

Formal Employment and Organized Crime: Regression Discontinuity Evidence from Colombia

Gaurav Khanna* Carlos Medina†
Anant Nyshadham‡ Jorge Tamayo§

April 9, 2022

Abstract

Canonical models of entry into crime emphasize occupational sorting on economic incentives. We attempt to isolate the occupation-choice dimension of criminal participation responses to disincentives for formal employment. We link administrative socioeconomic microdata with the universe of arrests in Medellín over a decade, and exploit exogenous variation in formal-sector employment around a socioeconomic-score cutoff, below which individuals receive benefits if not formally employed. We model the various mechanisms by which the policy variation we study could affect both idiosyncratic and occupational criminality. Regression discontinuity estimates confirm this policy unintentionally reduced formal-sector employment and generated a corresponding increase in arrests associated with criminal enterprise activity. Consistent with an occupational choice interpretation as modeled, we find no effects on crimes unlikely to be associated with organized entities, such as crimes of impulse or opportunity. Effects on arrests are strongest in neighborhoods with more opportunities to join criminal enterprises.

Keywords: *occupational choice, criminal enterprise, crime, informality, Colombia*

JEL Codes: J24, J46, K42

*University of California – San Diego, gakhanna@ucsd.edu, gauravkhanna.info

†Banco de la Republica de Colombia, cmedindu@banrep.gov.co

‡University of Michigan, BREAD, NBER, JPAL & Good Business Lab, nyshadha@umich.edu, anantnyshadham.com

§Harvard University, Harvard Business School, jtamayo@hbs.edu, jorge-tamayo.com

We thank Nicolas Arturo Torres Franco and Cristian Chica for excellent research assistance. We thank seminar participants at the NBER Summer Institute, UCSD, Michigan (H2D2), Rosario (Bogota), SoCCAM, PacDev, Empirical Studies of Conflict, Barcelona GSE Summer Forum, Economics of Risky Behavior (Bologna), Colombian Central Bank (Bogota), Minnesota (MWIEDC), Hawaii, American Economic Association (ASSAs), Niskanen Center, and the Inter American Development Bank (AL CAPONE) for feedback, and Achyuta Adhvaryu, Jorge Aguero, Prashant Bharadwaj, Chris Blattman, Charles Brown, Michael Clemens, Gordon Dahl, Rafael DiTella, Gordon Hanson, Mauricio Romero, Emilia Tjernstrom, Santiago Tóbon, Juan Vargas, Mauricio Vilamizar and Jeff Weaver for insightful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research, and do not commit Banco de la República or its Board of Directors.

1 Introduction

Many countries, particularly across the developing world and in much of Latin America, are plagued by coincident high degrees of informality in the labor market and criminal activity often controlled by organized enterprises (Arteaga, 2019; Brown and Velasquez, 2017; Buonanno and Vargas, 2018; Chimeli and Soares, 2017; Dell et al., 2018; DiTella et al., 2010; Sviatschi, 2018; Tobón, 2020). Classic models of entry into crime contend that individuals rationally weigh the expected costs and benefits of engaging in criminal activity relative to alternatives, primarily legitimate employment in the labor market (Becker, 1968; Ehrlich, 1973). If individuals are (perhaps unintentionally) disincentivized from participating in formal employment in these settings, will they be more likely to engage in criminal enterprise activity?

Some of the best evidence of individual-level criminal participation decisions comes from responses to interventions that raise the human capital of individuals via skills training (Blattman and Annan, 2015) or work experience (Davis and Heller, 2020; Gelber et al., 2016; Heller, 2014; Schochet et al., 2008). A few studies have focused on current criminals and recently released prisoners (Blattman et al., 2017; Cook et al., 2015; Tobón, 2020). However, the variation in human capital leveraged in these studies could change returns to many, if not all, types of economic activities (formal, informal legitimate, and criminal). And so, this variation may not specifically recover the degree of sorting into criminal occupations in response to incentives for formal employment.

A related literature identifies criminal responses to job destruction (Britto et al., 2020; Khanna et al., 2020) and legal restrictions to formal employment (Mastrobuoni and Pinotti, 2015; Pinotti, 2017). These studies come close to answering our question of interest. Yet, both of these sources of variation generate financial necessity or desperation which can lead to temporary or idiosyncratic criminality, not necessarily reflective of an occupational choice. Job destruction and legal restrictions can impact criminality through depression and stigma as well. Furthermore, legal restrictions to formal employment in particular (though job destruction as well by way of impacts on subsequent job search) limit the degree to which individuals can actually choose formal employment, and therefore do not yield ideal variation for studying occupational *choice* between formal employment and criminal enterprise activity.

In this study, we attempt to isolate the occupational choice dimension of criminal participation responses to disincentives for formal employment. We use rich administrative data between 2002 and 2013 from Medellín, Colombia to test the relationship between formal employment and participation in crime at the individual level. In this empirical context, where informality is common and criminal enterprise activity abounds, financially dissuading individuals (unintentionally) from engaging in formal employment could drive some to crime as their most lucrative option outside of the formal sector. Exploiting a discontinuity in the incentives to remain informal in Colombia, we investigate individual-level occupational sorting into crime.

We develop a model of occupational sorting between formal and informal employment with both legitimate and illegitimate (but lucrative) opportunities in the informal sector, and incorporate the specific policy variation we attempt to study in this context. This intuition behind participation in criminal enterprise activity as a long-term occupational choice is consistent with recent evidence on gang recruitment and career progression from the Latin American context (Sviatschi, 2018) in general, and specifically from Medellín (Blattman et al., 2018). The model also includes the choice to engage in idiosyncratic crimes which are not economically motivated separate from the occupational choice. We derive testable hypotheses from the model to illuminate various mechanisms by which the policy variation can induce changes in criminality, and motivate placebo tests using non-occupational crimes.

The Colombian government provides health benefits to all residents of households that have a socio-economic score (known as the *Sisben* score) below a certain threshold. Formal employment makes the individual and their dependents ineligible for this program, raising the relative benefits to informality as an unintended consequence of the eligibility criteria for this program. Those formally employed are automatically taxed a fraction of their wages to avail of comparable benefits. Accordingly, the usual benefits to formal employment (e.g., higher wages, job security, legal protections) are at least partially offset by the increased cost of health care coverage for those below the cutoff, who would be eligible for free coverage by the government if they were not formally employed. The importance of these incentives in our context is emphasized by the near-complete health care coverage in the population, despite costs representing large proportions of income for many households.

Using a regression discontinuity design, we find that the policy-induced a roughly 4 percentage point lower formal employment rate at the margin, consistent with estimates from previous studies (Camacho et al., 2014). These same individuals are arrested at higher rates for crimes “likely associated with criminal enterprises” (LACE). At the RD cutoff, we find a 0.45 percentage point rise in LACE violent crimes (e.g., firearms trafficking, homicide), a roughly 0.66 percentage point rise in LACE property crimes (e.g., car theft), and a less precisely estimated 0.1 percentage point rise in LACE drug crimes (e.g., cocaine and heroin distribution). The impact on LACE violent crimes is fairly novel, as most related studies only find impacts on property crimes, and is particularly indicative of organized criminal activity in this setting. Importantly, offenses not likely associated with economic motivations and organized criminal enterprises (e.g., rape and marijuana consumption) do not show significant increases at the cutoff, allowing us to rule out many alternative theories. For instance, if social benefits induce risky behavior it should increase non-LACE crimes as well.

At the cutoff, the program encouraged informality. High-crime environments like Medellín have an informal market that contains many lucrative opportunities to be “employed” by criminal enterprises. Additional results show that impacts on LACE crimes are strongest in neighborhoods known to have the highest gang activity at baseline. Our results suggest that increases in formal employment can lead to reductions in criminal enterprise activities. Our magnitudes are in line with related studies, as we measure an occupational-switching elasticity of crime-to-formal employment of about 1.93.

Our contributions lie in validating economic models of occupational choice and entry into crime (Becker, 1968; Ehrlich, 1973). We leverage individual-level variation in employment incentives and rich administrative data to establish a causal relationship between formal employment and participation in criminal enterprise activity. Such evidence has proven difficult to find in a literature that has mostly relied on aggregate shocks, such as local recessions or trade-shocks. Aggregate-level studies estimate a combination of individual sorting and general equilibrium and neighborhood effects. That is, for example, firm relocation and employment displacement, victim migration and the spatial reallocation of resources, and peer effects among criminals and potential criminals are difficult to disentangle from the role of individual sorting

responses to employment opportunities (Dustmann and Damm, 2014; Ihlanfeldt, 2007; Kling et al., 2005, 2007). In contrast, the variation we leverage compares otherwise similar individuals with no change to their aggregate environment.

Recent studies have highlighted how unemployment shocks, job loss, and employment restrictions lead to increases in criminal activity (Bell et al., 2018; Bennett and Ouazad, 2018; Khanna et al., 2020; Pinotti, 2017; Rose, 2019; Schnepel, 2018). We build upon this small set of recent studies by documenting occupational sorting as a result of exogenous variation in exposure to a tax on formal wages. Job losses and structurally imposed employment restrictions may induce effects on depression, subsequent job search, and social stigma that are less likely in our setting. These sources of variation also generate financial necessity or desperation which can lead to temporary or idiosyncratic criminality, not necessarily reflective of an occupational choice. This role of financial necessity is shown to be a mitigating factor in the relationship between job loss and criminality by Britto et al. (2020); Khanna et al. (2020) and is emphasized by studies of current criminals and recently released prisoners (Blattman et al., 2017; Munyo and Rossi, 2015). We stress the importance of distinguishing between different types of crime, as some are more likely reflective of occupational sorting into criminal enterprise activity (e.g., firearms trafficking and heroin distribution) whereas others are more likely to be crimes of impulse or opportunity (e.g., rape and drug consumption). In doing so, we establish a falsification test to rule out alternative mechanisms that have little to do with occupational choice.

Additionally, there are fewer studies in the developing world, as many look at the US, the UK or Scandinavian countries from which data are more readily available (Bhuller et al., 2018; Cook et al., 2015; Davis and Heller, 2020; Gelber et al., 2016; Hjalmarsson and Lindquist, 2019). We study a high crime and high informality environment similar to most parts of the developing world and, in particular, a city with a significant presence of organized crime, which has been shown to have particularly detrimental effects on growth and development (Alesina et al., 2017; Blattman et al., 2018; Melnikov et al., 2019; Velasquez, 2020), and broader consequences for child development (Arteaga, 2019). We build upon recent evidence from important high-crime environments in Latin America that leverages area-based variation from trade shocks (Dell et al., 2018; Dix-Carneiro et al., 2018), or district-level unemployment (Buonanno and Vargas,

2018; Cortes et al., 2016). In contrast, we isolate occupational-sorting responses to formal employment disincentives, from other mechanisms arising from job destruction (Britto et al., 2020; Khanna et al., 2020), short-term liquidity constraints (Munyo and Rossi, 2015), and criminal capital deriving from past prison experiences (Tobón, 2020).

Finally, we highlight an unintended, adverse consequence of a welfare policy. In other contexts, access to welfare may reduce criminal participation (Deshpande and Mueller-Smith, 2021). Yet, several recent papers have specifically documented that means-tested benefits programs in Latin American countries affect formal employment. While this literature does not document changes in criminal participation, in high-crime settings like Medellín, reductions in formal participation may concurrently lead to increases in criminal participation. Bergolo and Cruces (2021) show that formal employment is impacted in Uruguay where formal earnings increase the chances of a poverty-score rising above an eligibility cutoff for a cash transfer program. Similarly, de Brauw et al. (2015) and Bosch and Schady (2019) show in Brazil and Ecuador, respectively, that means-tested cash transfer programs do not reduce overall work but do lead to a resorting from formal to informal employment. Relatedly, Gerard and Gonzaga (2021) show that an unemployment insurance program in Brazil discourages formal-sector employment.

Studies prior to ours have also documented the impact (of health insurance payroll costs) on formal employment in Colombia (Camacho et al., 2014; Kugler and Kugler, 2009; Lamprea and Garcia, 2016). In fact, the government adjusted the costs of the formal-sector health benefits program in response leading to a significant increase in formal employment (Bernal et al., 2017; Fernández and Villar, 2017; Kugler et al., 2017; Morales and Medina, 2017). We add to this literature evidence that criminality can be impacted along with formal employment. In doing so, we also build on previous work on the interaction between public sector interventions and crime (Agan and Starr, 2018; Chimeli and Soares, 2017; Chioda et al., 2016; Sviatschi, 2020; Yang, 2008).

2 Background

2.1 Crime in Medellín

Located in the north-western region of Colombia, Medellín is the second largest city after the capital, Bogotá. The urban zone consists of 249 neighborhoods, divided into 21 (*comunas*), 5 of which are semi-rural townships (*corregimientos*). Economic incentives are intricately tied to criminal participation in Medellín (Khanna et al., 2020). Here we describe some features of crime in Medellín, with further details in Appendix D.

Although Colombian violence has traditionally been high, the emergence of drug cartels in the late 1970s and early 1980s, fueled the emergence of organized crime to support illegal businesses, and guerrilla or paramilitary groups to care for the entire production chain. From the mid 1980s to early 1990s, homicide rates rose rapidly driven by the boom of cartels, paramilitaries, and local gangs. Over the period covered by our data, Medellín was one of the most violent cities in the world (CCSPJP, 2009), placing our analysis among a handful that study motivations behind joining organized crime in high-crime environments. The high homicide rates are a result of fights among urban militias, local gangs, drug cartels, criminal bands, and paramilitaries based in surrounding areas.¹ Demobilized militias continue to be involved in extortion and trafficking, given their experience with guns and avoiding police (Rozema, 2018).

There are two features of the homicide rate that are pertinent for our analysis. First, it is predominantly male. In 2002, the first year of our data, the male homicide rate was 184 per 100,000 whereas the female homicide rate was about 12, less than one-tenth the rate of males. Over the entire sample period (2005-13), 12% of all males (across all age groups) were at some point arrested, while the arrest rate for females was only 1%. Second, youth are far more likely to be involved than other age groups. Approximately, 63% of all first observed arrests are between 13 and 26. Younger individuals are more likely to be engaged in drug trafficking and consumption, whereas slightly older individuals are involved in violent crimes (homicides, extortions, and kidnapping), and the oldest still in property crime. Irrespective of type of crime, however, arrests peak within the 13 to 26 age window depicted in Figure A1.

¹*Operacion Orion*, followed by the demobilization of paramilitary forces led to a decline in homicides, as the military clamped down on urban militias (Medina and Tamayo, 2011).

In ongoing research, [Blattman et al. \(2018\)](#) document Medellín’s criminal world as hundreds of well-defined street gangs (*combos*) controlling local territory, organized into hierarchical relationships of supply, and protected by the *razones* at the top of the hierarchy. They confirm that gangs are mainly profit-seeking organizations, earning money from protection, coercive services such as debt collection and drug sales. Anthropological studies and in person interviews show that economic incentives (such as the focus of our study) drive young men in Medellín to join organized crime ([Baird, 2011](#)). As many respondents highlight, the reason to join crime is mostly “economic” or for a profitable career (see, e.g., *interview with Gato, p264* and *interview with Armando, p197*). Knowing this, groups actively recruit idle youth that are *amurrao* (local slang: ‘sitting on the wall’) and without a formal sector job.

An interview with El Mono (*p191*) documents the recruitment process: “*those guys would hang out around here and be nice to me and say ‘come over here, have a bit of money’.*” Having a formal sector job means that one is not “hanging around the neighborhood” when the gangs come to recruit. A desirable outside option would be a job with benefits and social security, yet those with formal sector jobs pay extortion fees to gangs (see *interview with El Peludo, p184*). Indeed, the options are often presented as an occupational choice: “*are you gonna work [for the gang] or do a normal job?*” (see *interview with Notes, p193*).

Often remunerations for gang members are higher than legitimate jobs for those with similar levels of education ([Doyle, 2016](#)). New recruits are employed to run guns (*carritos*), before transitioning to extortion and trafficking. [Blattman et al. \(2018\)](#) estimate that foot soldiers of the combos receive well above national minimum wage whereas *combo* leaders earnings “put them in the top 10% of income earners in the city.” These anecdotes are consistent with our hypothesis: better benefits for informal workers discourage young men from joining the formal sector, which in turn leads many to be recruited by criminal enterprises. Internalizing this trade-off, during the demobilization of militias in the mid-2000s, many were encouraged to join the formal sector, given identity cards and medical cards ([Rozema, 2018](#)). Yet, this disparity in costs across social benefit regimes, discouraged formal sector re-integration.

We can link our arrests data to administrative data on the universe of formal sector workers from the *Planilla Integrada de Liquidación de Aportes* (PILA), to understand which industries

those who are ever arrested worked in. Those ever arrested, are more likely to be in construction, rental/leasing, and transportation and warehousing, than their ‘never-arrested’ counterparts. These may partly reflect the industries that criminal enterprises are also involved in.

For our sample of young men in the bandwidth of analysis, 21.5% were arrested over the period of study – 11.1% for drug crimes, 5.6% for property crime, and 4.8% for violent crimes. These numbers are high relative to most contexts, but are representative of cities in Latin America. The US has an incarceration rate more than six times the typical OECD nation, where one in ten youths from a low-income family may join a gang, 60% of crimes are committed by offenders under the age of 30, and 72% by males (Kearney et al., 2014). Accordingly, in some regards, arrests in our context are similar to high-crime regions in many parts of the developing world, and especially Latin America (Brown and Velasquez, 2017; Sviatschi, 2018).

2.2 Access to Health Benefits

In 1993, *Law 100* established two tiers of health insurance: the Contributive Regime (CR) and the Subsidized Regime (SR). In Appendices C1 to C5, we describe the details behind the policy, the costs of access, coverage, eligibility criteria, the corresponding policy variation, and changes over time. Here, we summarize the information relevant for the analysis.

The CR covers formal workers with a comprehensive set of health services that includes nearly all of the most common illnesses. The SR covers the families of the poorest informal workers and unemployed with a plan that initially covered fewer illnesses than CR, but was expanded to cover the same benefits.² Formal workers and employers fund workers’ insurance premiums for coverage by the CR. Between the 1993 reform and 1998, insurance coverage under both grew from 20% to 60%. By 2007, health coverage in Colombia was above 90% (Ministry of Health and Social Protection). Thus, health insurance was already high for our sample period in the country. In the case of Medellín, insurance was nearly universal, and it stayed at or above 95% by the end of the analyzed period (Lamprea and Garcia, 2016).

²In 2008, the Constitutional Court ordered that the basket of health services covered under SR be equal to that of CR. CR was slightly more comprehensive in the early years, which could have theoretically weakened our first-stage on being formal. Nevertheless, we find strong impacts on both SR enrollment and formal employment as presented below. At the beginning of our sample period, the subsidized program covered nearly 60% of health services that the full program covered – this fraction increased steadily to 100% of services.

Colombian employers are required by law to enroll all their employees in a Health Promoting Company, which gives them access to health insurance under the CR. Formal employees pay up to 12.5% of their wage, and cross-subsidize informal workers in SR.³ Authorities initially expected the formal sector population to rise and cover costs for SR. But the SR grew faster than the CR population, in part due to the generous nature of the SR (Lamprea and Garcia, 2016). Costs to formal workers rose substantially after the 1993 reforms, with strong evidence that such costs discouraged formal sector employment (Kugler and Kugler, 2009).

Self-employed workers are allowed to enroll in the CR themselves by paying a monthly fixed amount based on a percentage of the monthly minimum wage. Unemployed or inactive individuals (and informal workers) can either get health insurance as the self-employed do through the CR, or apply for access to the SR (if eligible). Individuals not covered by the CR or the SR use public hospitals, and are charged fees for both medicines and services. For instance, to the best of our knowledge, there are no meaningful differences in terms of lines at hospitals, or wait times for appointments under the two regimes.

To target the SR, roughly 70% of the poorest households in the country were interviewed between 1994 and 2003, and a welfare index (*Sisben* score) was calculated using a confidential formula based on respondent characteristics, incomes and assets, disability, education, and housing. Only households with a *Sisben* score below a certain cutoff and not formally employed were eligible to become beneficiaries of the SR. Households keep their *Sisben* score until it is updated by the government (expected to take place about every five years). Other public programs use the *Sisben* score, but the SR *Sisben* cutoff did not coincide with other major interventions, at the eligibility cutoff of *Sisben* during the study period.⁴ The SR health program is by far the largest that has eligibility determined by the *Sisben* score. The SR budget is nearly

³Formal sector workers make up about 54% of the urban labor force and pay 4% of their monthly wage for enrollment in the CR, while the employer pays the other 8.5%. This implies that effectively employees may bear a burden somewhere between 4 and 12.5% of their monthly wage depending on their bargaining power. Formal workers pay 1.5% of their salary to cover informal workers in SR. Independent workers earning above the cutoff (richer) are required to contribute 12.5% of the maximum between one monthly minimum wage and 40% of their earnings to the CR. As such, for those above the income cutoff (richer), the cheapest way to get health insurance is to be formal (and pay 4-12.5% of their wages); while for those below the cutoff, it is to not be formally employed and enroll in SR.

⁴See www.sisben.gov.co/Paginas/Noticias/Puntos-de-corte.aspx for programs by *Sisben* 3 cutoff. While the *Sisben* cutoff for SR enrollment may differ across counties, there is only one cutoff for the entirety of Medellín, and it did not coincide with other major programs. See also Appendix C5.

2% of GDP, while all other anti-poverty programs represent less than 0.4% of GDP.

2.3 Incentives for Informality

Between 2005 and 2013, informal workers made up 46% of the urban labor force in Medellín. Among informal workers, around 60% were own-account workers, 20.5% were private sector employees, 7.8% were domestic workers, 7.7% self-employed workers, and the rest were laborers, family workers, and unpaid workers. There is a stark education gradient, where more education is associated with a higher likelihood of formal employment (Medina et al., 2013).

Effectively, financial incentives embodied in the health coverage options switch from potentially promoting formal employment above the cutoff due to a partial defrayal of the costs of healthcare by the employer, to strongly discouraging formal employment below the cutoff due to a significantly more enticing full defrayal of these large costs by the government for individuals who are not formally employed. Near complete health care coverage in the population, despite costs representing large proportions of income, reveals the importance of these incentives.

The attractiveness of SR is compounded by the lower costs of access as well. Over and above their enrollment contributions (up to 12.5% of their wage), workers in the CR pay a copayment and a *Cuota Moderadora*. The copayments across health services are a portion of the costs covered by individuals out of pocket that vary with the workers' salary (number of minimum wages). The *Cuota Moderadora* is a fixed out-of-pocket amount paid by an insured for covered services and varies with the workers' salary (number of minimum monthly wages). Informal workers enrolled in the SR neither pay for their enrollment, nor the *Cuota Moderadora*, and have much lower copayments. For more details see Appendix C2.

That this policy led to a fall in formal-sector employment has been documented in both the academic literature and public discourse. The Minister of Social Protection, in a news article in *Presidencia de la Republica* (February, 2006), claimed that the people's valuation of SR was so high that it discouraged formal employment. Studying the effects on the entire country, Camacho et al. (2014) use individual-level data and control for both region and time fixed effects to show that informality increased by 4 percentage points as SR was rolled out across the country. This is a combination of workers dropping out of the formal sector, but

also fewer youth joining the formal sector over time (Lamprea and Garcia, 2016). Recognizing these adverse effects on formal employment, the government drastically lowered the costs of being enrolled in CR right at the end of our study period, when Law 1607 was enacted. Not surprisingly, this led to a significant increase in formal sector employment (Bernal et al., 2017; Fernández and Villar, 2017; Kugler et al., 2017; Morales and Medina, 2017).

Since the *Sisben* score is calculated at the family rather than individual level (Article 21, Decree 2353 of 2005), other work suggests that older family members may have reason to discourage youth within the family from joining the formal labor force for fear of losing access to benefits, and that large families stay informal in the hope of retaining benefits (Bergolo and Cruces, 2021; de Brauw et al., 2015; Joumard and Londono, 2013). The Ministry of Health maintains a census of all those enrolled in both CR and SR, and if someone in the household becomes formal, and so enrolls in CR, they and all dependents (including adult dependents) are made ineligible for SR. This administrative database is updated, and becoming formal will also increase your *Sisben* score the next time it is updated.

Similarly interviews in Baird (2011) highlight how being involved in crime can sometimes be a ‘family decision.’ This is confirmed by our other work in the same context that show spillovers in criminal activity across generations within the family (Khanna et al., 2020). Indeed, Santamaria et al. (2008) find that half of all SR recipients indicated that they would not switch to formal employment as it would mean losing benefits. These effects are not restricted to men, as women’s formal-sector participation also decreased in response to SR (Gaviria et al., 2007). Yet, we find that dis-employment effects on men are about four times larger than on women, consistent with the hypothesis that men have a lucrative alternative outside the formal sector: criminal enterprises.

We leverage the fact that the costs of accessing these benefits change discontinuously at the *Sisben* cutoff. Indeed, as most individuals are covered by one healthcare regime or the other, almost everyone has access to benefits on either side of the cutoff by the end of this period (Lamprea and Garcia, 2016). Yet, on one side of the cutoff these benefits are free only if you are not formally employed. The primary driving variation, therefore, is that being outside the formal sector allows you to not pay for benefits on one side of the cutoff.

3 Theoretical Framework

We first posit a simple model of occupational choice and generate empirical predictions which we test in the subsequent section. We model the heterogeneity in the utility that workers derive from living and working in different parts of the city following [Eaton and Kortum \(2002\)](#). There is a continuum of workers living across D different neighborhoods denoted with $d \in \{1, \dots, D\}$. In stage 0, workers receive their Sisben score $o \in [0, 1]$, which depends on the characteristics of their households. In stage 1, individuals decide on which sector, σ , they want to work in, that is, $\sigma \in \{f, n, c\}$, where f , n , and c denote the formal, informal (i.e., nonformal, noncriminal activities), and criminal sector, respectively. Only workers with a score below the threshold, $o^* \in (0, 1)$, can enroll in the subsidized regime (SR) if they do not choose to work in the formal sector in stage 1. Whereas, individuals with a score below the threshold o^* who choose $\sigma \in \{n, c\}$ in stage 1, enroll and enjoy the benefits of the SR.

Individual Occupational Choices and Consumption. In our framework, the criminal sector is an occupational choice, and so will affect whether these individuals are more likely to be engaged in enterprise-related LACE crimes. They also choose to consume/engage in other illicit activities K_d that are not related to their occupation, such as consuming marijuana or engaging in crimes of impulse/passion. The utility of a (d, o) -type worker (i.e., a worker with Sisben score o , living in neighborhood d) from choosing sector $\sigma \in \{f, n, c\}$ depends on the consumption of illicit activities K_d , non-illicit consumption $C_{\sigma do}$, and an occupation specific preference draw $\epsilon_{\sigma do}$.

Illicit activities, K_d , may be costly, with cost $r_{K,d}$ capturing say the cost of marijuana consumption, or bribery, or an opportunity cost of engaging in non-productive activities. Furthermore, the opportunity to engage in such crimes of impulse or passion \bar{K}_d , may be allowed to vary across neighborhoods d . Each worker observes their realizations of the idiosyncratic utility component for each occupational choice and solves the following problem:

$$\begin{aligned} \max_{C_{\sigma do}, K_d} U_{\sigma do} &= \max_{C_{\sigma do}, K_d} b_{\sigma do} C_{\sigma do} K_d \epsilon_{\sigma do} & (1) \\ \text{s.t.} \quad C_{\sigma do} &\leq w_{\sigma do} \quad \& \quad K_d r_{K,d} \leq \bar{K}_d, \end{aligned}$$

where $b_{\sigma do}$ captures the relative benefits from living in neighborhood d , having Sisben score o , and choosing working sector σ ; $w_{\sigma do}$ is the earnings of workers in sector σ earned by type (d, o) , and $\epsilon_{\sigma do}$ is an idiosyncratic term that captures individual preferences for choosing sector σ given the workers type. The term $\epsilon_{\sigma do}$ is drawn from the Frechet distribution $F(\epsilon) = e^{-E_\sigma \epsilon^{-\gamma}}$, where E_σ is the average utility derived from working in sector $\sigma \in \{f, n, c\}$, and the shape parameter $\gamma > 1$ defines the dispersion of the idiosyncratic utility term. From (1), it follows that individuals would consume their entire earnings, and so the indirect utility for a (d, o) -type worker is:

$$U = \frac{b_{\sigma do} \bar{K}_d}{r_{K,d}} w_{\sigma do} \epsilon_{\sigma do}. \quad (2)$$

The distribution of the indirect utility U determines the probability that a (d, o) -type worker chooses working in sector σ , and it is given by:⁵

$$\pi_{\sigma do} = \mathbb{P} \left(U \geq \max_{\sigma'} U \right) = \frac{E_\sigma (b_{\sigma do} w_{\sigma do})^\gamma}{\sum_{\sigma' \in \{f, n, c\}} E_{\sigma'} (b_{\sigma' do} w_{\sigma' do})^\gamma}. \quad (3)$$

Thus, the probability of choosing to work in sector σ , for type (d, o) depends simply on the return to working in sector σ , relative to all possible sectors. To close the model, we introduce the equations governing the wage in each of the sectors. We allow these wages to depend on neighborhood-level characteristics, such as the number of workers in each neighborhood, and the opportunity to be more productive in certain types of jobs.

Formal Sector. The production in the formal sector from people living in neighborhood d is given by:

$$Y_{fd} = a_{fd} H_{fd}^\theta, \quad (4)$$

where a_{fd} is the productivity of workers in neighborhood d in the formal sector. This is allowed to vary by neighborhood-level opportunities (captured by d). H_{fd} is the neighborhood d population working in the formal sector (i.e., $H_{fd} = \bar{H} \pi_{fd}$, where $\pi_{fd} = \pi_{fd[o \geq o^*]} + \pi_{fd[o < o^*]}$), and $\theta \in (0, 1)$ is the elasticity of output with respect to formal labor. We assume that the labor market is competitive, then, the wage in the formal sector is given by the marginal product

⁵See Appendix B1 for a derivation of $\pi_{\sigma do}$.

of labor:

$$w_{fdo} = a_{fd}\theta H_{fd}^{\theta-1}. \quad (5)$$

Informal Sector. Informal (non-illicit) wages follow a similar functional form as formal wages. We assume a competitive informal labor market where the total production is given by:

$$Y_{nd} = \frac{a_{nd}}{(\rho + 1)\bar{H}^\rho} H_{nd}^{\rho+1}, \quad (6)$$

where a_{nd} is the productivity of workers in neighborhood d in the informal sector, and H_{nd} is the population engaged in noncriminal non-formal activities in neighborhood d (i.e., $H_{nd} = \bar{H}\pi_{nd}$ and $\rho \in (-1, 0)$). Thus, wages are a function of the population in the informal sector of each neighborhood d and equal to:

$$w_{ndo} = a_{nd} \left(\frac{H_{nd}}{\bar{H}} \right)^\rho. \quad (7)$$

Crime Sector. When individuals engage in LACE crimes, with probability p_d they earn a wage $\bar{w}_{cdo} \equiv f(H_{cd})$ that is a function of the population engaged in LACE crimes, H_{cd} . Note that p_d can be a function of several elements, such as the efficacy of criminals in being successful at crime, or the size of institutions in the society: $1 - p_d$ can represent the probability of being caught while committing a crime. Thus, the expected salary of individuals engaging in LACE crimes is:

$$w_{cdo} = p_d f(H_{cd}) = a_{cd} \left(\frac{H_{cd}}{\bar{H}} \right)^\iota, \quad (8)$$

where $H_{cd} = \bar{H}\pi_{cd}$, and $\iota \in (-1, 0)$. a_{cd} is productivity of criminals living in neighborhood d , engaged in LACE crimes. We allow for this productivity to vary by neighborhood opportunities (such as the presence of gangs or militias). Note that the formulation for w_{cdo} can be generalized to one in which there is random shock ν accounting for the randomness of wages in the criminal sector. So that $\bar{w}_{cdo} \equiv f(H_{cd}, \nu)$. Thus, the expected salary of individual engaging in LACE crimes would be $w_{cdo} = p_d \mathbb{E}_\nu[f(H_{cd}, \nu)]$. Note that if ν is chosen to have a constant mean, and $f(H_{cd}, \nu) = \nu f(H_{cd})$ then both formulations would be similar.

Testable Hypotheses. We derive our testable hypotheses. Taking logarithms on both sides of Equation (3) and using the fact that $\pi_{fdo} = \frac{E_f(b_{fdo}w_{fdo})^\gamma}{\sum_{\sigma' \in \{f,n,c\}} E_{\sigma'}(b_{\sigma'do}w_{\sigma'do})^\gamma} = \frac{H_{fdo}}{\bar{H}}$, we get:

$$\log(\pi_{cdo}) = \gamma \log(w_{cdo}) - \gamma \log(w_{fdo}) - \log\left(\frac{\bar{H}}{H_{fdo}}\right) + \log\left(\frac{E_c b_{cdo}^\gamma}{E_f b_{fdo}^\gamma}\right). \quad (9)$$

We can use the wage equations $w_{fdo} = a_{fd}\theta H_{fd}^{\theta-1}$ and $w_{cdo} = p_d a_{cd} \left(\frac{H_{cd}}{H}\right)^\iota$, to express (9) as the relative probabilities of choosing crime-related occupations π_{cdo} and formal sector jobs π_{fdo} :

$$\log\left(\frac{\pi_{cdo}}{\pi_{fdo}}\right) = \gamma(\log(H_{cd}^\iota) + p_d) - (\theta - 1)\gamma \log(H_{fd}) + \log\left(\frac{E_c}{E_f \theta^\gamma \bar{H}^{\iota\gamma}}\right) + \gamma \log\left(\frac{a_{cd}}{a_{fd}}\right) + \gamma \log\left(\frac{b_{cdo}}{b_{fdo}}\right). \quad (10)$$

Equation 10 guides our empirical strategy. Notice here that the only terms that vary with the Sisben score o is the last term $\gamma \log\left(\frac{b_{cdo}}{b_{fdo}}\right)$. Note, while this equation motivates a logistic-regression, as we are doing an RD, we approximate with a linear-probability model. The regression discontinuity design helps control for all other terms on the right-hand side as they do not discontinuously change at the o^* cutoff. As such, the RD controls for differences in policing or the probability of capture p_d , the occupational structure of the neighborhood, local job opportunities, and so on. We can even further include neighborhood fixed effects to account for these factors. Estimation strategies that rely on area-based variation (say, from local recessions or trade shocks), would also capture the effects of changes in funds for policing p_d , criminal opportunities and neighborhood composition H_{cd} and H_{fd} , for instance, and make it challenging to recover γ .

The coefficient on the term $\log\left(\frac{b_{cdo}}{b_{fdo}}\right)$, γ , is the relative labor-supply (occupational choice) elasticity with respect to changes in relative benefits. As access to SR changes when $o < o^*$ if $\sigma \neq f$, we posit that b_{cdo} jumps discontinuously at the o^* cutoff. At this point, individuals may choose $\sigma \neq f$ as a result, and either engage in criminal or non-formal activities $\sigma \in \{c, n\}$ depending on the relative benefits to each. Equation 10 has a counterpart with $\log\left(\frac{\pi_{cdo}}{\pi_{ndo}}\right)$ as an outcome showing what determines the decision between criminal and non-formal jobs. As the benefits between these two options should change proportionally at the cutoff o^* , we are unlikely to see a change in the relative benefits $\log\left(\frac{b_{cdo}}{b_{ndo}}\right)$.

Illicit Consumption. Finally, note that non-LACE crimes of impulse or passion K_d do not change at the cutoff. This is because our model posits that these non-occupational illicit activities are best conceived of as additional consumption. What changes at the Sisben cutoff o^* is the benefit in being in a certain *occupation* (e.g., being involved with a gang); while consumption (both non-illicit and illicit) do not change discontinuously at the cutoff.

4 Data

Administrative data allow us to identify the relationship between incentives for informality and participation in criminal enterprise. We do not need to rely on self-reported or aggregate victim counts. As our data is at the individual-level, we isolate vulnerable demographics (young men), and test both employment outcomes and crime. Additionally, detailed information on the types of crime allow us to isolate mechanisms.

We combine two sources of data at the individual level using national identification numbers and dates of birth. One source is from successive *Sisben* surveys of the Medellín population in 2002 (baseline *Sisben I*), 2005 (*Sisben II*) and 2009-2010 (*Sisben III*). The *Sisben* dataset consists of cross sections from censuses of the poor, and we match household records across the three waves. These are municipality censuses of people living in the three poorest socioeconomic strata. In large cities, like Medellín, it amounts to 65-80% of the population. The second source is the census of individuals arrested between 2002-13 from the Judicial Police Sectional of the National Police Department. We describe these data in detail in Appendix D. We match 78% of arrests to the *Sisben* survey as many arrested may not be residents of Medellín.

We focus on the probability of ever being arrested over the period, which by and large in this young population is their first observed arrest. The literature often restricts data to first arrests, and repeat arrests are excluded as time spent under incarceration and the length of sentencing may be endogenous to other characteristics. Indeed, first arrests most closely map to the first decision node between legal and illegal activities. Once captured a criminal career begins, with subsequent decisions to repeat, escalate, or exit the criminal sector based on many factors we do not observe (including prison sentences). Accordingly, subsequent criminal behavior is outside the scope of this study. Nevertheless, we show that our results are robust

to including repeat arrests.

For similar reasons, we follow recent studies (Gronqvist, 2017; Kling et al., 2005) in focusing on young men in our analysis. Our primary sample was between 13 and 18 in 2005, and we follow these cohorts for the entire period of our analysis (i.e., till 2013, the last year of our arrest data). This means, we are only documenting arrests between the ages of 13 and 26, capturing more than 63% of all first observed arrests (as shown in Figure A1). Only about 4% of arrests are below the age of 16, and underage arrests are less likely to lead to prison sentences, and more likely to be short stays in special youth shelters (Guarin et al., 2013; Ibanez et al., 2017). In robustness checks, we expand the age criteria to include individuals who were as young as 10 in 2005 (or 18 in 2013). Expanding the age criteria produces similar results.

Of the individuals arrested more than once during the observation period for any crime (not just LACE), 40% are first arrested before the age of 27. At the same time, while incarcerated, individuals would not be able to be arrested for additional crimes and would, therefore, have lower measured propensities to be engaged in new criminal activity. Older individuals may have been arrested in their youth (or currently still be incarcerated) but as our crime data only begins in the early 2000s, we do not have their entire criminal history, and would miss their youth arrest. As such, we exclude older men. Focusing on ages when arrest rates peak reduces these concerns regarding the measurement of criminality, and allows us to emphasize the period when young men first make choices between crime and other jobs in Medellín (Doyle, 2016).

Figure A2 describes the timeline of our data. We use the 2002 *Sisben* as our baseline to create our running variable and predict eligibility for SR. The formula to compute the Sisben score and the eligibility cutoff varies across the waves (I, II, and III). Since our crime data begins in 2002, we use the 2002 Sisben score as our running variable. We test for SR enrollment, and for employment status and incomes in both the 2005 and 2009 *Sisbens*. We then follow the criminal histories of young men aged 13 to 18 in 2005, between 2005 (after we have a measure of SR enrollment from the second *Sisben*) and 2013.

Table 1 presents the 2002 baseline summary statistics of the complete *Sisben* survey and for the subsample of male youth only. The SR status is established based on the previously computed Sisben score, from the semi-decadal Sisben municipality census of the 70% of the

Table 1: Summary Statistics in 2002

Variable	<i>Complete Sample</i>		<i>Males</i>	
	Mean	Std. Dev.	Mean	Std. Dev.
<i>Individual Characteristics</i>				
Male	0.490	0.500	1.000	0.000
Subsidized Regime	0.319	0.466	0.312	0.463
Contributive Regime	0.228	0.420	0.222	0.416
Age 10-15	0.105	0.306	0.109	0.311
Age 15-20	0.105	0.306	0.110	0.313
Age 20-25	0.089	0.285	0.093	0.290
Age 25-30	0.068	0.251	0.068	0.251
Ever Arrested	0.062	0.242	0.114	0.318
<i>Household Head (HH) Characteristics</i>				
Female	0.387	0.487	0.308	0.462
Employed	0.628	0.483	0.643	0.479
Unemployed	0.106	0.308	0.107	0.309
Married	0.345	0.475	0.377	0.485
Attending School	0.009	0.097	0.008	0.089
Has CR	0.219	0.413	0.214	0.410
Age	43.237	14.302	43.869	14.159
Years of Education	4.542	2.451	4.480	2.454
Owns House	0.314	0.464	0.327	0.469
Sisben Stratum 1	0.271	0.444	0.273	0.446
Sisben Stratum 2	0.620	0.485	0.620	0.485
Sisben score	45.707	9.901	45.716	9.908
Number of members in household	4.090	1.709	4.215	1.709
N	1,161,446		568,923	

Summary tabulations using *Sisben* I survey, conducted in the year 2002, and police arrests data.

poorest population. After a *Sisben* survey, it takes around one or two years to get the new *Sisben* score and household eligibility. Accordingly, we use the lagged *Sisben* score as the running variable for our analysis. The arrests data include a detailed description of the person arrested (national identification number and date of birth), type of crime (e.g., homicide, rape, motor vehicle theft, etc.), the precise penal code article associated with the crime, a description of the act (e.g., trafficking cocaine), the date of arrest, the location of arrest, and a police flag for whether the officer knew the perpetrator to be gang affiliated. We also observe the location of the crime, whereas in the *Sisben* we observe the location of residence (which we use for neighborhood fixed effects).

4.1 Classifying Crimes

We classify the crimes into three categories – violent, property, and drug crimes – based on the US Bureau of Justice Statistics’ classifications in the Sourcebook of Criminal Justice Statistics (BJS, 1994). If, for example, an individual was first arrested for violent crime and later for property crime, they show up as an arrest for violent crime. We use the detailed description of the act (e.g., distribution/trafficking cocaine) to categorize the crimes (e.g., drug crimes).

Our theoretical framework encourages us to distinguish between occupation-related crimes “likely associated with criminal enterprises” (LACE), and the consumption-driven non-LACE crimes. To that end, we worked closely with senior police officials in Medellín to divide our crimes into LACE and those more likely to reflect impulse, passion, or opportunity (non-LACE). For about 30% of our data, the police used a system that flagged the arrest with whether the individual was known to be part of an organized criminal enterprise or not, as well as information on the specific organization to which the individual belonged. This organizational affiliation was based on extensive police intelligence and follow up interviews. The data span 284 street gangs, urban militias, *narcotraficante* (drug distributors linked to gangs), and other organized criminal entities.

The police discontinued this system after a period of time. As the gang-flag system was not available for the entirety of our period, we use the patterns revealed over the subset of arrests for which the flag is available to classify the full sample of arrests. Police officials advised us that the best way to classify arrests is along two dimensions: (1) the crime and (2) the location. Accordingly, in our main analysis, we classify a crime as likely associated with criminal enterprises (LACE) if more than 30% of recorded arrests for that crime had the gang flag. We use 30% as that is the fraction of the sample under which the police was using the flags, but our results are similar when using the median as the cutoff. As a result of this exercise, for example, we classify homicide arrests as violent LACE, and rape or domestic violence as violent non-LACE.⁶

In robustness checks, we use a method that relies on the association between these crimes and historically high-gang neighborhoods. In this alternative definition, we classify crimes as

⁶Gang rape gets classified as a gang crime. Our police contacts also describe how burglars are imbibed into gangs based on their work territories and would find it difficult to be a burglar without being a part of the gang.

LACE if they are more likely (above the median) to list any of these high-gang neighborhoods as the location of arrest; however, our results are similar when using other cutoffs from 30% to 60% of arrests. Thirdly, we also use methods that predict the likelihood that the crime is LACE given observables, after training it on the sample that reports gang flags. Here we use k-means logit to predict LACE crimes based on neighborhood, crime and individual characteristics.

While neither classification is perfect, the robustness across classification methods helps to validate the exercise. Additionally, using the crime-level classification (rather than the individual flags) of LACE crimes protects us against any police biases against specific individuals, or their characteristics (such as insurance status or who the police have more intelligence on).

In Appendix Table A1, we categorize the 25 (of 103) most prevalent crimes under each classification method. These data-driven methods line up with our priors on types of crime: homicides, motor vehicle theft, extortion, kidnapping, break-ins, and the manufacturing, delivery and trafficking of drugs fall under LACE crimes. The remaining crimes are often thought of as crimes of impulse or opportunity (like rape, simple assault, and drug consumption). Indeed, we can distinguish between nuanced details – such as trafficking cocaine (LACE) vs. consuming marijuana (non-LACE). The advantage of these classification approaches is that they are purely data driven. Additionally, they may speak to the types of activities that gangs in Medellín engage in: for example, they are more likely to engage in car theft than identity theft. The table also shows the division between LACE and non-LACE crimes for each category. About 53% of all arrests are for LACE crimes.

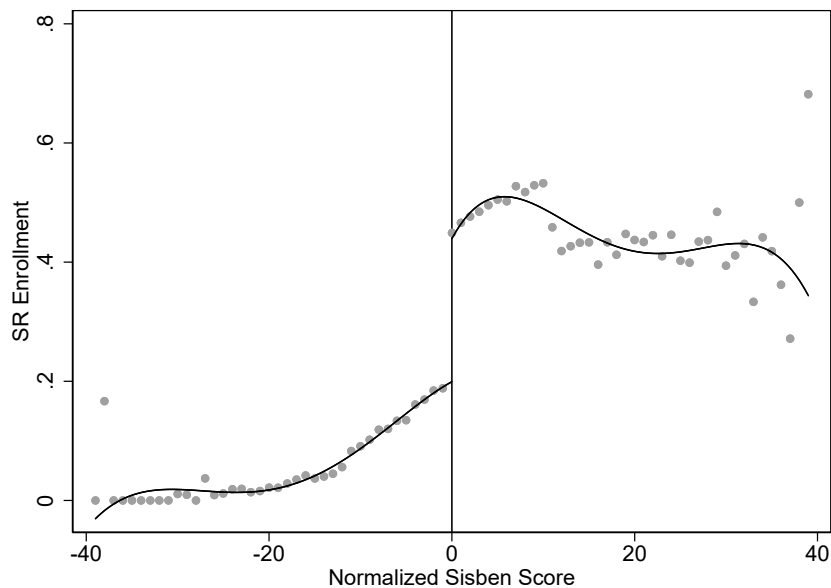
5 Enrollment in the Subsidized Regime (SR)

As only households in the two lowest levels of *Sisben* I (2002), a score below 47, could qualify for the SR, we compare households on either side of the cutoff to identify the effect of SR eligibility. First, we verify that there is a discontinuity in the probability of SR enrollment at the cutoff. Second, we examine how the likelihood of being in the formal sector changes at the cutoff. Last, we examine the effect on different types of criminal activity.

Following RD convention, we normalize the *Sisben* score so that treated units are individuals with positive values of our new score. Figure 1 presents the first stage: the discontinuity in

the probability of SR enrollment using the optimal binning procedure found in [Calonico et al. \(2014a\)](#). The probability of enrollment discontinuously increases by around 26 percentage points. Around 20% of households that have a high 2002 *Sisben* also avail of SR in 2005, as a fraction of households became eligible under a smaller 1998 *Sisben* survey, and the government allows them to keep their benefits for some time after they graduate out of eligibility. Not all eligible persons enroll in SR, as formal sector jobs may be valuable to some, but enrollment still jumps substantially to 42% at the cutoff.

Figure 1: Discontinuity in the Probability of SR enrollment at Cutoff.



SR enrollment is probability of being enrolled in the subsidized regime in 2005. RD graph using optimal binning procedure discussed in [Calonico et al. \(2014a\)](#). Normalized Sisben (from 2002) score on horizontal axis centered around cutoff. Higher values represent lower scores (higher poverty).

For two-staged least squares (2SLS) exercises we follow a fuzzy regression discontinuity design, where our running variable is the 2002 *Sisben* score. We use both parametric and non-parametric approaches to estimate the effect of SR eligibility at the cutoff. For the parametric approach we follow [Hahn et al. \(2001\)](#), where we instrument enrollment in the SR with the eligibility indicator $1[s_i < 47]$, and estimate the following equation as our first stage:

$$SR_{i,n} = \alpha + \alpha_1 1[s_{i,n} < 47] + X'_{i,n} \alpha_2 + A_i(s_{i,n}) \alpha_3 + \mu_n + \varepsilon_{i,n}, \quad (11)$$

where A_i is a vector of smooth polynomial functions of the *Sisben* score of each individual, $s_{i,n}$. We also estimate models conditioning on demographics and other baseline characteristics.

Here $X_{i,n}$ is a vector of demographic characteristics for individual i living in neighborhood n . μ_n corresponds to neighborhood fixed effects for the 249 neighborhoods. In robustness checks, we control for various characteristics of the household head in 2002, the baseline year. These controls include an indicator for female-headed households, employment status, years of education, marital status, attendance to any academic institute, year-of-birth fixed effects, socioeconomic strata of the household, and home ownership. In additional robustness exercises we control for community characteristics interacted with being above the *Sisben* cutoff.

An important issue in practice is the selection of the smoothing parameter. We use local regressions to estimate the discontinuity in outcomes at the cutoff point. In particular, we estimate local polynomial regressions conducted with a rectangular kernel and employing the optimal data-driven procedure suggested by [Calonico et al. \(2014b\)](#). We use two different optimal bandwidth procedures: the [Imbens and Kalyanaraman \(2012\)](#) method and the [Calonico et al. \(2014b\)](#) bandwidth. The optimal bandwidths from the different procedures lie between 5.5 and 6.2 points, on the 100-point *Sisben* I scale. We present our results for multiple bandwidths to highlight the robust nature of our estimates, varying them from below the optimal bandwidths to larger bandwidths. Specifically, we check for coefficient stability for results spanning these bandwidths ranging between 4 and 10 points around the cutoff. Varying the size of the bandwidth and the polynomial order do not affect the results.

Table 2: SR Enrollment at Sisben Cutoff (First Stage)

Bandwidths:	In 2005			In 2009		
	4	6	10	4	6	10
	Dependent variable: Enrollment in SR					
Above Cutoff	0.260*** (0.0138)	0.260*** (0.0132)	0.269*** (0.0110)	0.242*** (0.0122)	0.241*** (0.0112)	0.250*** (0.00999)
F-stat of IV	354.97	387.97	598.02	391.9	461.8	625.1
Observations	181,132	246,974	340,581	181,132	246,974	340,581
Sample mean			0.36			0.31

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Coefficient of indicator of being below *Sisben* cutoff, with linear controls for 2002 *Sisben* scores that vary flexibly at the cutoff. SR enrollment as measured in the 2005 and 2009 *Sisben* surveys. Standard errors clustered at the *comuna* level.

Our first stage results are shown in [Table 2](#), confirming the 24 to 26 percentage point

increase in SR enrollment shown in Figure 1. As we vary the bandwidths from 4 through 10 the coefficient is stable and both economically and statistically significant. The table also shows that the standard IV F-test suggests a strong instrument, and for our remaining outcomes we conduct two-staged least squares analyses using this as our first stage.

6 Impacts on Formal Employment and Reported Income

Guided by our theoretical framework, we test the simple hypothesis that the SR conditions disincentivized formal-sector employment and led to an increase in organized-crime activities. We first reproduce a well-established result and show that the program has a negative effect on formal employment (Camacho et al., 2014; Gaviria et al., 2007; Joumard and Londono, 2013; Santamaria et al., 2008). We exploit the discontinuity in enrollment rates at the cutoff, by using the eligibility indicator as an instrument for enrollment status to identify the effect of SR on formal employment and income. Here $Emp_{i,n}$ is 1 if the individual i from neighborhood n was formally employed. We include demographic controls in $X_{i,n}$, and neighborhood fixed effects μ_n . We show the reduced form relationship between employment and being above the RD cutoff:

$$Emp_{i,n} = \gamma_0 + \gamma_1 1[s_{i,n} < 47] + X'_{i,n} \gamma_2 + A_i(s_{i,n}) \gamma_3 + \mu_n + \varepsilon_{i,n} ,$$

We then instrument for SR enrollment, where $S\hat{R}_{i,n}$ is the predicted SR enrollment probability from the first stage estimated in equation 11. The second stage is:

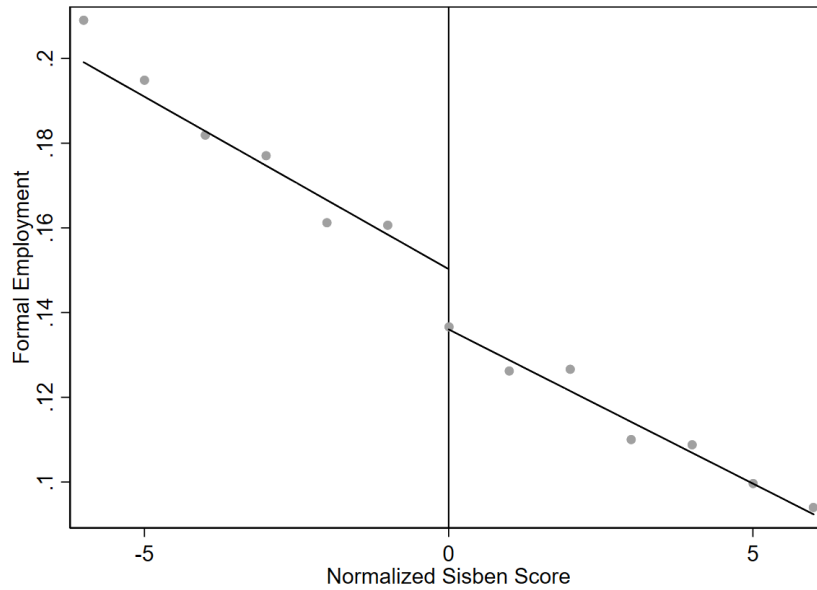
$$Emp_{i,n} = \beta_0 + \beta_1 S\hat{R}_{i,n} + X'_{i,n} \beta_2 + A_i(s_{i,n}) \beta_3 + \mu_n + \epsilon_{i,n} ,$$

Figure 2 captures the fall in formal sector employment at the cutoff, where formal employment is defined as a working individual making wage contributions to benefits as measured in the 2009 *Sisben* III survey.⁷ In all remaining RD figures, we focus on a bandwidth of 6 around the cutoff as it is the Calonico et al. (2014b) optimal bandwidth.

Table 3 presents the results for reported formal employment and incomes (from all sources,

⁷While this is a somewhat conservative measure of formal employment, paying contributions to health insurance is widely used as a measure of formal employment in Colombia (See Attanasio et al., 2017; Morales and Medina, 2017). The *Sisben* does not explicitly ask households whether members are in the formal sector.

Figure 2: Discontinuity in Formal Employment (2009).



RD Graph using optimal binning procedure discussed in [Calonico et al. \(2014a\)](#). Formal employment based on measures in 2009 *Sisben* survey. Subsample of males. Normalized *Sisben* (2002) score on horizontal axis centered around cutoff. Higher values represent lower scores (higher poverty).

not just salaries) in the *Sisben* survey. The table presents results for the reduced form change at the cutoff, and the two-staged least squares (2SLS) effect of enrolling in SR. These results show that the health insurance program had a negative impact of 4.2 percentage points (when using the optimal bandwidth in 2009) on the probability of being employed in the formal sector.⁸

Lower formal sector employment at the cutoff may be a combination of fewer youth joining the formal sector as they enter working-age, lower transition rates out of informal work, and higher transition probabilities out of formal work at the cutoff. As formal sector employment affects SR enrollment for the enrollee and their dependents, these are often family decisions, where older family members may discourage youth from joining the formal sector ([Joumard and Londono, 2013](#)). This effect is larger for men than it is for women (Appendix Table A2), perhaps once again suggesting that males have an outside option in organized crime. These differences may also reflect differences in formal sector options. Yet, we should note that data from the Gran Encuesta Integrada de Hogares (GEIH), shows that the formality rates (as a proportion of employed individuals, across all incomes groups) for men are about 52%, whereas

⁸Note, that we should not necessarily think of this result as a ‘first-stage’ on crime outcomes. Instead, crime and formal employment choices are jointly determined. Indeed, it is possible that, as we predict, the incentives lead individuals to leave the formal sector and join crime. After a few years in crime, an individual may wish to re-join the formal sector, but may be unable to do so given a criminal record.

Table 3: Reported Formal Employment and Income

Bandwidths:	In 2005			In 2009		
	4	6	10	4	6	10
Panel A: Formal Employment (Males)						
Above Cutoff	-0.00920**	-0.00652**	-0.00269	-0.0147***	-0.0111***	-0.00845***
Reduced Form	(0.00366)	(0.00285)	(0.00242)	(0.00467)	(0.00280)	(0.00217)
Enrolled in SR	-0.0349***	-0.0248**	-0.00984	-0.0551***	-0.0420***	-0.0308***
2SLS	(0.0133)	(0.0104)	(0.00868)	(0.0169)	(0.0105)	(0.00827)
Observations	162,942	221,761	304,895	133,067	180,742	247,886
Sample mean			0.07			0.14
Panel B: Annual Household Income (USD)						
Above Cutoff	-0.549	-0.0397	-0.870	-1.622	-0.758	8.295
Reduced Form	(1.275)	(0.998)	(0.616)	(2.376)	(2.253)	(7.780)
Enrolled in SR	-2.244	-0.161	-3.330	-6.680	-3.128	31.44
2SLS	(5.055)	(3.960)	(2.280)	(9.454)	(9.086)	(28.65)
Observations	64,963	88,358	122,105	46,797	63,457	87,510
Sample mean			112.00			171.24

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. We use the *Sisben* surveys of 2005 and 2009 to construct both outcome variables. Formal employment for males only. The results for women are presented in Table A2. ‘Enrolled in SR’ report Two-Stage Least Squares (2SLS) coefficients where the first stage is SR enrollment on being below the *Sisben* cutoff. Regressions control linearly for the *Sisben* score, flexibly around the cutoff. We cluster standard errors by *comuna*. Household-level income reported in *pesos* and converted to USD using the average 2009 exchange rate. Sample means for households in bandwidth.

for women are 46%; and formality rates as a fraction of the total population are 27.9% for men and 18.8% for women.

The impact on household-level income is statistically indistinguishable from zero and economically small (\$3.13 per household annually). Though we are wary of reading too much into impacts on self-reported income measures, these results suggest that even as workers drop out of the formal sector they find replacement sources of income. One caveat is that income is self-reported, and respondents may under-report assets and incomes in order to get a lower *Sisben* score. However, as respondents do not know the score formula, perfect manipulation is impossible, and so, as we show below, the density of respondents is smooth around the cutoff.

Note, if anything, we may particularly expect that incomes from illicit activities would be under-reported rather than over-reported. Accordingly, the absence of a negative impact

on income might suggest that desperation is less likely to be driving any increase in criminal activity we document below. Indeed, by revealed preference, they *choose* to be outside of the formal sector, and as such, should be better off. Nevertheless, we do not wish to emphasize the interpretation of these self-reported income results.

Canonical crime models (Becker, 1968; Ehrlich, 1973) stress the role of both income and substitution effects when wages in one sector change. In other analyses that focus on legitimate sector job-loss and unemployment shocks, the income and substitution effects work to both increase criminal activity. Interestingly, in contrast, here any gains from the subsidy essentially lower the likelihood of criminal activity. The results of this section confirm that the program encouraged informality among young men. The obvious question that this raises is how such discouragement of formal sector employment affects the likelihood of criminal activity.

7 Impacts on Crime

We turn our attention to outcomes on crime. One important distinction with the formal employment results is that we measure crime cumulatively over a decade. We interpret the impacts on crime as causally related to the incentives to leave the formal sector. Note that by the latter half of this period almost everyone had healthcare (under either one of the two regimes), and the benefits were similar. As such health benefits are not changing at the cutoff, only the incentives behind who pays for it changes. We show both the reduced form and 2SLS estimates of impacts on crime. In the second stage, we use the eligibility indicator as an instrument for enrollment status to identify the effect of SR enrollment on crime. Here $crime_i$ is 1 if the individual i was arrested between 2005 and 2013.⁹

$$Crime_{i,n} = \beta_0 + \beta_1 S\hat{R}_{i,n} + X'_{i,n}\beta_2 + A_i(s_{i,n})\beta_3 + \mu_n + \varepsilon_{i,n} ,$$

Our main results do not condition on other factors. In robustness checks, we control for various characteristics of the household head in 2002, the baseline year. These controls include

⁹Even as we have crime data for many years, we have formal employment only recorded at two points of time: in 2005 and 2009. This poses challenges when trying to simultaneously measure changes in employment and crime. Yet our results are robust to doing so (Appendix Table A12).

an indicator for female-headed households, employment status, years of education, marital status, attendance to any academic institution, year-of-birth fixed effects, socioeconomic strata of the household,¹⁰ home ownership, and neighborhood fixed effects.

A literature on neighborhood effects and crime (Cullen et al., 2006; Dustmann and Damm, 2014) highlight the perils of using area-based relationships (like differences in unemployment rates) to study individual-level occupational choice, and re-iterates the strength of our approach. Indeed, our theoretical model highlights how policing, returns to different types of employment, and non-monetary benefits may change across neighborhoods. There may still be general equilibrium effects of the policy that affect the entire country, but since our variation is not driven by differences across neighborhoods, unlike other studies, this is all netted out. Our results are unaffected by the inclusion of neighborhood fixed effects that absorb any neighborhood level characteristics (demographics, amenities, property values and police presence) that may affect crime rates. We cluster standard errors at the *comuna* level, but everywhere our results are robust to clustering at other geographic levels, like smaller neighborhoods (*barrios*), or households.

We present results for violent, property, and drug-related crimes, dividing each group between crimes “likely associated with criminal enterprises” (LACE) and those more reflective impulse or opportunity (non-LACE). We hypothesize that LACE crimes are reflective of an occupational choice across legitimate and illegitimate sectors, whereas non-LACE crimes should be less affected by incentives for remaining informal and hence serve as an effective falsification test. We expect the effects on the latter group to be zero, since as we posit in our model, crimes of impulse and passion are less directly related to occupational choice.

As we elaborate in a later discussion, over and above a falsification test, the lack of effects on non-LACE crimes also allows us to rule out alternative mechanisms. We do not classify crimes based on whether or not they are pecuniary, as that would capture crimes of desperation and necessity that arise out of poverty. Instead, we posit that the policy induced an *occupational choice* into crime, and as such use activities associated with criminal enterprises as a basis for classification. Alternative mechanisms (such as riskier behavior when having insurance) may

¹⁰Urban areas in Colombia are split into six socioeconomic strata, used by authorities to spatially target subsidies for piped water, sewage and electricity.

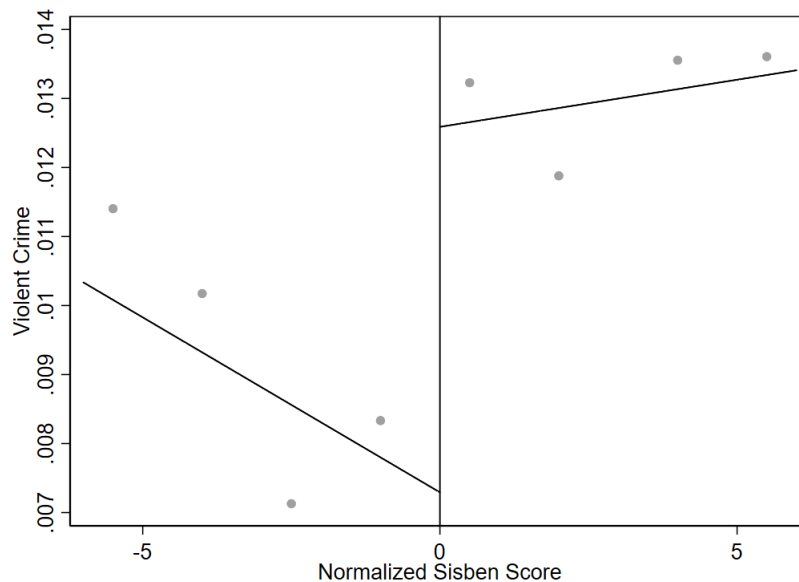
have weight if non-LACE crimes rose as well, but the lack of effects on non-LACE crimes allows us to rule them out.

We employ conservative outcome definitions: when looking at the impacts on violent crime, we exclude those whose first observed arrest was in property or drug crime, and do the same for each type of crime. Accordingly, the number of observations differs by type of crime. As such, our outcome will be 1 if the person’s first observed arrest was in violent crime, and 0 if they were never arrested in their youth. In robustness checks, we include the other types of crimes as 0s, and our results are simple more precisely estimated (see Appendix Table A10).

7.1 Violent Crime

We first start with the probability of being arrested for violent criminal activities. Based on the police flags for gang-related activity, violent LACE crimes include homicides, extortion, kidnapping, and firearms trafficking. Non-LACE violent crimes include domestic violence and rape. Figure 3 and Table 4 present the results.

Figure 3: LACE Violent Crimes



RD Graph using optimal binning procedure discussed in [Calonico et al. \(2014a\)](#). Normalized Sisben (2002) score on horizontal axis centered around cutoff. Higher values represent lower scores (higher poverty). LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations.

Figure 3 shows the jump in arrests for violent LACE crimes at the *Sisben* cutoff, concentrating on an optimal bandwidth of 6 points on the 100 point scale. In Table 4 we present the

Table 4: Violent Crimes

	Bandwidths:	4	6	10
Panel A: LACE Violent Crimes				
Above Cutoff Reduced Form		0.00722*** (0.00236)	0.00649** (0.00249)	0.00456** (0.00164)
Enrolled in SR No Covariates		0.0257*** (0.00873)	0.0231*** (0.00838)	0.0158*** (0.00539)
Enrolled in SR Including pre-treatment covariates		0.0274*** (0.00950)	0.0232** (0.00937)	0.0149** (0.00583)
Number of observations		18,052	24,272	33,027
Sample mean (men 13-26 years old in bandwidth)				0.014
Sample mean for those enrolled in SR and in high-gang comuna				0.020
Panel B: Non-LACE Violent Crimes				
Above Cutoff Reduced Form		0.00279 (0.00454)	0.000988 (0.00304)	-0.000581 (0.00326)
Enrolled in SR No Covariates		0.00994 (0.0158)	0.00349 (0.0104)	-0.00201 (0.0110)
Enrolled in SR Including pre-treatment covariates		0.00791 (0.0168)	0.00118 (0.0111)	-0.00322 (0.0125)
Number of observations		18,419	24,768	33,702
Sample mean (men 13-26 years old in bandwidth)				0.034
Sample mean for those enrolled in SR and in high-gang comuna				0.039

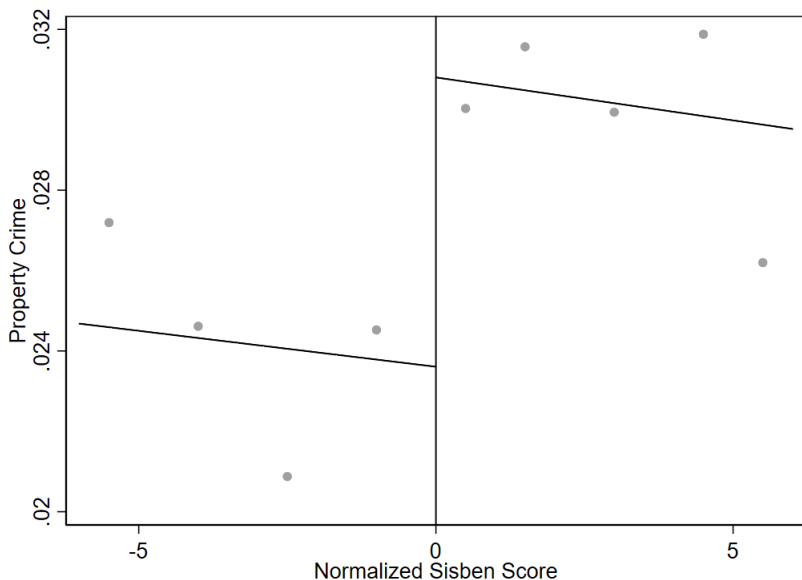
Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations. Non-LACE crimes are the remaining, more likely representing crimes of impulse or opportunity. Tables report reduced form and two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measured in 2002, and SR enrollment in 2005. We measure crime between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. We cluster standard errors by *comuna*. We consider only males who were between 13 to 18 years old in 2005. For regressions that have pre-treatment covariates, we include household characteristics, year of birth fixed effects, and neighborhood fixed effects. The sample excludes anybody whose first observed arrest was a property or drug crime (Appendix Table A10 includes these observations as a robustness).

regression discontinuity results varying the bandwidth and specifications. The reduced form results (first row in each panel) show an increase in LACE violent crime (Panel A), but no corresponding change in non-LACE violent crime (Panel B). Within a bandwidth of 10 points on the *Sisben* scale, and measuring arrests over a decade, these results amount to a 32% increase (or a 0.45 percentage point increase) in violent crime arrests from the mean around the cutoff. These magnitudes are both economically meaningful and similar to those from recent studies in this context (Khanna et al., 2020).

Our 2SLS results (next two rows of each panel) show an economically and statistically significant increase in the probability of arrest for LACE violent crimes for individuals enrolled in SR. We do not find any meaningful effect on the probability of arrest for non-LACE violent crimes. A comparison of the various rows in each panel shows that the estimates are robust to including controls, whereas a comparison across columns shows the robustness to bandwidths. The difference between coefficients for LACE and non-LACE crimes are statistically significantly different for all bandwidths.

7.2 Property Crime

Figure 4: LACE Property Crimes



RD Graph using optimal binning procedure discussed in Calonico et al. (2014a). Normalized Sisben (2002) score on horizontal axis centered around cutoff. Higher values represent lower scores (higher poverty). LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations.

Table 5: Property Crimes

	Bandwidths:	4	6	10
Panel A: LACE Property Crimes				
Above Cutoff Reduced Form		0.0106** (0.00387)	0.00930** (0.00389)	0.00666* (0.00350)
Enrolled in SR No Covariates		0.0380*** (0.0123)	0.0331*** (0.0126)	0.0232** (0.0113)
Enrolled in SR Including pre-treatment covariates		0.0408*** (0.0139)	0.0341*** (0.0131)	0.0240** (0.0108)
Number of observations		18,426	24,740	33,625
Sample mean (men 13-26 years old in bandwidth)				0.032
Sample mean for those enrolled in SR and in high-gang comuna				0.040
Panel B: Non-LACE Property Crimes				
Above Cutoff Reduced Form		-0.00263 (0.00554)	-0.00217 (0.00425)	-0.00205 (0.00336)
Enrolled in SR No Covariates		-0.00941 (0.0194)	-0.00772 (0.0149)	-0.00712 (0.0112)
Enrolled in SR Including pre-treatment covariates		-0.0116 (0.0212)	-0.00872 (0.0156)	-0.00854 (0.0119)
Number of observations		18,240	24,523	33,358
Sample mean (men 13-26 years old in bandwidth)				0.024
Sample mean for those enrolled in SR and in high-gang comuna				0.028

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations. Non-LACE crimes are the remaining, more likely representing crimes of impulse or opportunity. Tables report reduced form and two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measured in 2002, and SR enrollment in 2005. We measure crime between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. We cluster standard errors by *comuna*. We consider only males who were between 13 to 18 years old in 2005. For regressions that have pre-treatment covariates, we include household characteristics, year of birth fixed effects, and neighborhood fixed effects. The sample excludes anybody whose first observed arrest was a violent or drug crime (Appendix Table A10 includes these observations as a robustness).

In Figure 4 and Table 5 we analyze the effects on property crimes. LACE property crimes include motor vehicle theft and burglary of businesses and residences. Crimes like fraud and identify theft are classified as non-LACE. Once again, we see that LACE property crimes increase, with little change to non-LACE property crimes. The reduced-form estimate, over

the entire decade, constitutes a 21% increase (or a 0.66 percentage point increase) from the mean around the cutoff within a bandwidth of 10 points. In 2SLS results, we also find an economically and statistically significant increase for LACE property crime arrests, and no strong effect for property crimes less associated with organized entities. Again, our estimates are quite robust to the inclusion of controls and the choice of the bandwidth.

It is interesting to note that many non-LACE property crimes may also be income generating (even if they do not reflect occupational choices), and as such, impacts on these non-LACE crimes may be consistent with early economic models of crime (Becker, 1968). Yet, we find that it is the decision to join a criminal enterprise that seems to be the driving force. This is consistent with the anthropological interviews discussed above, which illustrate how gangs recruit idle youth and document that working for a gang is lucrative. The difference in coefficients between LACE and non-LACE crimes are statistically significantly different for all bandwidths.

7.3 Drug Crime

We analyze the impact on the probability to engage in drug-related crimes in Figure 5 and Table 6. LACE drug crimes include the manufacturing, distribution, and trafficking of hard drugs like cocaine and heroin. Non-LACE drug crimes include possession and consumption of marijuana, as these are mostly indicative of personal recreational use.

In Figure 5, even though the discontinuity in drug crime arrests is minor, there is a change in the slope of the relationship. The binned averages suggest a somewhat imprecise effect at the cutoff. In Table 6 the direction of effects are what we may expect, but our results are not precisely estimated. In Appendix Table A12, we use an alternative joint outcome definition of arrested and informal in 2009. Those results are precise enough to measure a significant increase in LACE drug crimes at the cutoff.

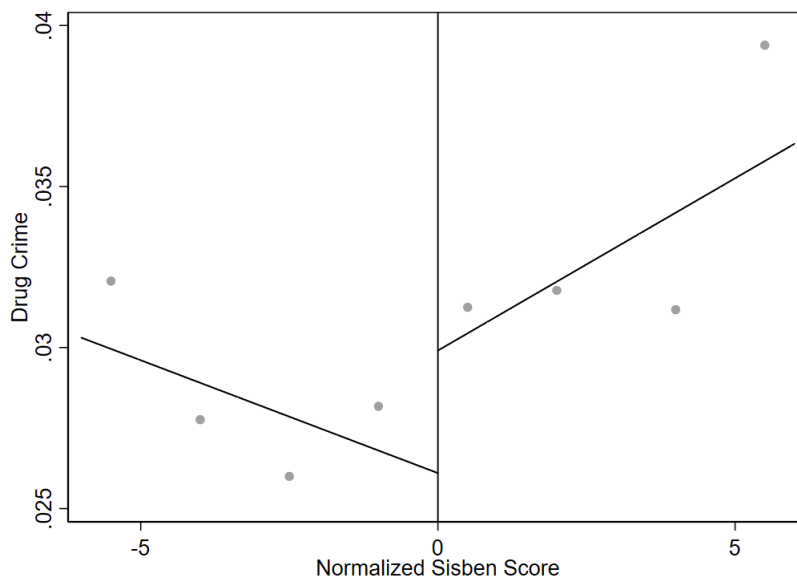
One possibility for the reduced precision in arrests for drug crimes is measurement error associated with the classification of such crimes. That is, the difficulty in classifying possession of drugs as consumption versus trafficking or distribution likely introduces noise for administrative data reasons. Indeed, offenses related to the trafficking of marijuana are problematic as small amounts of personal possession were not prosecuted during this period. While homicides,

Table 6: Drug Crimes

	Bandwidths:	4	6	10
Panel A: LACE Drug Crimes				
Above Cutoff Reduced Form		0.00799 (0.00721)	0.00348 (0.00492)	0.00133 (0.00458)
Enrolled in SR No Covariates		0.0285 (0.0240)	0.0124 (0.0169)	0.00461 (0.0155)
Enrolled in SR Including pre-treatment covariates		0.0303 (0.0270)	0.0135 (0.0180)	0.00524 (0.0159)
Number of observations		18,463	24,857	33,851
Sample mean (men 13-26 years old in bandwidth)				0.038
Sample mean for those enrolled in SR and in high-gang comuna				0.045
Panel B: Non-LACE Drug Crimes				
Above Cutoff Reduced Form		-0.00976 (0.00774)	-0.0129 (0.00798)	-0.00788 (0.00629)
Enrolled in SR No Covariates		-0.0348 (0.0280)	-0.0458 (0.0293)	-0.0274 (0.0218)
Enrolled in SR Including pre-treatment covariates		-0.0385 (0.0299)	-0.0501 (0.0329)	-0.0277 (0.0230)
Number of observations		19,150	25,740	35,104
Sample mean (men 13-26 years old in bandwidth)				0.073
Sample mean for those enrolled in SR and in high-gang comuna				0.088

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations. Non-LACE crimes are the remaining, more likely representing crimes of impulse or opportunity. Tables report reduced form and two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measured in 2002, and SR enrollment in 2005. We measure crime between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. We cluster standard errors by *comuna*. We consider only males who were between 13 to 18 years old in 2005. For regressions that have pre-treatment covariates, we include household characteristics, year of birth fixed effects, and neighborhood fixed effects. The sample excludes anybody whose first observed arrest was a property or violent crime (Appendix Table A10 includes these observations as a robustness).

Figure 5: LACE Drug Crimes



RD Graph using optimal binning procedure discussed in Calonico et al. (2014a). Normalized Sisben (2002) score on horizontal axis centered around cutoff. Higher values represent lower scores (higher poverty). LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations.

assaults and theft involve victims as clear evidence of crimes, drug crimes are often difficult to detect and record. Not having any evidence of a crime actually being committed (e.g., a victim) may also allow authorities to under-report, especially if cartels pressure authorities to do so.

In sum, our results indicate that the drop in formal employment as a result of the subsidized benefits for informal workers raised the likelihood of being arrested for LACE violent and property crimes. The magnitudes of the estimated impacts are also economically meaningful. The pattern of results is similar but imprecise for drug crimes. Importantly, the results also show that non-LACE crimes of each type are not impacted by SR enrollment, ruling out many alternative mechanisms. In the following section, we investigate whether impacts are strongest in *comunas* that were historically associated with high organized crime activity as further evidence in support of our occupational choice interpretation.

8 Heterogeneity, Specification Tests and Robustness

8.1 Heterogeneity by Comuna: the Importance of Neighborhoods

Previous studies have emphasized that the opportunities in a neighborhood affect how easy it is to induce youth into crime (Kling et al., 2005). Understanding the heterogeneity by neighborhood helps us speak to much of the literature which relies on area-based variation. Consistent with our theoretical framework, high crime neighborhoods may have more policing and higher detection rates that may lower the employment-crime elasticity, but may also have more opportunities to join a gang and thereby raise the elasticity.

We investigate if *comunas* with a high incidence of gangs demonstrate stronger impacts on LACE arrests at the RD cutoff. If the policy induces men to join organized crime, then we may expect that neighborhoods that have more such opportunities would have a larger impact. Figure A3 shows the spatial distribution of the locations where criminals were arrested in the act between 2005 and 2013, by type of crime. In our main results we already show specifications that include neighborhood fixed effects, and we cluster errors at spatial levels larger than neighborhoods. Our results are robust to clustering at smaller spatial levels, like the neighborhood.

We select the five *comunas* with the highest number of gang members captured by the police (i.e., most arrests with gang-flags as a ratio of total crimes), and create an indicator variable for whether individuals lived in these *comunas* in 2002, our baseline year. These are not necessarily high crime areas, as the mean arrest rates for young men is 18% in both gang and non-gang *comunas*. Yet, LACE crimes make up 43% of arrests in gang *comunas*, and 37% of arrests in non-gang *comunas*.

We interact this variable with the cutoff to analyze the heterogeneity in effects by area-level gang activity. Table 7 presents the results. Since we have an interaction term, we report the IV first stage F-statistics as well. The effects on crime are present in both high and low gang activity areas, but for property crime are larger in areas that have more gang activity. For violent crime the interaction term is strongly positive for the optimal bandwidth, but less robust than results for property crimes.

Table 7: Heterogeneity by Comuna

	Bandwidths:	4	6	10
Panel A: LACE Violent Crimes				
Enrolled in SR		0.0267*** (0.00892)	0.0211** (0.00914)	0.0150*** (0.00538)
Enrolled* Gang Comuna		-0.00152 (0.00464)	0.0141*** (0.00376)	0.00563 (0.00537)
F stat		90.2	154.6	232.9
Number of observations		18,052	24,272	33,027
Panel B: LACE Property Crimes				
Enrolled in SR		0.0344** (0.0134)	0.0273** (0.0137)	0.0190* (0.0115)
Enrolled* Gang Comuna		0.0282 (0.0209)	0.0364** (0.0167)	0.0258** (0.0116)
F stat		86.7	145.5	249.6
Number of observations		18,426	24,740	33,625
Panel C: LACE Drug Crimes				
Enrolled in SR		0.0282 (0.0248)	0.0131 (0.0177)	0.00296 (0.0166)
Enrolled* Gang Comuna		0.000310 (0.0136)	-0.00590 (0.0135)	0.00690 (0.0129)
F stat		96	149.1	201.7
Number of observations		18,463	24,857	33,851

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff and an interaction between high-gang comunas and being below the cutoff. The Sisben score is measure in 2002, SR enrollment in 2005, and crime outcomes are measured between 2005 and 2013. Regressions include comuna fixed effects and an interaction between high-gang comunas and indicators for SR enrollment. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males who were between 13 to 18 years old in 2005. We cluster errors by comuna. The mean arrest rate across all five gang comunas are 18%, which is also the mean arrest rate in non-gang comunas. See Appendix Table A3 for the non-LACE crimes.

Appendix Table A3 shows the results for non-LACE crimes. Again, there is no evidence of SR enrollment being associated with non-LACE crimes in either gang *comunas* or non-gang *comunas*. Notice that our identification strategy protects against increases in policing activity in gang *comunas*, as we are comparing one side of the *Sisben* cutoff to the other. Additionally, throughout the paper our tables show a row of results that include *comunas* fixed effects.

8.2 Density Tests and Balance Tests

In our study, identification relies on the assumption that all other determinants of the outcome vary smoothly at the cutoff. The SR *Sisben* cutoff did not coincide with other interventions, at the eligibility cutoff of *Sisben* during the study period in Medellín. Importantly, we show that an extensive set of observables display no systematic patterns in discontinuities. In Table A4, we show that baseline characteristics from 2002 (three years before our crime data begins) are balanced for the entire sample and for the sample used in the regressions (young males), respectively. We consider two sets of baseline characteristics: one for household-level socioeconomic variables, and the other for individuals. We find no evidence of systematic discontinuities in covariates at the threshold. In the first row of Table A4, we report a summary measure in which we collapse these variables by taking their first principal component and repeat the same RD analysis that we do for our main results. Again, there is no detectable difference in this composite measure of baseline characteristics. In all our regression tables we show effects both with and without these variables as controls. Finally, at the bottom of Table A4 we show that the match rate across *Sisben* waves was high (about 94% for the sample).

Additionally, for the empirical strategy to be valid, households must not be able to manipulate their score to cross the cutoff. Work by Camacho and Conover (2011) highlights politically motivated manipulation in 1998 in other parts of Colombia during mayoral elections. This includes both under-reporting of wealth (not necessarily a threat to our design), but also manipulation of the score. We use the raw survey data and the 2002 *Sisben* score only for Medellín, and so are not concerned with any manipulation of the final score. Indeed, our tests of balance in the large set of baseline characteristics of the household are indicative of a lack of systematic manipulation in this context. Importantly, the cutoff for SR eligibility was determined well

after the 2002 *Sisben* scores were released, and as such the cutoffs were not known to anybody during the survey. As such, we are confident that nobody could manipulate the 2002 score, even if they joined a life of crime in later years.

We test whether there was a discontinuity in the density of scores at the cutoff for the particular context of Medellín after 2002. We do this by following two methods used in the literature: the McCrary (2008) test and a test recently developed by Cattaneo et al. (2017). The Cattaneo et al. (2017) test yields a conventional t-statistic of 0.0489 or a p-value of 0.961, and a robust bias corrected p-value of 0.940, confirming that there is no statistically detectable evidence of manipulation. Figure A4 shows the distribution of the *Sisben* score for males (non-criminals and criminals) and for the full sample, conducts McCrary (2008) tests, and shows a closeup of the distribution around the cutoff. The distribution is smooth with no evidence of bunching before the cutoff between *Sisben* levels 2 and 3 (red line).

8.3 Robustness

Alternative Crime Classifications: We conduct a number of robustness checks. First, we re-classify crimes into LACE and non-LACE groups based on the *location* where these types of crimes are more likely to occur. We calculate the relative propensity of each crime in each neighborhood. The crimes that have a higher propensity to take place in neighborhoods that were traditionally associated with organized crime are classified as LACE crimes. These are neighborhoods that also have the highest proportion of gang flags associated with them. To be specific, we sort the crimes by the fraction of first observed arrests that happen in a high-gang neighborhood. The top half of this list is classified as LACE crimes.

This Neighborhood Classification Method of crimes produces a list similar to the one where we use the police generated flags, with minor differences. The only big difference among the top 25 crimes is with “conspiracy to commit murder,” which is classified as LACE under the original classification but not LACE under the neighborhood definition. Any other differences appear outside the 25 most prevalent crimes. The lists of the most prevalent crimes by classification method can be found in Appendix Table A1. In Table A5, we re-examine our main results using the alternative classification for LACE crimes. These results are similar, but with the

added statistical significance of drug crimes under some specifications.

We do another machine-learning exercise to classify crimes based on the police flags and the characteristics of each crime, by predicting the probability of LACE crimes based on these observable features. We then re-weight the outcomes based on these probabilities. To be specific, we first train our data on the sample that has the police classification, and use k-means logit to predict the likelihood that the crime committed was a LACE crime. We use neighborhood, crime and individual characteristics to predict LACE crimes as an outcome. This first step provides us with a weight for each crime, where the weight captures the propensity that the crime was LACE. Then, we re-run our regressions, this time weighting by the propensity to be a LACE crime. We report the predictive power of this process with the help of the mean squared error and pseudo R2. Appendix Table A6 shows qualitatively similar patterns to our main results.

RD Bandwidths, Nonparametrics and Inference: Next, we re-examine our main results using the bias-correction methods suggested by Calonico et al. (2014a). In Table A7, we show results that conduct a polynomial bias correction at a larger bias-correction bandwidth (reported in the table). Once again, our results show an economically and statistically significant increase in LACE violent and property crimes, but the effects on drug crimes are small and imprecise.

We also test robustness to various ways to aggregate the sample and conduct inference at the household level. In Appendix Table A8, we cluster standard errors at the household level, and in Table A9 we conduct the entire analysis after aggregating to the household level. Our results remain similar to before.

Alternative Samples and Specifications: In our main specifications, when looking at a specific type of crime, we exclude arrests from other crimes. For instance, when studying violent crimes, we exclude property and drug crimes from the sample altogether. In the specification shown in Appendix Table A10, we include the other categories along with the non-criminals; and in Appendix Table A11, we include repeat arrests. Our results are similarly robust to these alternatives. In Appendix Figure A5, we vary the bandwidth through a much wider range – every integer between 2 and 10. LACE violent and property crimes consistently display

a positive RD coefficient, whereas drug crimes are not statistically indistinguishable from zero, even as the coefficients are positive and fairly large for smaller bandwidths.

We also examine the robustness to varying the cohorts under analysis. In our main specifications, we follow the cohort of 13-18 year-olds in 2005 till 2013. In Appendix Figure A6, we expand the range of ages under study, examining impacts on those as young as 10 in 2005 (or 18 in 2013). We are reluctant to study older ages as we would not have previous arrests for them if they were ever arrested before 2002.

Finally, as our hypotheses involve the intersection of both informality and violent crime, we present a specification in Appendix Table A12 that simultaneously captures both. The advantage of our individual-level data is that we can measure both employment and criminal behavior for the same individual. However, since we do not have annual data on formal employment, we use the 2009 *Sisben* to measure formal employment. Here the dependent variable is an indicator that equals one when the individual was not formally employed and arrested for a crime, and zero otherwise. Our results again show an increase in LACE criminal activity, with even the drug crimes now being economically and statistically significant. This result allows us to address any concerns that the increase in informality and increase in arrests were coincidental.

Additional Controls for Pre-treatment Covariates and Community Characteristics: Our theory highlights that crimes are not randomly allocated across communities, and so we may also consider including further community-level controls. While we already show results with neighborhood fixed effects, we can also do robustness checks controlling for pre-treatment covariates and community characteristics interacted with the treatment indicator. These characteristics include the share formally employed, and average wages for formal sector workers in the neighborhood. Tables A13 to A15 in the appendix show our results are robust to such controls.

8.4 Human Capital Investments and Family Structure

We conduct two additional analyses which help to support the intuition of the proposed mechanisms. First, we examine changes to investments in human capital. One possible implication

of eligibility for SR would be that individuals are less likely to acquire education if they are unlikely to join the formal sector. We link individual-level schooling data (*Matricula en Linea*) to our *Sisben* survey and examine school enrollment for our cohorts. Appendix Table A16 shows that school enrollment is indeed discontinuously lower for youth eligible for SR.

Second, we may expect that the family structure determines the sensitivity to SR eligibility. For instance, we may expect that the results are muted for individuals who have adult family members already in the formal sector, as those individuals are unlikely to be eligible for SR (if they are dependents), and so less likely to change employment decisions. In additional analysis, we fail to detect any impact on youth who have an adult family member working in the formal sector. Though encouraging, we do not present these results here for the sake of brevity and because we believe one should interpret these results with caution, as formal status is an endogenous choice.

9 Alternative Mechanisms

The simultaneous decrease in formal sector employment and rise in LACE-related arrests supports our model in which discouragement from formal employment induces an occupational choice into a life of crime (Becker, 1968). While we do not discuss in detail specific pathways, anthropological evidence lends credence to active recruitment by gang members of young men that ‘hang around’ in neighborhoods with idle time, and are not in the formal sector (Baird, 2011). Indeed, the results for joint informality and LACE arrest outcomes, presented in Appendix Table A12 and discussed above, validate the interpretation of our results as being driven by individual-level occupational choice.

However, we discuss three of the most likely alternative theories in light of the full pattern of empirical results. First, better health benefits at the cutoff may induce one to engage in riskier behaviors. However, it is difficult to support why these riskier behaviors would not also (or even primarily) include non-LACE crimes (e.g., drug consumption). Furthermore, as health coverage is near universal by the end of this period, individuals on both sides of the cutoff have nearly identical coverage, with the only difference being who pays. If anything, any early differences in coverage, favored better care for formal employees, which would lead to effects of

the opposite sign of what we find. This suggests that the driving force is the cost of coverage under formal sector employment. Second, formal workers vesting more into the health system may fear losing their jobs if arrested, reducing their criminal activities as a result. However, again this should be just as true for non-LACE crimes.

Finally, the police may falsely target informal workers even if they are not criminals, but it is unlikely that they would be booked disproportionately under LACE crimes. Indeed, it may be easier to falsely target potential criminals for petty crime rather than more serious offenses like homicide or auto theft. The distinction between LACE and non-LACE crimes powerfully helps exclude alternative mechanisms and lends credence to the occupational choice story we posit. Additionally, police are unlikely to know a person's Sisben score or formal status. Importantly, 91% of the arrests occur *in flagrante* (in the act) and so are unlikely to be associated with individual characteristics. In Table A17 we demonstrate robustness of all our main results when restricting the sample to the subset of *in flagrante* crimes.

Another possibility we consider is that arrests may be driving the fall in formalization. While the decision to drop out of the formal sector and join a criminal enterprise is likely to be made simultaneously (rather than staggered over time), we take the order of events seriously. We try to think of possibilities that may first lead SR eligibility to drive arrests (and then occupational choice), but have difficulty in doing so. For instance, we would need to think of alternatives where SR drives arrests for reasons *unrelated* to occupational choice. While without fine-grained variation over time, it is challenging to unpack the order of events (first occupational choice, and then arrests), the results for joint informality and LACE arrest outcomes, presented in Table A12 and discussed above, help suggest that our results are being driven by individual-level occupational choice. This discussion emphasizes how difficult it is to find alternative explanations that reconcile the full pattern of empirical results, including particularly the lack of effects for non-LACE crimes.

10 Interpretation and Magnitudes

Our estimates here suggest that for 100 male youth that are less likely to be formally employed, 12 are later arrested for a crime at least once over the subsequent 9 year period.¹¹ As a reminder, the raw data shows that on average 12.5% of male youth are arrested at some point. The model describes this parameter as an occupational-choice elasticity.

The crime-formal employment elasticity we obtain from our results is roughly 1.93, and the crime elasticity with respect to the wage costs is about 3.31.¹² A recent review by [Bennett and Ouazad \(2018\)](#) discusses how a 1pp increase in unemployment usually corresponds to a 3-7% increase in crime. In other work on Medellín, we show that for mass-layoffs among adults across all genders, a 1% increase in unemployment raises arrests by 2.12%, and over the medium-run, the elasticity with respect to changes in earnings is about 2.02 ([Khanna et al., 2020](#)).

Note, however, that the parameter we estimate in this paper differs slightly from others in the literature. [Becker \(1968\)](#) discusses both income and substitution effects when wages rise in one sector. For much of the other literature, job loss or unemployment in the formal sector may produce income and substitution effects that both accentuate criminal activity. For instance, losing one's job reduces not just the expected returns to formal jobs (encouraging workers to substitute away from the formal sector), but also lowers their income directly, and the income loss may drive individuals into crime. In contrast, we may think that the gains from the subsidy in our case (income effect) may actually reduce criminal activity. Additionally, while we are directly testing the choice of occupation, most of the related work studies reduced access to legitimate jobs as a consequence of job losses or work restrictions, and thereby recover different elasticities. For example, job losses and structurally imposed employment restrictions may additionally induce effects on depression, subsequent job search, social stigma etc., whereas our

¹¹We compare a 4.2pp (IV-2SLS) drop in formality for the high-risk group (male youth) based on the number of formal-sector workers in 2002, and the 4.3pp (IV-2SLS) increase in crime rates for the high-risk group (4.3pp is the sum of the 2SLS coefficients across all violent, drug, and property crimes over the 9 year period). For the baseline crime-arrest rates, we use the average between 2002 and 2005.

¹²For the crime-formal employment occupational-switching elasticity, we first calculate the overall percentage change in crime across all three LACE categories: violent property and drug, by taking the sum of the 2SLS coefficients (0.0435) with respect to the sum of the baseline arrest rates (0.105). We then do the same for the employment change (a 21% change) to get the denominator of the employment-crime elasticity. For the elasticity with respect to wage costs, in the denominator, we use the 12.5% of wages formal workers must pay for SR.

variation in relative costs of employment should not, as individuals here are choosing to leave the formal sector.

As our estimates are similar to those obtained from other recent causal analysis leveraging individual-level variation (Khanna et al., 2020; Rose, 2019), we interpret our results to be not only economically meaningful, but also plausible. Yet, these magnitudes should be understood to be context-specific. We study a high-crime environment, similar only to other developing countries and especially Latin America. Recent evidence from Latin America suggests much larger employment-crime elasticities (Dell et al., 2018; Dix-Carneiro et al., 2018).

Furthermore, we estimate a Local Average Treatment Effect (LATE) on marginal workers in the neighborhood of an income cutoff. It is plausible that at higher income levels, healthcare is a less important fraction of expenditures, and is less likely to induce such behavior. Finally, it should be noted that the newly induced marginal criminals may be unlike the average criminal along many dimensions, including skill for avoiding arrests, and as such our results may not be widely generalizable for other sub-populations. Nevertheless, since the exogenous probability of getting caught conditional on committing a crime has no reason to be discontinuous at the cutoff, our estimates are unbiased even in the presence of such heterogeneity in criminal “skill.”

11 Conclusion

In this paper, we argue that disincentivizing formal employment can lead to substantial increases in criminal activity when informal opportunities include employment by criminal enterprises. We evaluate this claim in the context of the high-crime environment of Medellín, Colombia. We first provide strong evidence showing that the criteria behind the health benefits policy unintentionally led to a sharp decrease in formal sector employment. At the margin, the policy encouraged workers to remain in the informal labor market.

In Medellín, this informal market contains significant opportunities related to organized crime. We follow these same individuals over a decade and show that this decrease in formal sector employment led to an increase in the probability of being arrested for criminal-enterprise related activities. On the other hand, crimes less likely to be associated with criminal economic enterprises, like crimes of impulse or opportunity, show no such impacts at the eligibility

threshold, lending credence to the occupational choice mechanism we advance. Together, our simple calculations suggest that as the policy pushed workers out of the formal sector, a meaningful fraction of these workers (i.e., more than 1 in 10 among young males) were drawn into criminal enterprises. These effects were largest in neighborhoods that had, at baseline, greater opportunities to join organized crime.

Crime deterrence may have limited benefits if the supply elasticity to criminal activity is high (Freeman, 1999). Investigating the decisions behind choosing a life of crime, as we do here, is essential in the fight against crime. Importantly, our work speaks to the occupational sorting dimension of the criminal participation decision in response to economic incentives. We do this using a unique data set that matches the census of arrests with socio-economic outcomes over a decade, in the context of one of the most violent cities in the world. We find a source of exogenous variation generated by policy rules, and use an RD design to estimate our effects.

We conclude that Colombia's well-intentioned and broad-based subsidies for healthcare had the unintended consequence of incentivizing participation in criminal enterprise by way of its distortionary provision rules. The program being important for providing subsidized health access to low income families implies that there is little reason to do away with it. Yet, the formality-clause governing the selection into the program is distortionary, and as such warrants examination. Recognizing these adverse effects, policy-makers lowered the costs of CR enrollment at the end of our study period, when Law 1607 was enacted, leading to a significant increase in formal sector employment (Bernal et al., 2017; Kugler et al., 2017; Morales and Medina, 2017).

Removing the emphasis on informality (but still targeting the poor) may negate the increase in criminal activity around the cutoff. The costs underlying such a change would be a larger fiscal burden as even low-income formal sector workers would be eligible for SR. The benefits are far reaching: less crime, less policing and incarceration, and fewer negative externalities on families and children. This has welfare implications for the design of many programs across the developing world which often have far-reaching and under-studied consequences on seemingly unrelated outcomes and behaviors. Our results provide guidance for how impactful improving access to and incentives for formal sector employment can be for deterring criminal activity.

References

- Agan, A. and S. Starr (2018). Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment. *The Quarterly Journal of Economics* 133(1), 191–235.
- Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015). The economics of density: Evidence from the berlin wall. *Econometrica* 83(6), 2127–2189.
- Alesina, A., S. Piccolo, and P. Pinotti (2017). Organized Crime, Violence, and Politics. *NBER Working Paper*.
- Arteaga, C. (2019). The Cost of Bad Parents: Evidence from the Effects of Parental Incarceration on Children’s Education. *Working Paper*.
- Attanasio, O., A. Guarín, C. Medina, and C. Meghir (2017). Vocational Training for Disadvantaged Youth in Colombia: A Long-Term Follow-up. *American Economic Journal: Applied Economics* 9(2), 131–43.
- Baird, A. (2011). Negotiating Pathways to Manhood: Violence Reproduction in Medellín’s Periphery. *Thesis, University of Bradford*.
- Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76(2), 169–217.
- Bell, B., A. Bindler, and S. Machin (2018). Crime scars: Recessions and the making of career criminals. *Review of Economics and Statistics* 100(3), 392–404.
- Bennett, P. and A. Ouazad (2018). Job Displacement, Unemployment, and Crime: Evidence from Danish Microdata and Reforms. *Journal of the European Economic Association*. forthcoming.
- Bergolo, M. and G. Cruces (2021). The anatomy of behavioral responses to social assistance when informal employment is high. *Journal of Public Economics* 193.
- Bernal, R., M. Eslava, M. Melendez, and A. Pinzon (2017). Switching from Payroll Taxes to Corporate Income Taxes: Firms’ Employment and Wages after the Colombian 2012 Tax Reform. *Inter-American Development Bank No. 1268*.
- Bhuller, M., G. Dahl, K. Loken, and M. Mogstad (2018). Incarceration, Recidivism and Employment. *NBER Working Paper No. 22648*.
- BJS, U. B. o. J. S. (1994). Sourcebook of Criminal Justice Statistics. *University at Albany*.
- Blattman, C. and J. Annan (2015). Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State. *American Political Science Review* 10(1), 1–17.
- Blattman, C., G. Duncan, L. Benjamin, and S. Toben (2018). Gangs of Medellín: How Organized Crime is Organized. Technical report.
- Blattman, C., J. C. Jamison, and M. Sheridan (2017). Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia. *American Economic Review* 107(4), 1165–1206.
- Bosch, M. and N. Schady (2019). The effect of welfare payments on work: Regression discontinuity evidence from ecuador. *Journal of Development Economics* 139, 17–27.
- Britto, D. G., P. Pinotti, and B. Sampaio (2020). The effect of job loss and unemployment insurance on crime in brazil. *CEPR Discussion Paper DP14789*.
- Brown, R. and A. Velasquez (2017). The Effect of Violent Crime on the Human Capital Accumulation of Young Adults. *Journal of Development Economics* 127, 1–12.
- Buonanno, P. and J. Vargas (2018). Inequality, crime, and the long run legacy of slavery. *Journal of Economic Behavior and Organization*. forthcoming.

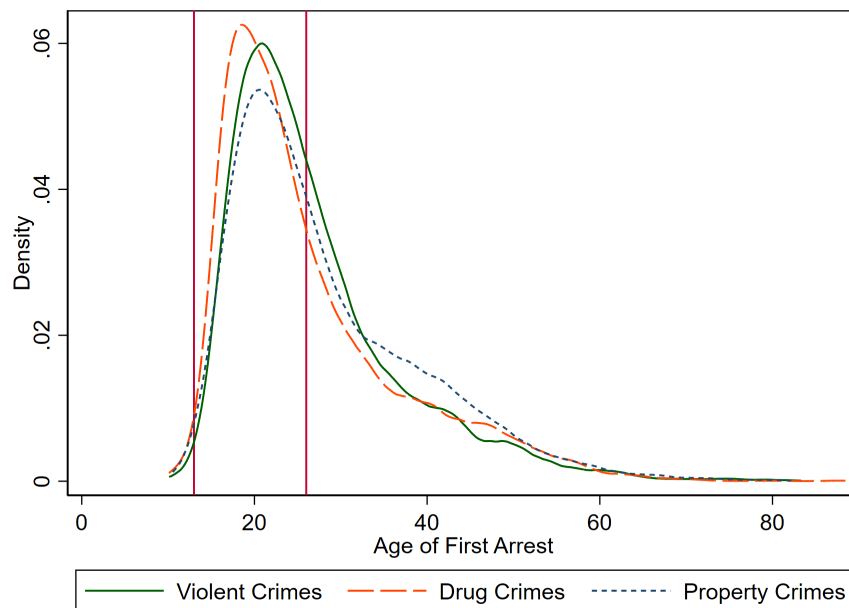
- Calonico, S., M. Cattaneo, and R. Titiunik (2014a). Robust Data-Driven Inference in the Regression Discontinuity Design. *Stata Journal* 14(4), 909–946.
- Calonico, S., M. Cattaneo, and R. Titiunik (2014b). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82(6), 2295–2326.
- Camacho, A. and E. Conover (2011). Manipulation of Social Program Eligibility. *American Economic Journal: Economic Policy* 3(2), 41–65.
- Camacho, A., E. Conover, and A. Hoyos (2014). Effects of Colombia’s Social Protection System on Workers’ Choice between Formal and Informal Employment. *World Bank Economic Review* 28(3), 446–466.
- Cattaneo, M., M. Jansson, and X. Ma (2017). Simple Local Regression Distribution Estimators with an Application to Manipulation Testing. *Mimeo Michigan*.
- CCSPJP (2009). Consejo Ciudadano para la Seguridad Publica y la Justicia Penal.
- Chimeli, A. B. and R. R. Soares (2017). The use of violence in illegal markets: Evidence from mahogany trade in the brazilian amazon. *American Economic Journal: Applied Economics* 9(4), 30–57.
- Chioda, L., J. De Mello, and R. R. Soares (2016). Spillovers from Conditional Cash Transfer Programs: Bolsa Familia and Crime in Urban Brazil. *Economics of Education Review* 54, 306–320.
- Cook, P. J., S. Kang, A. A. Braga, J. Ludwig, and M. E. O’Brien (2015). An experimental evaluation of a comprehensive employment-oriented prisoner re-entry program. *Journal of Quantitative Criminology* 31, 355–382.
- Cortes, D., J. Santamaria, and J. Vargas (2016). Economic shocks and crime: Evidence from the crash of ponzi schemes. *Journal of Economic Behavior and Organization* 131, 263–275.
- Cullen, J. B., B. A. Jacob, and S. D. Levitt (2006). The Effect of School Choice on Participants: Evidence from Randomized Lotteries. *Econometrica* 74(5), 1191–1230.
- Davis, J. M. and S. B. Heller (2020). Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. *The Review of Economics and Statistics* 102(4), 664–677.
- de Brauw, A., D. O. Gilligan, J. Hoddinott, and S. Roy (2015). Bolsa familia and household labor supply. *Economic Development and Cultural Change* 63(3), 423–457.
- Dell, M., B. Feigenberg, and K. Teshima (2018). The Violent Consequences of Trade-Induced Worker Displacement in Mexico. *American Economic Review: Insights forthcoming*.
- Deshpande, M. and M. Mueller-Smith (2021). Does welfare prevent crime? the criminal justice outcomes of youth removed from ssi. *Working paper*.
- DiTella, R., S. Galiani, and E. Schargrodsky (2010). Crime distribution and victim behavior during a crime wave. In *The economics of crime: Lessons for and from Latin America*, pp. 175–204. University of Chicago Press.
- Dix-Carneiro, R., R. Soares, and G. Ulyssea (2018). Economic Shocks and Crime: Evidence from the Brazilian Trade Liberalization. *American Economic Journal: Applied Economics forthcoming*.
- Doyle, C. (2016). Explaining Patterns of Urban Violence in Medellin, Colombia. *Laws* 5(3).
- Dustmann, C. and A. P. Damm (2014). Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior? *American Economic Review* 104(6), 1806–1832.
- Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica* 70(5), 1741–1779.

- Ehrlich, I. (1973). Participation in Illegitimate Activities: A Theoretical and Empirical Investigation. *Journal of Political Economy* 81(3), 521–565.
- Fernández, C. and L. Villar (2017). The Impact of Lowering the Payroll Tax on Informality in Colombia. *Economía* 18(1), 125–155.
- Freeman, R. B. (1999). The Economics of Crime. *Handbook of Labor Economics* 3(c). edited by O. Ashenfelter and D. Card. Elsevier Science.
- Gaviria, A., C. Medina, and C. Mejia (2007). Assessing Health Reform in Colombia: From Theory to Practice. *Economía* 7(1).
- Gelber, A., A. Isen, and J. B. Kessler (2016). The effects of youth employment: Evidence from new york city lotteries. *The Quarterly Journal of Economics* 131(1), 423–460.
- Gerard, F. and G. Gonzaga (2021). Informal labor and the efficiency cost of social programs: Evidence from unemployment insurance in brazil. *American Economic Journal: Economic Policy* 13(3), 167–206.
- Gronqvist, H. (2017). Youth Unemployment and Crime: Lessons from Longitudinal Population Records. *Mimeo*.
- Guarin, A., C. Medina, and J. Tamayo (2013). The Effects of Punishment of Crime in Colombia on Deterrence, Incapacitation and Human Capital Formation. *Borradores de Economía* 774.
- Hahn, J., P. Todd, and W. van der Klaauw (2001). Identification and Estimation of Treatment Effects with a Regression Discontinuity Design. *Econometrica* 69(1), 201–209.
- Heller, S. (2014). Summer Jobs Reduce Violence among Disadvantaged Youth. *Science* 346(6214), 1219–1223.
- Hjalmarsson, R. and M. Lindquist (2019). The Causal Effect of Military Conscription on Crime. *The Economic Journal* 129(622), 2522–2562.
- Ibanez, A. M., A. Ritterbusch, and C. Rodriguez (2017). Impact of the Judicial System Reform on Police Behavior: Evidence on Juvenile Crime in Colombia. *Mimeo*.
- Ihlanfeldt, K. (2007). Neighborhood Drug Crimes and Young Males Job Accessibility. *The Review of Economics and Statistics* 89(1), 151–164.
- Imbens, G. W. and K. Kalyanaraman (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *The Review of Economic Studies* 79(3), 933–959.
- Joumard, I. and J. Londono (2013). Income Inequality and Poverty in Colombia: The Role of the Labour Market. *OECD Economics Department Working Papers No. 1036* (1).
- Kearney, M. S., B. Harris, E. Jacome, and L. Parker (2014). Ten Economic Facts about Crime and Incarceration in the United States. *The Hamilton Project Policy Memo*. Brookings Institute.
- Khanna, G., C. Medina, A. Nyshadham, C. Posso, and J. Tamayo (2020). Job Loss, Credit and Crime in Colombia. *American Economic Review: Insights*. Forthcoming.
- Khanna, G., C. Medina, A. Nyshadham, D. Ramos-Menchelli, J. Tamayo, and A. Tiew (2022). Spatial mobility, economic opportunity, and crime. Technical report.
- Kling, J., J. Ludwig, and L. Katz (2005). Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment. *The Quarterly Journal of Economics* 120(1), 87–130.
- Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica* 75(1), 83–119.
- Kugler, A. and M. Kugler (2009). Labor Market Effects of Payroll Taxes in Developing Countries: Evidence from Colombia. *Economic Development and Cultural Change* 57(2), 335–358.

- Kugler, A., M. Kugler, and L. Herrera Prada (2017). Do Payroll Tax Breaks Stimulate Formality? Evidence from Colombia's Reform. *NBER Working Paper No. 23308*.
- Lamprea, E. and J. Garcia (2016, December). Closing the Gap Between Formal and Material Health Care Coverage in Colombia. *Health and Human Resources Journal*.
- Mastrobuoni, G. and P. Pinotti (2015). Legal status and the criminal activity of immigrants. *American Economic Journal: Applied Economics* 7(2), 175–206.
- McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics* 142(2), 698–714.
- Medina, C., J. Núñez, and J. Tamayo (2013). The unemployment subsidy program in colombia: An assessment. Technical report.
- Medina, C. and J. Tamayo (2011). An Assessment of How Urban Crime and Victimization Affects Life Satisfaction. in *Dave, Webb and Eduardo, Wills-Herrera (Eds.), Subjective Well-Being and Security*. Springer, Social Indicators Research Series.
- Melnikov, N., C. Schmidt-Padilla, and M. M. Sviatschi (2019). Gangs, labor mobility, and development. *Working Paper*.
- Morales, L. and C. Medina (2017). Assessing the Effect of Payroll Taxes on Formal Employment: The Case of the 2012 Tax Reform in Colombia. *Economia Journal* 0, 75–124. LACEA.
- Munyo, I. and M. A. Rossi (2015). First-day criminal recidivism. *Journal of Public Economics* 124, 81–90.
- Pinotti, P. (2017). Clicking on Heaven's Door: The Effect of Immigrant Legalization on Crime. *American Economic Review* 107(1), 138–168.
- Rose, E. (2019). The Effects of Job Loss on Crime: Evidence From Administrative Data. *Working Paper*.
- Rozema, R. (2018). Urban DDR-processes: Paramilitaries and Criminal Networks in Medellin, Colombia. *Journal of Latin American Studies* 40, 423–452.
- Santamaria, M., F. Garcia, and A. V. Mujica (2008). Los Costos No Laborales y el Mercado Laboral: Impacto de la Reforma de Salud en Colombia. *Working Paper 43. Fedesarrollo, Bogota, Colombia*.
- Schnepel, K. T. (2018). Good jobs and recidivism. *The Economic Journal* 128(608), 447–469.
- Schochet, P., J. Burghardt, and S. McConnell (2008). Does job corps work? impact findings from the national job corps study. *American Economic Review* 98(5), 1864–1886.
- Sviatschi, M. M. (2018). Making a Narco: Childhood Exposure to Illegal Labor Markets and Criminal Life Paths. *Working paper*.
- Sviatschi, M. M. (2020). Spreading Gangs: Exporting US Criminal Capital to El Salvador. *Mimeo*.
- Tobón, S. (2020). Do better prisons reduce recidivism? evidence from a prison construction program. *The Review of Economics and Statistics*, 1–47.
- Velasquez, A. (2020). The Economics Burden of Crime: Evidence from Mexico. *The Journal of Human Resources*. Forthcoming.
- Yang, D. (2008). Can Enforcement Backfire? Crime Displacement in the Context of Customs Reform in the Philippines. *Review of Economics and Statistics* 90(1), 1–14.

Online Appendix A: Additional Tables and Figures

Figure A1: Distribution of Age at Arrest (Males)



Source: Policía Nacional de Colombia. Vertical lines represent ages 13 and 26.

Figure A2: Timeline of Data Used

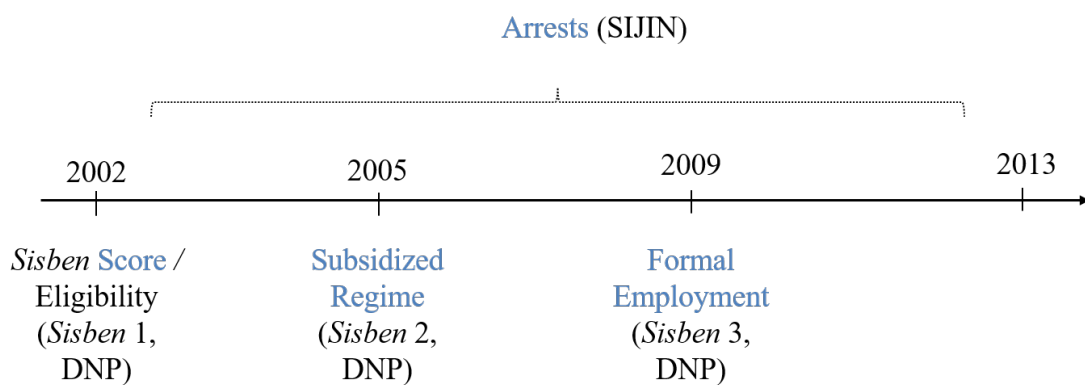
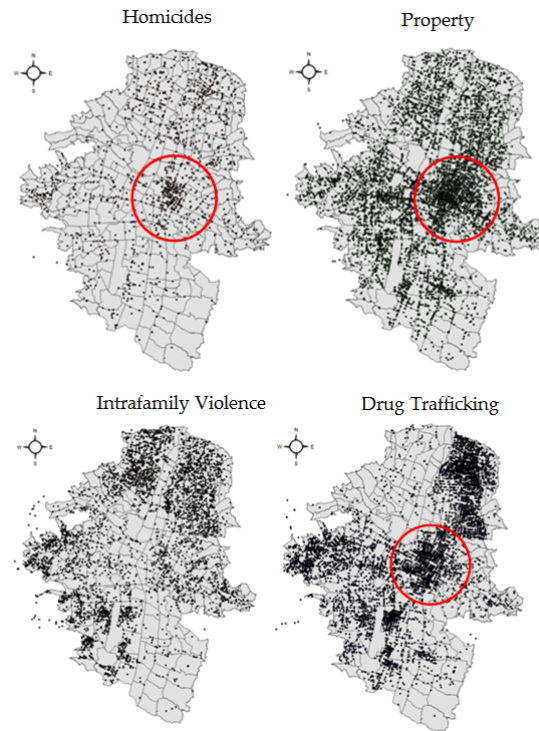
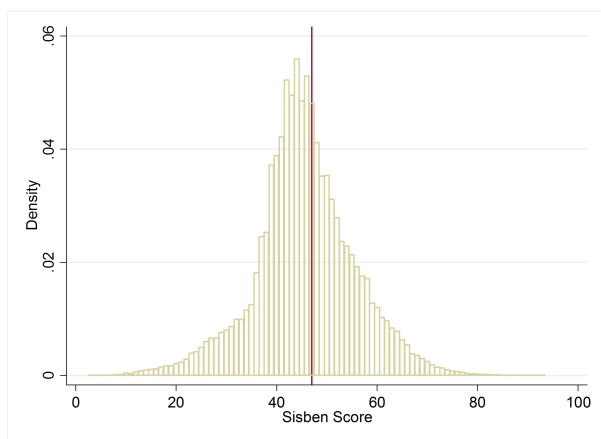


Figure A3: Location of 'in-the-act' arrests by type of crime, 2005-2013.

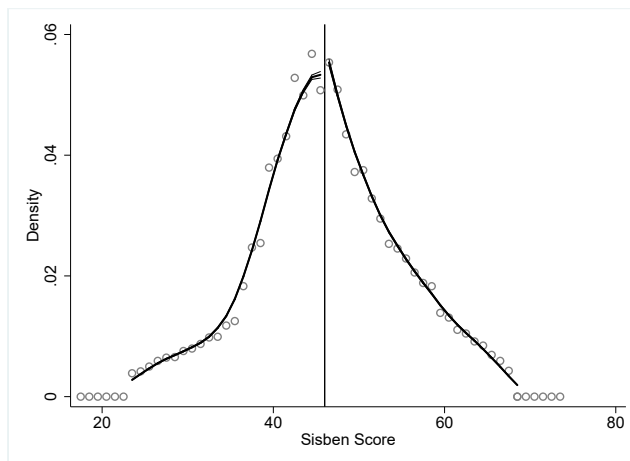


Source: [Medina and Tamayo \(2011\)](#) using Policía Nacional de Colombia. Dots indicate arrests. Bold lines are neighborhood boundaries. The red circle specifies the downtown of the city.

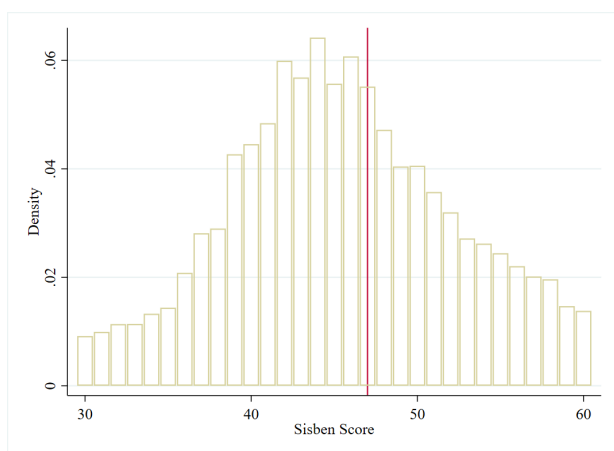
Figure A4: Sisben score Distribution



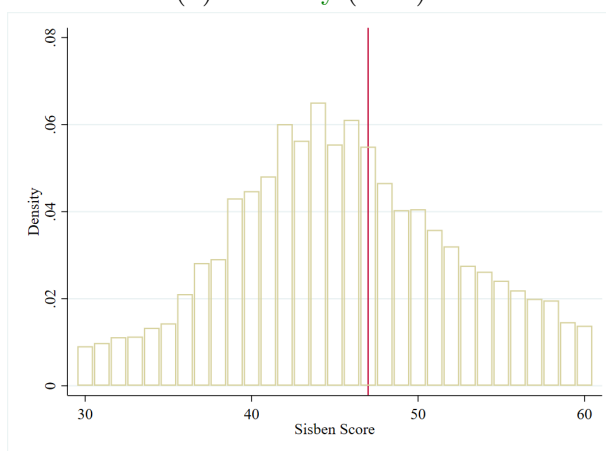
(a) Score distribution for males



(b) McCrary (2008) test



(c) Distribution for males (short bandwidth)



(d) Distribution for full sample (short bandwidth)

Source: *Sisben* survey of 2002. Figures A4a-A4c includes all males (i.e. both non-criminals and arrested individuals). Figure A4d is for the full sample (not just males). Figure A4b conducts a McCrary (2008) test.

Table A1: Top 25 Crimes (of 103) by Data-driven LACE Classifications

Crime	Gang Flags	Neighborhood Def	Incidence of crimes
Drug Crimes			46%
Non LACE			30%
LACE			16%
Drug Consumption / Possession	No	No	50%
Drug trafficking / Distribution - Marijuana	No	No	15%
Drug trafficking / Distribution	Yes	Yes	21%
Drug trafficking / Distribution - Cocaine paste	Yes	Yes	8%
Drug trafficking / Distribution Heroin	Yes	Yes	0%
Property Crime			28%
Non LACE			8%
LACE			20%
Use of Fake Identification, false document	No	No	9%
Motor vehicle theft (Motorcycles)	No	No	4%
Receiving illegal goods	No	No	4%
Receiving Bribes (as officials)	No	No	0.3%
Illegal public monopoly activity	No	No	2%
Copyright/Fraud	No	No	4%
Identity Theft	No	No	0.3%
Fraud	No	No	3%
Theft / Assault	Yes	Yes	31%
Robbery (To Businesses, firms)	Yes	Yes	17%
Property Vandalism	Yes	Yes	11%
Motor Vehicle Theft - Cars	Yes	Yes	5%
Burglary	Yes	Yes	3%
Violent Crime			23%
Non LACE			9%
LACE			14%
Simple Assault/Battery	No	No	2%
Rape/Sexual Assault	No	No	4%
Personal injuries	No	No	13%
Domestic/ Family Violence	No	No	17%
Conspiracy to commit murder	Yes	No	0.4%
Homicide	Yes	Yes	9%
Extortion	Yes	Yes	5%
Assault / Battery - Against Police	Yes	Yes	3%
Manufacture, Trafficking Firearms / Weapons	Yes	Yes	37%
Intimidation and Stalking	Yes	Yes	0.3%
Terrorism	Yes	Yes	0.4%
Kidnapping	Yes	Yes	2%

List of top crimes by type and enterprise classification, out of 103 crimes. LACE crimes are those “likely associated with criminal enterprises,” and as such most reflective of individual sorting into criminal occupations. Non-LACE crimes are the remaining, more likely representing crimes of impulse or opportunity. The ‘Gang Flags’ method lists whether the crime has a high propensity to receive a police reported flag of known gang affiliation at the time of arrest. The ‘Neighborhood Method’ classifies crimes that have a high propensity to be in neighborhoods known to have high gang activity.

Table A2: Formal Employment By Gender

	Bandwidths:	4	6	10
Panel A: Men Formal Employment in 2009				
Enrolled in SR		-0.0551*** (0.0169)	-0.0420*** (0.0105)	-0.0308*** (0.00827)
Number of observations		133,067	180,742	247,886
Sample mean				0.14
Panel B: Women Formal Employment in 2009				
Enrolled in SR		0.00560 (0.00757)	-0.0130* (0.00786)	-0.0169* (0.00889)
Number of observations		156,942	213,755	292,980
Sample mean				0.09

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. We use the *Sisben* survey of 2009 to construct formal employment. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the *Sisben* cutoff. Regressions control linearly for the 2002 *Sisben* score, flexibly around the cutoff. We cluster standard errors by *comuna*.

Table A3: Non-LACE Crimes: Heterogeneity by Comuna

	Bandwidths:	4	6	10
Panel A: Non-LACE Violent Crimes				
Enrolled in SR		0.00886 (0.0167)	0.00333 (0.0114)	-0.00421 (0.0123)
Enrolled* Gang Comuna		0.00773 (0.0240)	0.00236 (0.0150)	0.0126 (0.0144)
F stat		92.6	134.6	204.3
Number of observations		18,419	24,768	33,702
Panel B: Non-LACE Property Crimes				
Enrolled in SR		-0.00571 (0.0168)	-0.00526 (0.0120)	-0.00503 (0.00969)
Enrolled* Gang Comuna		-0.0176 (0.0180)	-0.0122 (0.0153)	-0.0100 (0.0122)
F stat		100.2	161.8	241.4
Number of observations		18,240	24,523	33,358
Panel C: Non-LACE Drug Crimes				
Enrolled in SR		-0.0379 (0.0299)	-0.0494 (0.0307)	-0.0292 (0.0222)
Enrolled* Gang Comuna		0.0132 (0.0292)	0.0155 (0.0211)	0.00281 (0.0195)
F stat		94.8	135.1	197.3
Number of observations		19,150	25,740	35,104

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Non-LACE, as determined by the data-driven classifications summarized in Table A1, are those more likely representing crimes of impulse or opportunity rather than activity of criminal enterprises. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff and an interaction between high-gang comunas and being below the cutoff. The Sisben score is measure in 2002, SR enrollment in 2005, and crime outcomes are measured between 2005 and 2013. Regressions include comuna fixed effects and an interaction between high-gang comunas and indicators for SR enrollment. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males who were between 13 to 18 years old in 2005. We cluster errors by comuna. The mean arrest rate across all five gang comunas are 18%, which is also the mean arrest rate in low-gang comunas.

Table A4: Match rate and Baseline (2002 *Sisben* Survey) Balance Tests

	Whole sample	Male Youth
First Principal Component	0.110 (0.0917)	-0.00246 (0.0446)
Years of Education	0.00248 (0.0606)	0.0275 (0.0552)
Age	0.0255 (0.0300)	0.00976 (0.0113)
Age Specific Education Gap	0.0159 (0.0602)	-0.0208 (0.0600)
HH Head Years of Education	-0.0832 (0.0592)	-0.0560 (0.0583)
Unemployed	0.0113 (0.00738)	0.0111 (0.00697)
Married	0.0275 (0.0175)	0.0254 (0.0175)
Employed	-0.0115 (0.0102)	-0.0139 (0.0106)
Attending School	0.000756 (0.00270)	0.000788 (0.00271)
Socioeconomic Stratum 2	-0.0129 (0.0144)	-0.00477 (0.0112)
Socioeconomic Stratum 1	0.0236* (0.0126)	0.00362 (0.00916)
Own House	0.0218 (0.0127)	0.00752 (0.0110)
Less than 6 years Olds	0.0131 (0.0129)	0.0122 (0.0126)
HH Head Age	0.000756 (0.00270)	0.338** (0.136)
Match Sisbens	0.00362* (0.00207)	0.00560 (0.00361)
Match rate	0.9241	0.9414
Observations	246,974	28,675

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report reduced form coefficients on being below the *Sisben* cutoff, where *Sisben* score is measured in 2002. Regressions are for the optimal bandwidth of 6 percentage points. Regressions control linearly for the *Sisben* score, flexibly around the cutoff. All variables are measured in 2002. We cluster standard errors by *comuna*. Neighborhood strata indicate the official socioeconomic strata of the neighborhood. First Principal Component takes the first principal component of all other variables.

Table A5: Neighborhood Classification Method

	Bandwidths:	4	6	10
Panel A: LACE Violent Crimes				
Enrolled in SR		0.0171** (0.00751)	0.0121* (0.00684)	0.00891** (0.00387)
Number of observations		17,995	24,198	32,931
Panel B: LACE Property Crimes				
Enrolled in SR		0.0335** (0.0131)	0.0271** (0.0122)	0.0192* (0.0107)
Number of observations		18,426	24,740	33,625
Panel C: LACE Drug Crimes				
Enrolled in SR		0.0284* (0.0163)	0.0108 (0.0126)	-0.00197 (0.0115)
Number of observations		18,909	25,447	34,661

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measured in 2002 and SR enrollment in 2005. Crime data are from 2005 to 2013. Results use the neighborhood classification method described in the text to classify crimes. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males who were between 13 to 18 years old in 2005.

Table A6: LACE-predicted Regressions

Bandwidths	4	6	10
Panel A: Property Crimes			
Enrolled in SR	0.00804** (0.00391)	0.00454** (0.00189)	0.00311** (0.00135)
Number of observations	18,714	25,138	34,168
Sample mean			0.0450
MSE			0.0753
Pseudo R2			0.0809
Panel B: Violent Crimes			
Enrolled in SR	0.0234** (0.0118)	0.0208** (0.00841)	0.0101* (0.00557)
Number of observations	18,572	24,977	33,983
Sample mean			0.0391
MSE			0.1069
Pseudo R2			0.1187
Panel C: Drug Crimes			
Enrolled in SR	-0.0244 (0.0531)	-0.0583 (0.0459)	-0.0433 (0.0397)
Number of observations	19,679	26,476	36,120
Sample mean			0.0943
MSE			0.0497
Pseudo R2			0.2176

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measured in 2002, SR enrollment in 2005, and crime outcomes are measured between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males who were between 13 to 18 years old in 2005. We cluster errors by comuna. We weight observations with the predicted likelihood that the crime committed was a LACE crime based on the police flags and the characteristics of each crime. We use a k-mean logit to make the prediction. For non criminals we impute a weight equal to 1.

Table A7: Semi-parametric RD with Bias Correction

	Type of Crime	Violent	Property	Drug
Enrolled in SR		0.0164	0.02794	0.00768
Standard error		(0.00972)	(0.01636)	(0.02069)
Bias corrected p-value		0.077	0.052	0.57
Bandwidth		5.2	5.8	6.6
Bias correction bandwidth		9.9	8.9	9.6
Number of observations		24,206	26,511	29,102

Note: Results using the [Calonico et al. \(2014a\)](#) CCT method for estimation, where the primary estimation uses a linear functional form and the bias correction uses a quadratic form. Arrests are restricted to LACE crimes, which are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in [Table A1](#), and as such most reflective of individual sorting into criminal occupations. Tables report fuzzy RD two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff, where Sisben score is measured in 2002. Crime data is from 2005 to 2013. We consider only males who were between 13 to 18 years old in 2005.

Table A8: Standard Errors Clustered at Household Level

Bandwidths	4	6	10
Panel A: LACE Property Crimes			
Above Cutoff Reduced Form	0.0106* (0.00573)	0.00930** (0.00459)	0.00666* (0.00361)
Enrolled in SR No Covariates	0.0380* (0.0206)	0.0331** (0.0164)	0.0232* (0.0126)
Enrolled in SR Including pre-treatment covariates	0.0408* (0.0224)	0.0341* (0.0175)	0.0240* (0.0130)
Number of observations	18,426	24,740	33,625
Sample mean			0.0259
Panel A: LACE Violent Crimes			
Above Cutoff Reduced Form	0.00722** (0.00351)	0.00649** (0.00287)	0.00456** (0.00229)
Enrolled in SR No Covariates	0.0257** (0.0126)	0.0231** (0.0102)	0.0158** (0.00793)
Enrolled in SR Including pre-treatment covariates	0.0274** (0.0139)	0.0232** (0.0110)	0.0149* (0.00822)
Number of observations	18,052	24,272	33,027
Sample mean			0.0114
Panel A: LACE Drug Crimes			
Above Cutoff Reduced Form	0.00799 (0.00585)	0.00348 (0.00479)	0.00133 (0.00378)
Enrolled in SR No Covariates	0.0285 (0.0209)	0.0124 (0.0171)	0.00461 (0.0132)
Enrolled in SR Including pre-treatment covariates	0.0303 (0.0232)	0.0135 (0.0184)	0.00524 (0.0137)
Number of observations	18,463	24,857	33,851
Sample mean			0.0314

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report reduced form and two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measured in 2002, SR enrollment in 2005, and crime outcomes are measured between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males who were between 13 to 18 years old in 2005. We cluster errors by household.

Table A9: Sample Aggregated to the Household Level

Bandwidths	4	6	10
Panel A: LACE Property Crimes			
Enrolled in SR	0.0454*** (0.0151)	0.0428*** (0.0135)	0.0388*** (0.0122)
Number of observations	15,168	20,392	27,815
Sample mean			0.0334
Panel B: LACE Violent Crimes			
Enrolled in SR	0.0134* (0.00700)	0.0172** (0.00790)	0.00703 (0.00460)
Number of observations	14,408	19,398	26,472
Sample mean			0.0160
Panel C: LACE Drug Crimes			
Enrolled in SR	0.0328* (0.0185)	0.0151 (0.0152)	0.0116 (0.0136)
Number of observations	14,794	19,937	27,214
Sample mean			0.0300

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measure in 2002, SR enrollment in 2005, and crime outcomes are measured between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only households with youths between 13 and 18 in 2005. We cluster errors by comuna.

Table A10: Robustness Check: Including Other Crimes in the Sample

Bandwidths	4	6	10
Panel A: LACE Property Crimes			
Above Cutoff Reduced Form	0.00899** (0.00322)	0.00809** (0.00317)	0.00577* (0.00282)
Enrolled in SR No Covariates	0.0320*** (0.0100)	0.0289*** (0.00998)	0.0204** (0.00894)
Enrolled in SR Including pre-treatment covariates	0.0333*** (0.0112)	0.0280*** (0.0107)	0.0203** (0.00876)
Number of observations	21,720	29,235	39,877
Sample mean			0.0259
Panel B: LACE Violent Crimes			
Above Cutoff Reduced Form	0.00598*** (0.00194)	0.00547** (0.00201)	0.00385*** (0.00128)
Enrolled in SR No Covariates	0.0213*** (0.00718)	0.0195*** (0.00676)	0.0136*** (0.00416)
Enrolled in SR Including pre-treatment covariates	0.0226*** (0.00761)	0.0203*** (0.00751)	0.0134*** (0.00430)
Number of observations	21,720	29,235	39,877
Sample mean			0.0114
Panel C: LACE Drug Crimes			
Above Cutoff Reduced Form	0.00667 (0.00593)	0.00298 (0.00403)	0.00115 (0.00377)
Enrolled in SR No Covariates	0.0238 (0.0195)	0.0107 (0.0138)	0.00279 (0.0132)
Enrolled in SR Including pre-treatment covariates	0.0238 (0.0220)	0.0114 (0.0148)	0.00507 (0.0132)
Number of observations	21,720	29,235	39,877
Sample mean			0.0314

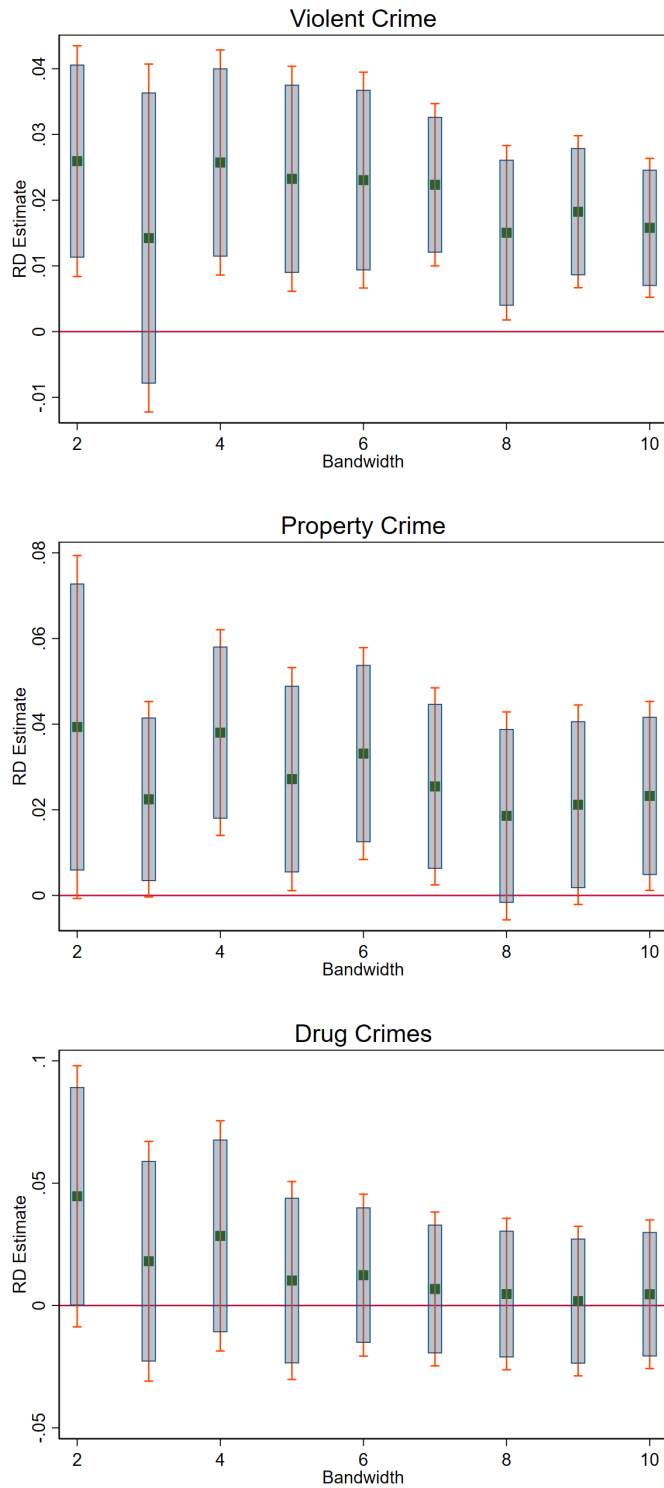
Note: The sample includes other crimes. For instance, when looking at violent enterprise-crime arrests as the outcome of interest, property crime, drug crime and violent non-enterprise crime arrests are also in the sample grouped with the people never arrested in this period. LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations. Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males who were between 13 to 18 years old in 2005.

Table A11: Robustness Check: Including Repeat Arrests

Bandwidths:	4	6	10
Panel A: LACE Property Crimes			
Enrolled in SR	0.0449*** (0.0123)	0.0414*** (0.0145)	0.0284** (0.0128)
Number of Observations	18,488	24,822	33,739
Panel B: non-LACE Property Crimes			
Enrolled in SR	-0.0158 (0.0217)	-0.0136 (0.0172)	-0.0118 (0.0115)
Number of Observations	18,265	24,558	33,409
Panel C: LACE Violent Crimes			
Enrolled in SR	0.0175** (0.0088)	0.0154* (0.0088)	0.0101* (0.0052)
Number of Observations	18,071	24,294	33,060
Panel D: non-LACE Violent Crimes			
Enrolled in SR	0.0145 (0.0176)	0.0066 (0.0123)	0.0016 (0.0144)
Number of Observations	18,480	24,842	33,800
Panel E: LACE Drug Crimes			
Enrolled in SR	0.0369 (0.0270)	0.0090 (0.0191)	-0.0028 (0.0177)
Number of Observations	18,524	24,944	33,973
Panel F: non-LACE Drug Crimes			
Enrolled in SR	-0.0336 (0.0310)	-0.0585 (0.0322)	-0.0428 (0.0245)
Number of Observations	19,327	25,982	35,450

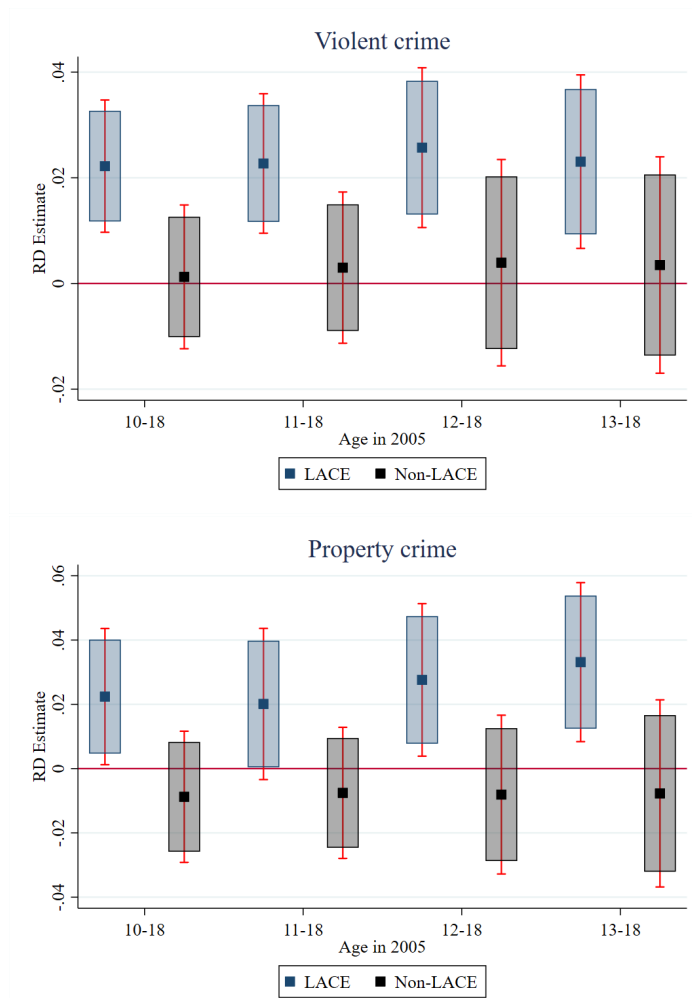
Note: The sample includes repeat arrests. LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations. Non-LACE crimes are the remaining, more likely representing crimes of impulse or opportunity. Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report reduced form coefficients. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males who were between 13 to 18 years old in 2005. We do not include controls.

Figure A5: Robustness to Bandwidths (LACE Crime)



Note: Coefficients of RD 2SLS regressions where the first stage is SR Enrollment on being below the Sisben cutoff. Sample of LACE crimes only. LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations. Grey bars indicate 90% confidence intervals. Red lines indicate 95% confidence intervals.

Figure A6: Robustness to Different Cohorts (LACE Crime)



Note: Coefficients of RD 2SLS regressions where the first stage is SR Enrollment on being below the Sisben cutoff. Sample of LACE crimes only. LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations. Grey bars indicate 90% confidence intervals. Red lines indicate 95% confidence intervals. We restrict the sample to individuals who were a specific age in 2005 (say, ages 10 to 18), and follow them till 2013 (when they are between 18 to 26 years old).

Table A12: Simultaneously both Informal (in 2009) and Arrested

	Bandwidths:	4	6	10
Panel A: LACE Violent Crimes				
Enrolled in SR		0.0104 (0.00771)	0.00997** (0.00405)	0.00990** (0.00436)
Number of observations		12,015	16,023	21,733
Panel B: LACE Property Crimes				
Enrolled in SR		0.0393** (0.0184)	0.0331*** (0.0122)	0.0183 (0.0128)
Number of observations		12,244	16,319	22,074
Panel C: LACE Drug Crimes				
Enrolled in SR		0.0207 (0.0177)	0.0299*** (0.00983)	0.0291** (0.0126)
Number of observations		12,253	16,368	22,199

Note: The outcome is arrests only for those not formally employed as measured in 2009. We exclude all arrests post 2009. Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. LACE crimes are those “likely associated with criminal enterprises,” as determined by the data-driven classifications summarized in Table A1, and as such most reflective of individual sorting into criminal occupations. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measured in 2002, and SR enrollment in 2005. Crime is measured between 2005 and 2009. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males who were between 13 to 18 years old in 2005.

Table A13: Robustness to Including Additional Control Variables (Property Crimes)

Bandwidths	4	6	10
Panel A: LACE Property Crimes			
Enrolled in SR	0.0408***	0.0341***	0.0240**
Including pre-treatment covariates	(0.0139)	(0.0131)	(0.0108)
Enrolled in SR	0.0415***	0.0351***	0.0261**
Including pre-treatment covariates and community characteristics	(0.0137)	(0.0132)	(0.0115)
Number of observations	18,426	24,740	33,625
Sample mean			0.0259
Panel B: Non-LACE Property Crimes			
Enrolled in SR	-0.0116	-0.00872	-0.00854
Including pre-treatment covariates	(0.0212)	(0.0156)	(0.0119)
Enrolled in SR	-0.0112	-0.00765	-0.00771
Including pre-treatment covariates and community characteristics	(0.0209)	(0.0155)	(0.0107)
Number of observations	18,240	24,523	33,358
Sample mean			0.0190

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measure in 2002, SR enrollment in 2005, and crime outcomes are measured between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. For regressions that have pre-treatment covariates, we include household characteristics, year of birth fixed effects, and neighborhood fixed effects. For regressions that have community characteristics we include share formally employed and average income for formal workers interacted with being above the Sisben cutoff and Sisben score. We consider only males who were between 13 to 18 years old in 2005. We cluster errors by comuna.

Table A14: Robustness to Including Additional Control Variables (Violent Crimes)

Bandwidths	4	6	10
Panel A: LACE Violent Crimes			
Enrolled in SR	0.0274***	0.0232**	0.0149**
Including pre-treatment covariates	(0.00950)	(0.00937)	(0.00583)
Enrolled in SR	0.0280***	0.0232**	0.0154**
Including pre-treatment covariates and community characteristics	(0.00997)	(0.00946)	(0.00601)
Number of observations	18,052	24,272	33,027
Sample mean			0.0114
Panel B: Non-LACE Violent Crimes			
Enrolled in SR	0.00791	0.00118	-0.00322
Including pre-treatment covariates	(0.0168)	(0.0111)	(0.0125)
Enrolled in SR	0.00815	0.00242	-0.00236
Including pre-treatment covariates and community characteristics	(0.0167)	(0.0115)	(0.0125)
Number of observations	18,419	24,768	33,702
Sample mean			0.0277

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measure in 2002, SR enrollment in 2005, and crime outcomes are measured between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. For regressions that have pre-treatment covariates, we include household characteristics, year of birth fixed effects, and neighborhood fixed effects. For regressions that have community characteristics we include share formally employed and average income for formal workers interacted with being above the Sisben cutoff and Sisben score. We consider only males who were between 13 to 18 years old in 2005. We cluster errors by comuna.

Table A15: Robustness to Including Additional Control Variables (Drug Crimes)

Bandwidths	4	6	10
Panel A: LACE Drug Crimes			
Enrolled in SR	0.0303	0.0135	0.00524
Including pre-treatment covariates	(0.0270)	(0.0180)	(0.0159)
Enrolled in SR	0.0299	0.00954	0.00380
Including pre-treatment covariates and community characteristics	(0.0272)	(0.0186)	(0.0161)
Number of observations	18,463	24,857	33,851
Sample mean			0.0314
Panel B: Non-LACE Drug Crimes			
Enrolled in SR	-0.0385	-0.0501	-0.0277
Including pre-treatment covariates	(0.0299)	(0.0329)	(0.0230)
Enrolled in SR	-0.0370	-0.0496	-0.0270
Including pre-treatment covariates and community characteristics	(0.0298)	(0.0320)	(0.0236)
Number of observations	19,150	25,740	35,104
Sample mean			0.0629

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measure in 2002, SR enrollment in 2005, and crime outcomes are measured between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. For regressions that have pre-treatment covariates, we include household characteristics, year of birth fixed effects, and neighborhood fixed effects. For regressions that have community characteristics we include share formally employed and average income for formal workers interacted with being above the Sisben cutoff and Sisben score. We consider only males who were between 13 to 18 years old in 2005. We cluster errors by comuna.

Table A16: Effect on Enrollment in Education

Bandwidths	4	6	10
Enrollment in 2006 or 2007			
Above Cutoff	-0.0232* (0.0111)	-0.0143 (0.00909)	-0.0195** (0.00865)
Number of observations	22,125	29,778	40,642
Sample mean			0.436
Enrollment in 2007			
Above Cutoff	-0.0299** (0.0112)	-0.0248** (0.0103)	-0.0226** (0.00887)
Number of observations	22,125	29,778	40,642
Sample mean			0.436
Enrollment in 2006			
Above Cutoff	-0.0218* (0.0125)	-0.0173 (0.0113)	-0.0208* (0.0102)
Number of observations	22,125	29,778	40,642
Sample mean			0.402

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report coefficient of indication of being above the Sisben cutoff. The Sisben score is measured in 2002, SR enrollment in 2005, and crime outcomes are measured between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males who were between 13 to 18 years old in 2005. We cluster errors by comuna.

Table A17: Robustness to Using Only *In Flagrante* Crimes

Bandwidths	4	6	10
Panel A: LACE Property Crimes			
Enrolled in SR	0.0374*** (0.0128)	0.0342*** (0.0125)	0.0242** (0.0119)
Number of observations	18,406	24,714	33,593
Sample mean			0.0254
Panel B: LACE Violent Crimes			
Enrolled in SR	0.0143 (0.00934)	0.0182* (0.00981)	0.0110** (0.00555)
Number of observations	17,993	24,190	32,912
Sample mean			0.00872
Panel C: LACE Drug Crimes			
Enrolled in SR	0.0302 (0.0234)	0.0149 (0.0163)	0.00539 (0.0154)
Number of observations	18,458	24,848	33,840
Sample mean			0.0312

Note: Standard errors in parentheses. *** significant at 1%; ** significant at 5%; * significant at 10%. Tables report two-staged least squares (2SLS) coefficients where the first stage is SR enrollment on being below the Sisben cutoff. The Sisben score is measured in 2002, SR enrollment in 2005, and crime outcomes are measured between 2005 and 2013. Regressions control linearly for the Sisben score, flexibly around the cutoff. We consider only males who were between 13 to 18 years old in 2005 and in flagrante crimes. We cluster errors by comuna.

Online Appendix B: Proofs

B1 Deriving $\pi_{\sigma do}$

To derive the equation for $\pi_{\sigma do}$, we need the distribution of the indirect utility U . This distribution determines the probability that a (d, o) -type worker chooses working in sector σ . First, notice that the distribution of the utility of a worker choosing σ is also Frechet:

$$G_{\sigma do}(u) \equiv \mathbb{P}(U \leq u) = F\left(\frac{r_{K,d}u}{b_{\sigma do}\bar{K}_d w_{\sigma do}}\right) = e^{-\Phi_{\sigma do}u^{-\gamma}}, \quad (12)$$

where $\Phi_{\sigma do} = E_{\sigma}\left(\frac{\bar{K}_d}{r_{K,d}}\right)^{\gamma} (b_{\sigma do}w_{\sigma do})^{\gamma}$. As such, the distribution across all possible σ 's follows a Frechet distribution:

$$G_{do}(u) \equiv \mathbb{P}\left(\max_{\sigma} U \leq u\right) = e^{-\Phi_{do}u^{-\gamma}} \text{ where } \Phi_{do} = \sum_{\sigma} \Phi_{\sigma do}.$$

Therefore, the probability that a (d, o) -type worker chooses working in sector σ is¹

$$\pi_{\sigma do} = \mathbb{P}\left(U \geq \max_{\sigma'} U\right) = \frac{E_{\sigma}(b_{\sigma do}w_{\sigma do})^{\gamma}}{\sum_{\sigma' \in \{f, n, c\}} E_{\sigma'}(b_{\sigma' do}w_{\sigma' do})^{\gamma}}. \quad (13)$$

Online Appendix C: Details on the Context and Policies

In this appendix we summarize the context and policies, with a particular focus on the rules, eligibility requirements, costs of access, services covered under each regime, the corresponding policy variation, and other benefits policies.

C1 Rules

According to Law 100 of 1993, enrollment in the General System of Social Security in Health (GSSSH) is mandatory: all employers must enroll their employees to the CR, and the government must provide such enrollment in the SR to anyone with no employer or that lacked the ability to contribute to the CR. Nonetheless, the evidence shows that there are informal workers not enrolled in the GSSSH. Formal workers (regardless of income) must be enrolled in the CR. Similarly, independent workers earning more than a monthly minimum wage must enroll in CR. Moreover, any informal worker belonging to a household whose Sisben score is below the cutoff and who cannot become a beneficiary of a formal worker in her household who contributed to the CR, must enroll in the SR. Yet, the mandatory SR enrollment is not always enforced.

If one leaves the SR and joins the CR, while de facto, getting back into SR is not automatic everywhere, for our analyzed period in Medellin it was extremely likely for someone to re-enroll in SR if they wanted and if, most importantly, they were still eligible according to their most recently calculated score. However, if one leaves SR to CR because they become formal, it will likely raise their Sisben score and might put them above the cutoff. If they then quit the formal sector and want access to SR again, it may be a while before they can access SR because they are stuck with the higher Sisben score until the next update. Thus, even though nobody is

¹This last step follows from [Ahlfeldt et al. \(2015\)](#).

aware of the exact Sisben formula, the household head and members may be afraid of losing their eligibility to the SR if a member of the household becomes a formal employee, as their Sisben score may increase above the eligibility cutoff.

C1.1 Eligibility

The Sisben survey is carried out on all the people who live in a house, whether or not they are related. The Sisben groups the people surveyed into family groups, which can be made up of one or more people who have a close kinship relationship such as parents, children, spouses, etc. The Sisben classifies family groups according to kinship relationships and not the score. For example, suppose a new family unit was created in a home where a person lives due to the arrival of a new member. In that case, people must go to the Sisben offices in their municipality to include the new member and ask to survey this new member.

As we mentioned above, the Sisben score was calculated using a confidential formula based on respondent socioeconomic characteristics, disability, and housing characteristics. The housing characteristics are the same for all the family groups in a house. Thus, individuals cannot offset changes in household eligibility by moving, registering as a separate household, changing their dependent status.

Finally, note that in a household that has all its members enrolled in the SR, if one of the members becomes formally employed, that person must begin to contribute to the CR and enroll as beneficiaries of the CR all dependents. Dependents include the member's spouse, any children below 18, any children 18-25 who are studying, and under certain conditions, parents. For example, suppose an individual who is 19 years old becomes formally employed, and they do not have a partner/spouse to enroll in the CR. In that case, they must enroll their parents as beneficiaries of the CR. Of course, any other household member not eligible to be a beneficiary (e.g., their 15-year-old brother/sister) who was already enrolled in the SR, could continue to be enrolled in the SR. However, their Sisben score is likely to be affected by the enrollment of the 19 years old household member (as it is explained below).

C2 Costs

Workers enrolled in the CR need to pay two types of health insurance expenses: enrollment contributions and copayments for the health services demanded. Workers in Colombia need to pay between 4% and 12.5% of their salaries to the CR. They are officially charged 4% of their wages and employers are charged 8.5%, but likely pass on this cost to workers in competitive labor markets. These payment amounts are regardless of whether they are below or above the Sisben cutoff (i.e., workers' payment is independent of their Sisben score and depends only on their salary). For the health services demanded, individuals enrolled in the CR pay a copayment and a Cuota Moderadora (fixed out-of-pocket amount). The copayments across health services are a portion of the costs covered by individuals out of pocket that vary with the workers' salary (number of minimum wages). The Cuota Moderadora is a fixed out-of-pocket amount paid by an insured for covered services and varies with the workers' salary (number of minimum monthly wages). Both the copayments and Cuotas Moderadoras are independent of the Sisben score. Thus, the only alternative for someone above the cutoff is to enroll in the CR and pay 4-12.5% of their salary as well as the copayments and Cuotas Moderadoras for the health services demanded. Thus, the total out-of-pocket could go from 58% to around 460% of

the legal monthly minimum wage (depending on the worker's salary) on a calendar year basis.² For example, in a household of four members, it would imply the payment of at least 4% of the salary of all the formal workers, plus the payment of the copayments and Cuotas Moderadoras of all health services demanded by all of the four household members.

Informal or unemployed individuals enrolled in the SR do not pay for their enrollment. Similarly, there are no Cuotas Moderadoras, and the copayments could go from 0 to 100% of a legal minimum wage on a calendar year basis. Thus, in a household of four members, it would imply no enrollment fees or Cuota Moderadora for all the household members, plus the copayment of the health services demanded by all of the four household members. In this example, the differential amount explained merely by the 4% of the wages of two members would be $4\% \times 2 \times \text{USD}171.24$ (average wage at the cutoff from Table 3) $\times \$2,150.3/\text{USD}$ (average exchange rate in 2009) = \$29,457, that is, about 6% of a monthly minimum wage (monthly minimum wage in 2009: \$496,900). Finally, individuals enrolled in the SR do not have to pay for prenatal checkups, deliveries (and related complications), and child health care during their first year of life.

C3 Services covered

Initially (in 2002), those with SR could access any hospital (with insurance agreements) for 55% of the health services covered by CR. Therefore, the only option to demand the services not covered by the SR would be to be covered by the network of public hospitals, which might imply some additional non-monetary costs (longer waiting times and a narrower basket of hospitals and physicians, among others). Over time, different groups in SR could start accessing any hospital for the full basket of services: first for children under 12 in 2009 (CRES Accord 4), then for children under 18 in 2010 (CRES Accord 11), for adults 60 and older in 2011 (CRES Accord 27), and finally, for adults 18 to 59 in 2012 (CRES Accord 32).

C4 Policy Variation

The policy studied in the paper increases the benefits for not being formally employed discontinuously at the cutoff, by reducing the cost of access to health benefits. Note that the CR cost and coverage does not change at the threshold. At the threshold, there is a new option to access SR, which, as we describe above is free enrollment and much lower copayments. As we mentioned before, the cost of access changes substantially between the two regimes. Those in the CR are mandatorily taxed, and so need to pay for their enrollment contributions as well as copayments and Cuotas Moderadoras for the health services demanded. Individuals enrolled in the SR do not pay for their enrollment or Cuotas Moderadoras, and the copayments are substantially smaller.

C5 Other Policies

While the Sisben score is used for other programs, the cutoff for this policy does not determine eligibility for any other large or meaningful policy over the time period of our analysis. The CONPES social policy documents numbers 55 of 2001 and 100 of 2006 describe the programs. The other large policies: the PACES program (CONPES 55), Rural Youths Program (CONPES

²See for example the data provided by the Ministry of Health (<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VP/DOA/cuotas-moderadoras-copagos-2021.pdf>) and providers of the CR, or EPS (<https://coosalud.com/cuotas-moderadoras-y-copagos-de-2021/>).

100) and *Familias en Acción* (CONPES 100) are targeted to other populations. For instance, PACES is targeted spatially, the Rural Youths Program is irrelevant for urban cities, and the *Familias en Acción* uses the Sisben score, but has a different cutoff. There is a Housing program that depends on a broad formula (which includes your Sisben levels) among other inputs. The Housing Program is small (less than 6,000 people in the lower Sisben levels, as opposed to 467,000 people in SR).

There are two other smaller programs that vary around the SR cutoff. The *Programa de Protección Social al Adulto Mayor* (PPSAM), sends unconditional transfers to the elderly population. But the coverage (since targeted to the old) is only 1% of the SR beneficiaries, and the budget is 2.2% of the SR budget. Most of the budget is spent on the poorest Sisben level 1 individuals (a different cutoff), or on those living alone, or with physical and mental disabilities. The final program is the *Jóvenes en Acción* (JeA), targeted about 20% of 18-25 year-olds in Sisben levels 1 and 2. The program was piloted in 2002, and continued till 2005. Except, the 18 year-olds in 2005, no other age groups in the various years overlap with our cohorts of study. Because the 18 year-olds in 2005 only had one shot at enrolling in JeA, this constitutes only 0.1% of our youth SR target sample.

D Online Appendix: Details on the Crime Data

We use the raw arrest data from the Judicial Police Sectional of the National Police Department. The data are different from the data on prosecutions, convictions and penalties, and are unrelated to court hearings. These are arrests, regardless of whether or not there is an eventual court hearing. Moreover, about 91% of the arrests occur *in flagrante* (in the act). At the time of the arrest, police officers are unaware of the individual's age, and only when they are booked at the station does their age become revealed. Finally, police are unlikely to know a person's Sisben score or formal status.

D1 Criminal justice system

The minimum legal age for Colombians to be subject of the criminal justice system is 14. Colombians above 18 years of age are subject to the Colombian Penal Code, youths between 14 and 18 years old are subject to the Penal Responsibility System for Adolescents. If a youth between 14 to 18 years old commits a violent crime subject to the Colombian Penal Code, the system could condemn the youth to detention (depending on the crime). However, the detention would not be in a regular prison for adults but in a specialized detention center under the Colombian Family Welfare Institute (ICBF by its acronym in Spanish). We may observe arrests for 13 year olds as their ages may not be known to arresting officers before they are entered into the database.

D2 Classifications of crimes

We use the exact crime definition that the individual was arrested for, and then classify those crimes using the US BJS data (rather than the Colombian data). Even though the Colombian system may not split thefts into various categories, since we can see a description of each act, we can classify the crimes using the refined BJS system.

In addition to the variable that classifies crimes using the Colombian penal code, we have a separate variable that provides a detailed description of the crimes individuals were involved in (data comes from the police). Thus, we can identify in our data set if individuals were

arrested because they committed a drug-related crime (i.e., a crime related to Article 376 of the Colombian Penal Code), but we can also identify whether that crime was drug consumption or drug trafficking and if it was for trafficking cocaine, heroin, or marijuana.

We use this to classify the crimes into LACE and non-LACE. In our data, individuals are booked under various drug crimes: distinguishing trafficking, consumption and production. Yet, classifying drug crimes is extremely challenging, as when caught with drugs, it is difficult for the police to know whether the individual was selling, trafficking or consuming.

Finally, according to Article 376 of the Penal Code, drug-related crimes are bringing into or out of the country, carrying, storing, producing, selling, offering, acquiring, funding or providing, any positive amount of narcotic, psychotropic, or synthetic drug. For our sample period (between 2002 and 2013), carrying or consuming the minimum dose was not penalized. The minimum dose is defined as 20 grams of marijuana and/or 1 gram of cocaine or coca base.

D3 Match between Crime Data and Sisben

We match 78% of arrests, as many arrested may not be residents of the primary metropolitan region of Medellin. The matched arrests are neither systematically more likely to be gang-related crimes nor in flagrante crimes. Those who are arrested are more likely to be enumerated as poor. While all that matters for our empirical strategy is the match rate at the cutoff, we can examine the probability of showing up in the arrests data across the Sisben score distribution. For instance, for someone around the cutoff, there is a 7% chance they will be matched to the crime data; whereas for someone with a low poverty score (i.e., relatively richer) there is about 1% chance of being arrested. This suggests that the unmatched arrests are unlikely to be not in the census of the poor, but rather from neighboring regions outside Medellin.

D4 Classifications of Neighborhoods

One of the main advantages of our data is that it allows us to observe both the origin of (where criminals live) and destination of (where the crime was committed) the crime. When we classify neighborhoods based on the incidence of crimes, we classify individuals based on where they live (residence), which is the most relevant for what activities they get recruited into. It also allows us to include controls and neighborhood fixed effects for the residence of the individuals that were captured.

From the model presented in Section 3, various factors vary across neighborhoods – job opportunities, peers in each occupation, policing, and the productivity levels of each type of job. As such, ones neighborhood may be important. Indeed, work that simply relies on area-based variation simultaneously picks up changes to relative returns with these other factors.

In [Khanna et al. \(2022\)](#), we combine transit surveys, along with GIS procedures to calculate the travel time between each person’s residence, and location of crime or formal employment. We show that most crimes are committed in near distance from ones residence – with the exception of property crime. The activity for which individuals travel the farthest is for legitimate employment. As such, most gang members live, and engage in, crime within their neighborhoods.