\tau$ is the average treatment effect after the event time; the untreated mean is $\frac{1}{n}\sum_{j=1}^{2n} \bar{Y}_{j,\tau}(1 - D_j)$.
Randomization-based p-value calculated based on 10,000 simulations.

The first row reports the results of the analysis conducted in the smallest possible window: one day before and one day after the adoption of the new CCP. On the day before the new code was adopted, the average number of crime reports per neighborhood was 5.306; this average increased by 0.839 to 6.145 the day after the intervention, but the difference is not distinguishable from zero according to our randomization-based test of $\mathsf{H}_{0,\tau}$ (p-value is 0.127).

In contrast, the second row shows that, when comparing the seven days before to the seven days after the intervention, the switch from an inquisitorial to an adversarial criminal procedure seems to have resulted in a statistically significant increase in the total number of crimes reported (sharp null hypothesis is rejected with p-value 0.016). The average of the neighborhood-level average daily crimes reported in the seven days before the intervention is 4.892, and this increases to 4.892+0.403=5.295. Since there are 62 neighborhoods, this corresponds to an increase from approximately $4.892 \times 62 \times 7 \approx 2,123$ reports in the seven days before the intervention to $(4.892 + 0.403) \times 62 \times 7 \approx 2,298$, an increase of about 8.2 percent. The pattern is similar for a window within 14 days of the start of the intervention (sharp null rejected with p-value 0.022).

In addition to analyzing the three windows (of 1, 7, and 14 days) discussed above, we also investigate the treatment effects in many windows simultaneously, accounting for multiple hypothesis testing within our randomization inference framework. Table 3 shows the results of joint randomization-based tests for two collections of windows: all windows between day 1 and day 7, $\mathcal{W}_1, \ldots, \mathcal{W}_7$, and all windows between day 1 and day 14, $\mathcal{W}_1, \ldots, \mathcal{W}_{14}$. The Fisherian p-values are calculated using three different test statistics: the maximum of the average difference in each window, Hotelling's T^2 statistic using the vector of average differences in each window with their respective covariance matrices, and the mean of the average difference in each window. The results show that, when jointly considering the treatment effect in the first seven or the first fourteen windows around the event time, we reject the null hypothesis of no treatment effect in most cases at 6% or 7% significance level. For example, according to the maximum average difference across all windows, the hypothesis of no effect has p-value 0.065 (for the first seven windows) and 0.67 (for the first fourteen windows), and according to Hotelling's T^2, these p-values are, respectively, 0.051 and less than 0.001. Thus, similarly to the individual window analyses reported in Table 2, these joint hypothesis tests suggest evidence of a significant effect very soon after the reform is introduced.

Table 3: Joint Inference for Effects of CCP Reform for Different Windows Around Event Time
(crime by neighborhood; event time: 12 a.m. on November 1st, 2017)

Randomization-based inference Joint test of hypotheses $\mathsf{H}_{0,\tau} : \bar{y}_{i,\tau}(1) = \bar{y}_{i,\tau}(0)\ \forall i$, for $\tau = 1, \ldots, K$	Max		Hotelling		Mean	
	Statistic	p-val	Statistic	p-val	Statistic	p-val
First 7 windows (1–7 days before/after event, $K = 7$)	0.839	0.065	13.102	0.051	0.292	0.172
First 14 windows (1–14 days before/after event, $K = 14$)	0.839	0.067	38.057	<0.001	0.302	0.054

Sample size is 62 neighborhoods in Montevideo, Uruguay, each observed before and after the event time.
Outcome is the daily average number of crimes reported to police in each neighborhood.
Randomization-based p-value calculated based on 10,000 simulations.

4.1 Assessing Assumptions: Falsification Analysis

In order to interpret the increases in crime reported in the prior section as the causal effect of the implementation of the new CCP, the assumption of no time trends must hold. Although this assumption is fundamentally untestable, we present results from a falsification analysis that estimates effects and p-values for our main outcome but moves the time of intervention.

Table 4 presents results with the event time artificially set to midnight on November 1st for the years 2015, 2016, and 2018. Theses falsification results show that there are no distinguishable differences in crime reports around artificial intervention times in the same windows considered in Table 2: the sharp null hypothesis of no effect for any unit fails to be rejected in all cases except for $\tau = 1$ in 2016, but this corresponds to a negative estimated effect, $\widehat{\theta}_\tau = -0.855 < 0$. Considering that November 1st, 2017, was a Wednesday, we can also evaluate the robustness of the results in Table 4 by setting the event time to the first Wednesday of November in 2015, 2016, and 2018. Once again, the estimated effects for these alternative dates are not statistically significant and most are considerably smaller than those seen for the correct date.

The falsification results in Tables 4-5 remain robust when we consider alternative values

Table 4: Effects of CCP Reform for Different Windows Around *Placebo* Event Times

	Event Time: 12 a.m. on November 1st					
	2015		2016		2018	
	$\widehat{\theta}_\tau$	P-value	$\widehat{\theta}_\tau$	P-value	$\widehat{\theta}_\tau$	P-value
1 day before/after event ($\tau = 1$)	0.048	0.962	-0.855	0.036	0.661	0.164
7 days before/after event ($\tau = 7$)	0.090	0.550	-0.041	0.819	0.002	0.995
14 days before/after event ($\tau = 14$)	0.015	0.913	0.035	0.743	-0.113	0.401

Sample size is 62 neighborhoods in Montevideo, Uruguay, each observed before and after the event time.
Outcome is the daily average number of crimes reported to police in each neighborhood.
$\widehat{\theta}_\tau$ is the average treatment effect after the event time; the untreated mean is $\frac{1}{n}\sum_{j=1}^{2n}\bar{Y}_{j,\tau}(1-D_j)$.
Randomization-based p-value calculated based on 10,000 simulations.

Table 5: Effects of CCP Reform for Different Windows Around *Placebo* Event Times

	Event Time: 12 a.m. on the First Wednesday of November					
	2015		2016		2018	
	$\widehat{\theta}_\tau$	P-value	$\widehat{\theta}_\tau$	P-value	$\widehat{\theta}_\tau$	P-value
1 day before/after event ($\tau = 1$)	-0.194	0.698	-0.065	0.898	0.129	0.793
7 days before/after event ($\tau = 7$)	-0.106	0.522	0.046	0.762	-0.081	0.666
14 days before/after event ($\tau = 14$)	-0.015	0.921	0.142	0.215	0.046	0.712

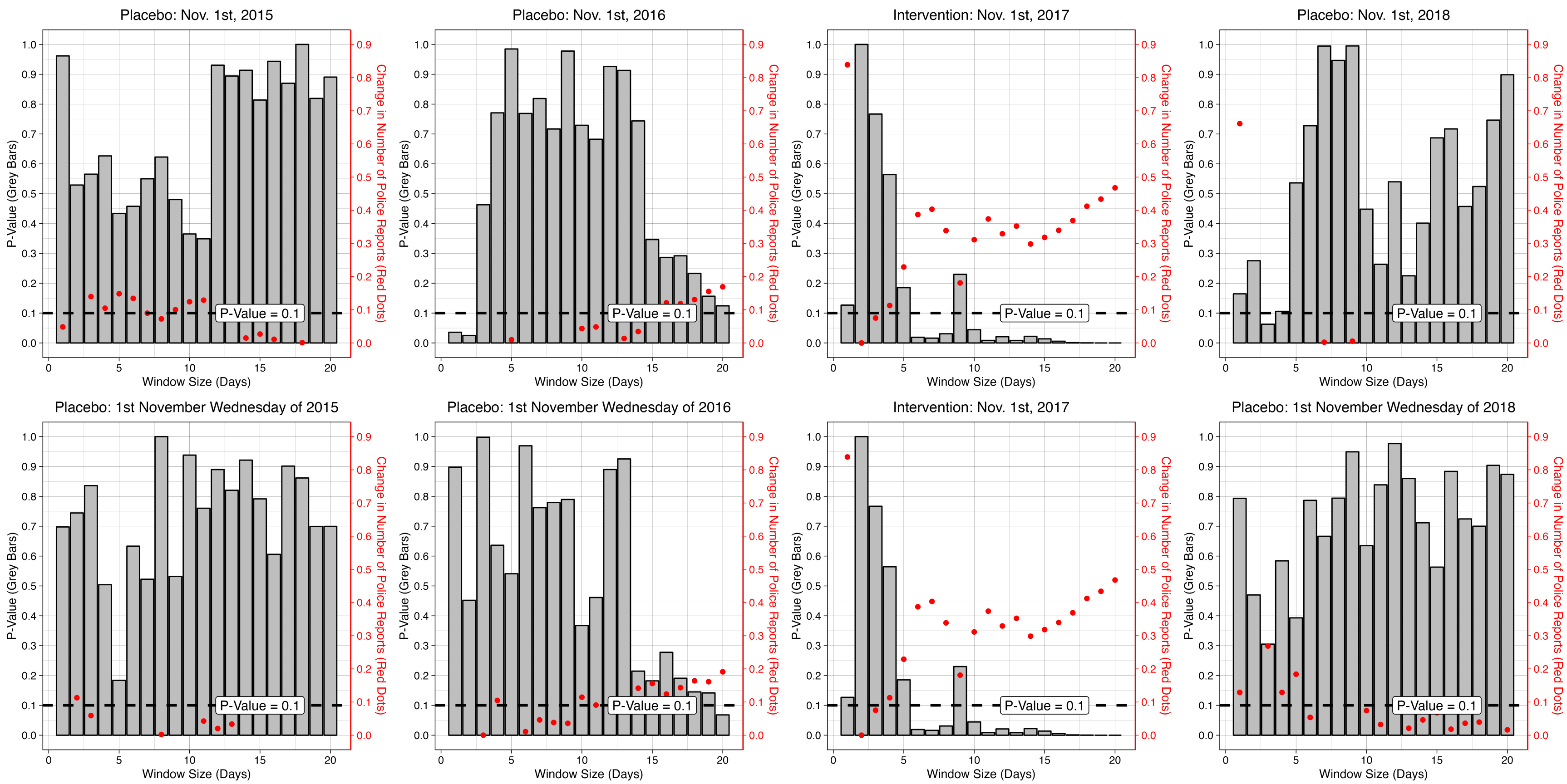
Sample size is 62 neighborhoods in Montevideo, Uruguay, each observed before and after the event time.
Outcome is the daily average number of crimes reported to police in each neighborhood.
$\widehat{\theta}_\tau$ is the average treatment effect after the event time; the untreated mean is $\frac{1}{n}\sum_{j=1}^{2n}\bar{Y}_{j,\tau}(1-D_j)$.
Randomization-based p-value calculated based on 10,000 simulations.

of τ. Figure 2 shows the estimates, $\widehat{\theta}_\tau$, and the randomization-based p-values for all values of τ between 1 and 20 (not just 1, 7, and 14, as reported in the tables), including the different placebo dates considered as time of intervention in the before-and-after estimations reported above. Not only is the average treatment effect statistically significant (grey bars) at 5% level or below for almost all values of τ greater than 5 when we use the time of the real event, but also the average treatment effects (red dots) are always positive and well above the values recorded for the same period in each of the placebo event time analyses. In contrast, in the falsification analyses, p-values are generally above 10% and, in several cases, the effects are negative (note that the red dots correspond only to positive estimates; the negative ones would be below the bottom horizontal line and are omitted).

In sum, our falsification tests suggest that the assumptions of no time trends or confounders are plausible within a small window around the implementation date of the new CCP for the total number of police reports in Montevideo, as the treatment effects that we see in close windows around the time of the actual intervention are not seen when we consider windows of the same length around the same date/day in prior years. And because our analysis is based on the exact same units before and after the time of the event in all falsification and actual analyses, our results cannot be explained by any time-constant unit-specific characteristics or fixed effects.

In the following section, we discuss the results for the main types of crime (theft, robbery, and domestic violence), exploring heterogeneities that might help identify possible mechanisms behind these outcomes.

Figure 2: Effects of CCP Reform for Different Windows Around *Placebo* Event Times



4.2 Potential Mechanisms

We document an average increase of about 25 ($\approx 0.403 \times 62$) police reports per day in the week immediately following the implementation of the new CCP, suggesting that the change from an inquisitorial to an accusatory CCP had immediate consequences on crime. Even though a code of criminal procedure neither defines the recognized offenses nor sets the corresponding penalties, our evidence suggests that a change in the adjudication process of criminal law unexpectedly decreased public safety.

Auxiliary evidence indicates that this increase in the number of police reports was not anticipated as a potential consequence of the implementation of the new CCP in Uruguay. Our findings are consistent with those of Zorro Medina et al. (2020), who document that a similar procedural reform in Colombia caused an increase in overall crime of 22%-34%. While our result is smaller in magnitude, our design identifies causal effects within narrow time periods. To the best of our knowledge, our paper is the first to document that the transition from an inquisitorial system to an adversarial system can lead to sudden and almost immediate increases in the number of police reports.

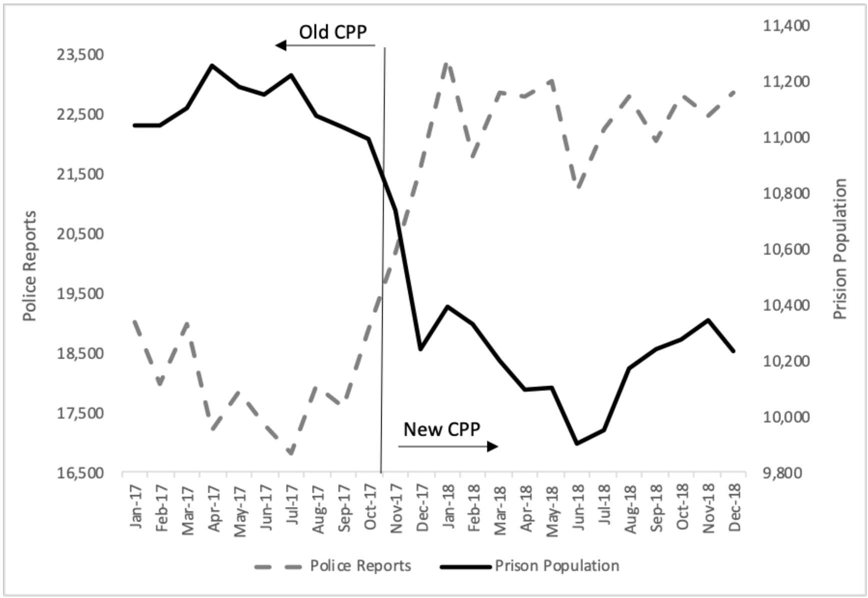
These results could have been caused by more lenient punishments and a lower probability of being caught and punished (Becker, 1968). First, the new adversarial system seems to have resulted in less severe penalties than the old inquisitorial one, as it introduced procedural alternatives to oral trials that are associated with lighter sentences. Procedural alternatives, which were employed in 90% of solved cases, help lawyers and defendants resolve their cases faster and face a more lenient sentence while allowing public prosecutors to avoid lengthy and demanding criminal trials. Most of these solved cases are the result of the use of the abbreviated process (i.e., a type of plea bargaining introduced by the new criminal procedure) that implies an agreement between defendant and prosecutor whereby the former pleads guilty to a particular charge in return for a more lenient sentence from the prosecutor. Moreover, preventive detention (i.e., detention while the process lasts until there is a sentence) ceased to be the norm, in contrast to the inquisitorial system, which might have created the expectation of a less severe punishment. Second, the new legal system might have resulted in a lower probability of conviction. Under the new system, prosecutors faced a significant increase in their workload, as they were now exclusively responsible for leading the investigation and carrying the evidence to judges. Moreover, the police must conduct investigations under new supervisors (prosecutors instead of judges) and different rules, which could have created coordination challenges.

Figure 3 provides some evidence in line with the hypothesis that the reform affected crime incentives. For the first time in more than a decade, the average number of people in prisons decreased in 2018, from 11,005 to 10,179 inmates (a 7.5% decrease). According to Figure 3(a), this sharp reduction in prison population starts at the same time as the implementation of the new CCP and contrasts with the sudden rise in the number of crimes reported to police that we documented in the prior section. Meanwhile, the number of criminal indictments (i.e., formal accusations made by public prosecutors) also experienced a strong month-to-month decrease in November 2017, from 1,001 to 584 cases (i.e., a 42% reduction). Figure 3(b) illustrates the evolution of indictments relative to police reports. The average ratio for the first two months of the new CCP is 3.2% (i.e., November and December 2017), well below the corresponding figure of 5.9% for the rest of 2017 in the last months of the old CCP.

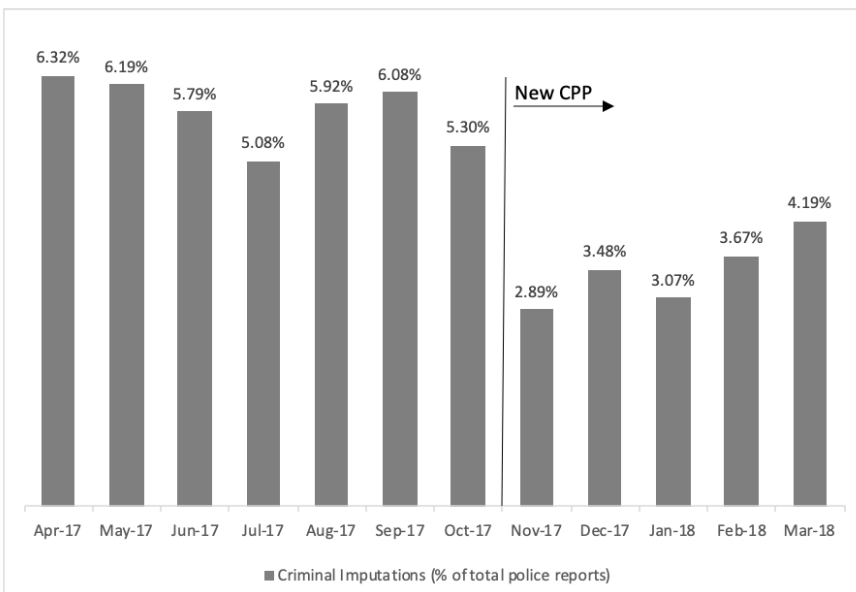
The above trends suggest that less severe crimes could have increased due to a change in crime incentives (i.e., less deterrence), and also could have responded to an increase in the number of active criminals (i.e., less incapacitation). The reduction in prison population might be explained by the new constraints imposed on the use of preventive detention by prosecutors. As mentioned above, preventive prison was used extensively under the inquisitorial system, in particular for cases of recidivism (55% of the cases). This was the norm for both property crimes (e.g., thefts) and violent crimes (e.g., robberies and domestic violence). However, under the new adversarial system, preventive detention is applicable only when there is sufficient evidence suggesting that the defendant might attempt to escape,

Figure 3: Police Reports, Prison Population and Criminal Imputations

(a) Police Reports and Prison Population



(b) Criminal Imputations



obstruct the investigation, or pose a risk to society (i.e., severe and typically violent crimes). Consequently, a convicted offender who re-offends by committing a lesser crime could be immediately released under the adversarial system, whereas under the old CCP, they would likely have served preventive detention.

To test the hypothesis that the reduction in the use of preventive detention might be be related to the increase in crime reports, we compare the immediate impact of the new CCP on the number of reported thefts to its impact on the number of reported robberies (i.e., violent thefts) and domestic violence (i.e., the most frequent crime against persons). If the effect is due to a more selective use of preventive detention, we should not observe any impact on domestic violence (this is a crime that is considered very severe by prosecutors under both procedural regimes), while the effects should be present for thefts and robberies—and should be higher for thefts than for robberies, as thefts are classified as non-violent and thus might avoid preventive detention.

Table 6 presents results for the most frequent violent crimes (robbery and domestic violence) and the most frequent non-violent crime (thefts), which together account for more than seventy percent of the crimes reported to the Uruguayan police in Montevideo. As expected, the number of domestic violence incidents reported to the police does not exhibit an immediate response to the change in CCP. In fact, negative point estimates are observed for $\tau = 1$ and $\tau = 7$ (sharp null hypothesis not significant). The local estimated effects for robberies also appear weak, with only one p-value below 0.10. This pattern suggests that the increase in the total number of police reports is not driven by the two most frequent violent crimes. The results for thefts suggest that the reform might have increased thefts immediately after implementation. However, these effects are not as strong as the effects for total crime reported in Table 2, which suggests that the decrease in preventive detention is likely only part of the story depicted by Figure 1. For instance, there is evidence that police officers may change their tactics when norms that restrict their actions are imposed, resulting in reductions of imprisonment rates (Hausman and Kronick, 2021). As it was previously discussed, the new CCP introduced substantial changes in the work of police officers, so this is a plausible channel.

In sum, we believe the increase in the number of police reports we observe is due to a combination of the reduction in incapacitation and the weakening of deterrence that was created by the change in CCP.

Table 6: Effects of CCP Reform for Different Types of Crime
(event time: 12 a.m. on November 1st, 2017)

	Property		Violent			
	Theft		Robbery		Domestic Violence	
	$\widehat{\theta}_\tau$	P-value	$\widehat{\theta}_\tau$	P-value	$\widehat{\theta}_\tau$	P-value
1 day before/after event ($\tau = 1$)	0.500	0.116	-0.258	0.211	-0.097	0.553
7 days before/after event ($\tau = 7$)	0.194	0.064	0.097	0.091	-0.035	0.490
14 days before/after event ($\tau = 14$)	0.150	0.164	0.021	0.672	0.020	0.582

Sample size is 62 neighborhoods in Montevideo, Uruguay, each observed before and after the event time.
Outcome is the daily average number of crimes reported to police in each neighborhood.
$\widehat{\theta}_\tau$ is the average treatment effect after the event time; the untreated mean is $\frac{1}{n}\sum_{j=1}^{2n} \bar{Y}_{j,\tau}(1 - D_j)$.
Randomization-based p-value calculated based on 10,000 simulations.

5 Conclusion

We presented a randomization-based framework to analyze before-and-after event studies where an intervention is given to all cross-sectional units at the same time, and all units are observed for several periods before and after the intervention. Our setup assumes that the units' potential outcomes are non-stochastic, and considers an assignment mechanism where each unit could have been treated either in the pre-intervention or in the post-intervention period with equal probability. Our approach is based on the assumption that the average potential outcomes under no intervention would have been the same in the period after the intervention, but we only make this assumption for a small window of time around the time of the intervention, which increases its plausibility. We also proposed a falsification approach to validate this assumption. Our methodology adds to the toolkit on program evaluation and causal inference (Abadie and Cattaneo, 2018; Rosenbaum, 2002b, 2010), in particular complementing more standard approaches based on linear panel data models with two-way fixed effects, trend, and other covariate adjustments, which rely on flexible parametric global modelling of the time series (Freyaldenhoven et al., 2019; Miller, 2023; Freyaldenhoven et al., Forthcoming). Incorporating covariate-adjustments to our local randomization inference methodology is possible (Rosenbaum, 2002a), but often unnecessary if the window chosen $\mathcal{W}_\tau$ is small.

We used our framework to study the effects of a reform to the code of criminal procedure introduced in Uruguay on November 1, 2017, which simultaneously implemented in the entire country. Our analysis of the $n = 62$ neighborhoods in Montevideo shows that the adoption of the new code increased total crime reports by about 8.2 percent in the week after the reform compared to the week immediately before the change from an inquisitorial to an accusatory code of criminal procedure. Using our approach to conduct statistical inferences in a small window around the time of the intervention, we show that there is strong evidence to distinguish this effect from zero. Moreover, falsification analyses show that the same effects indistinguishable from zero when the analysis is conducted in analogous windows of time in prior years, before the intervention is active.

References

Abadie, A., and Cattaneo, M. D. (2018), "Econometric Methods for Program Evaluation," *Annual Review of Economics*, 10, 465–503.

Abrams, D. S. (2011), "Is Pleading Really a Bargain?" *Journal of Empirical Legal Studies*, 8, 200–221.

Acemoglu, D., and Johnson, S. (2005), "Unbundling Institutions," *Journal of Political Economy*, 113, 949–995.

Atkins, R. A., and Rubin, P. H. (2003), "Effects of Criminal Procedure on Crime Rates: Mapping Out the Consequences of the Exclusionary Rule," *The Journal of Law and Economics*, 46, 157–179.

Becker, G. S. (1968), "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76, 169–217.

Bushway, S. D., and Redlich, A. D. (2012), "Is Plea Bargaining in the "Shadow of the Trial" a Mirage?" *Journal of Quantitative Criminology*, 28, 437–454.

Bushway, S. D., Redlich, A. D., and Norris, R. J. (2014), "An Explicit Test of Plea Bargaining in the "Shadow of the Trial"," *Criminology*, 52, 723–754.

Cattaneo, M. D., Frandsen, B., and Titiunik, R. (2015), "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate," *Journal of Causal Inference*, 3, 1–24.

Cattaneo, M. D., and Titiunik, R. (2022), "Regression discontinuity designs," *Annual Review of Economics*, 14, 821–851.

Cattaneo, M. D., Titiunik, R., and Vazquez-Bare, G. (2017), "Comparing Inference Approaches for RD Designs: A Reexamination of the Effect of Head Start on Child Mortality," *Journal of Policy Analysis and Management*, 36, 643–681.

Cook, T. D., Campbell, D. T., and Shadish, W. (2002), *Experimental and quasi-experimental designs for generalized causal inference*, Houghton Mifflin.

Dalla Pellegrina, L. (2008), "Court Delays and Crime Deterrence," *European Journal of Law and Economics*, 26, 267–290.

Duŝek, L. (2015), "Time to Punishment: The Effects of a Shorter Criminal Procedure on Crime Rates," *International Review of Law and Economics*, 43, 134–147.

Fandiño, M., and González Postigo, L. (2020), "Adversarial Criminal Justice in Latin America: Comparative Analysis and Proposals," *Justice Studies Center of the Americas (JSCA)*.

Frazier, A., Shockley, K., Keenan, J. M., Wilford, M. M., and Gonzales, J. E. (2018), "When a Plea is no Bargain at All: Comparing Sentencing Outcomes for Massachusetts Defendants in Non-Sexual and Sexual Crimes," *Alb. L. Rev.*, 82, 775.

Freyaldenhoven, S., Hansen, C., Pérez, J. P., and Shapiro, J. M. (Forthcoming), "Visualization, identification, and estimation in the linear panel event-study design," *Advances in Economics and Econometrics: Twelfth World Congress*.

Freyaldenhoven, S., Hansen, C., and Shapiro, J. M. (2019), "Pre-event trends in the panel event-study design," *American Economic Review*, 109, 3307–3338.

Hausman, C., and Rapson, D. S. (2018), "Regression discontinuity in time: Considerations for empirical applications," *Annual Review of Resource Economics*, 10, 533–552.

Hausman, D., and Kronick, D. (2021), "When Police Sabotage Reform by Switching Tactics," *SSRN*.

Ho, D. E., and Imai, K. (2006), "Randomization Inference with Natural Experiments: An Analysis of Ballot Effects in the 2003 Election," *Journal of the American Statistical Association*, 101, 888–900.

Imbens, G. W., and Rosenbaum, P. (2005), "Robust, accurate confidence intervals with a weak instrument: Quarter of birth and education," *Journal of the Royal Statistical Society, Series A*, 168, 109–126.

Institute for Crime & Justice Policy Research (2021), *The World Prison Brief*, https://www.prisonstudies.org/.

Kang, H., Peck, L., and Keele, L. (2018), "Inference for instrumental variables: a randomization inference approach," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 181, 1231–1254.

Kronick, D. (2019), "The Legal Origins of State Violence," *Unpublished Manuscript*.

Langer, M. (2007), "Revolution in Latin American Criminal Procedure: Diffusion of Legal Ideas from the Periphery," *The American Journal of Comparative Law*, 55, 617–676.

Listokin, Y. (2007), "Crime and (with a Lag) Punishment: The Implications of Discounting for Equitable Sentencing," *Am. Crim. L. Rev.*, 44, 115.

Maier, J., and Struensee, E. (2000), "Introducción: Las Reformas Procesales Penales en América Latina," *J. Maier, K. Ambos, & J. Woischnik (Edits.), Las Reformas Procesales Penales en América Latina*, 17–34.

Miller, D. L. (2023), "An introductory guide to event study models," *Journal of Economic Perspectives*, 37, 203–230.

Palumbo, I. (2018), "Para Bonomi, el nuevo código trajo aparejado contradicciones entre jueces, fiscales y policías," *Semanario Crónicas*, July 27th.

Prieto Curiel, R. (2021), "Weekly Crime Concentration," *Journal of Quantitative Criminology*, 1–28.

Rosenbaum, P. R. (2002a), "Covariance adjustment in randomized experiments and observational studies," *Statistical Science*, 17, 286–327.

Rosenbaum, P. R. (2002b), *Observational Studies* (2nd ed.), New York: Springer.

Rosenbaum, P. R. (2007), "Interference between units in randomized experiments," *Journal of the american statistical association*, 102, 191–200.

——— (2010), *Design of Observational Studies*, Vol. 10, Springer.

Soares, Y., and Sviatschi, M. M. (2010), "Does Court Efficiency Have a Deterrent Effect on Crime? Evidence for Costa Rica," *Unpublished Manuscript*.

Solomita, M. (2019), "Fiscales Agobiados: Turnos Interminables, Licencias por Estrés y Jubilaciones Anticipadas," *Diario El País*, Suplemento Qué Pasa, May 18th.

Ulmer, J. T., and Bradley, M. S. (2006), "Variation in Trial Penalties among Serious Violent Offenses," *Criminology*, 44, 631–670.

Ulmer, J. T., Eisenstein, J., and Johnson, B. D. (2010), "Trial Penalties in Federal Sentencing: Extra-Guidelines Factors and District Variation," *Justice Quarterly*, 27, 560–592.

Zorro Medina, A., Acosta, C., and Mejia, D. (2020), "The Unintended Consequences of the U.S. Adversarial Model in Latin American Crime," *SSRN*.