Who expects to join criminal gangs and why? Occupational choice among 5,000 teenage boys in Medellín*

Christopher Blattman Arantxa Rodriguez-Uribe Santiago Tobon†

[Ongoing work, do not cite or circulate]

Abstract

Across the Americas, criminal gangs are among the largest forced recruiters of children and adolescents into armed groups. What techniques do they use? Which adolescents are most at risk? And what NGO and government interventions can prevent and disrupt this forced recruitment? We are currently running a survey targeting 5,000 13-year-old adolescent males in Medellin's highest-risk gang recruitment neighborhoods. We will use the survey to assess risk factors associated with recruitment. To mitigate the identification problem concerning the separation of preferences, expectations, and structural barriers, we use rich data on subjective expectations, with direct measures of financial constraints, to estimate a life-cycle model of preferred career path. In this preliminary paper, we describe the model, report preliminary descriptive statistics, and discuss intervention design. By May, we expect to present descriptive statistics on the full sample and report results of survey experiments that will inform our field experimental interventions.

*For comments and feedback we thank Thomas Abt, Leo Bursztyn, Scott Decker, Manasi Deshpande, Rebecca Dizon-Ross, Greg Lane, Raul Sánchez de la Sierra, Maria Micaela Sviatschi, and Chris Roth. Innovations for Poverty Action coordinated all research activities. For research assistance we thank Verónica Abril, Bruno Aravena-Maguida, Felipe Fajardo, Ana María Leal, Felipe Martínez, Juan Pablo Mesa-Mejía, Angie Mondragón, Helena Montoya, Juan Sebastián Ortiz, and México Vergara. We thank the Secretariat of Education of Medellín for their cooperation. For financial support, we thank the Templeton World Charity Foundation, the Ministry of Science and Technology of Colombia, Towards Kindness, and Innovations for Poverty Action Peace and Recovery Program.

†Blattman: University of Chicago (Harris School of Public Policy), blattman@uchicago.edu; Rodriguez-Uribe: Princeton University (Department of Political Science), mauribe@princeton.edu; Tobón: Universidad EAFIT (School of Finance, Economics and Government), stobonz@eafit.edu.co

1 Introduction

Across the Americas, organized crime and gangs are the main users of force, fraud, and deception to recruit children and adolescents into armed groups. Consider Medellin—a prosperous city of roughly 4 million people in the metropolitan area, and Colombia's commercial and industrial heartland. There, we estimate there are 12,000 current gang members plus thousands more victims in jail, killed, or retired. Most of these young men were recruited between the ages of 13 and 17 (Blattman et al., 2021).

When it comes to recruiting children into non-state armed groups, most governments have focused on insurgent groups (Department of State, 2022). As in most Latin American countries, however, criminal groups are far larger recruiters of young men and women. For example, our data suggest that in the low- and middle-income neighborhoods of Medellin, about 10% of all males aged 15–34 are affiliated with one of the roughly 400 local armed gangs known (Blattman et al., 2021). This situation seems to be similar in other contexts such as El Salvador or Rio de Janeiro (Melnikov et al., 2020; Carvalho and Soares, 2016).

With criminal gangs being among the largest forced recruiters of children and adolescents into armed groups, this work aims to shed light on several questions: What techniques do they use? Which adolescents are most at risk? And what NGO- and government-based interventions can prevent and disrupt this forced recruitment?

We are currently running a panel survey targeting at least 5,000 13-year-old adolescent males in Medellin's highest-risk gang recruitment neighborhoods. We will use the survey to assess risk factors associated with recruitment and, through future rounds of survey and administrative data, to track gang and criminal justice involvement among these youth for years to come. We will also experimentally evaluate two behavioral interventions in this sample. Each is designed to counter common forms of deception coercing youth into gangs: countering misperceptions about the returns to non-criminal careers and countering misperceptions about the risks of death and incarceration from joining a gang. In this preliminary paper, we develop a life-cycle model of preferred career path for adolescents, report initial descriptive statistics on roughly 15% of our target sample, and discuss intervention design. By May, we expect to present descriptive statistics on the full sample and report results of survey experiments that will inform our field experimental interventions.

Medellín's gangs are highly organized criminal structures. Most have 15 to 45 salaried members engaged in retail drug sales. This business can be extremely profitable. A middle-ranking member of a drug corner—typically a young man in their twenties—can be in the top half or quartile of earnings in the city. Yet while most current members are currently adults, almost all gang members were recruited between the ages of 10 and 17 (Blattman

et al., 2021; Carvalho and Soares, 2016). Typically, children as young as 10 begin socializing with the gang and doing part-time jobs as lookouts and carriers before eventually becoming one of the more formal and salaried members as an adolescent between the ages of 13 and 17. Few make it to the top and the high-earning levels, of course. Many are jailed, killed, or escape. As a result, gang membership is a cycle of a select tier of victims recruiting victims.

On occasion, gangs coerce recruits, in the sense of being forced to join and stay under the threat of violence. Most recruitment, however, is forced in the sense that gangs use deception and narcotics to recruit members, alongside the promise of future rewards. Anthropological studies situated in Latin America understand engagement in crime and violence by youth as a way, not to counter, but rather to try to fit into the mainstream for marginalized youth (Bourgois et al., 2003). They understand engagement in gangs as a way of obtaining successful adulthood in contexts where the standard pathways to adulthood—such as acquiring education or obtaining formal jobs—are not really available (e.g., Baird, 2018; Butti, Butti).

Just as importantly, once they join, gang members often are not allowed to leave without the severe menace of penalty. While many members may "age out" without penalty (especially after long periods of incarceration), younger and active gang members (i.e., those not in prison) may be killed if they indicate a desire to leave. One reason is the fear that they could be used as informants by police or could ally with a rival criminal organization. Older adolescents (under 18) and young adults are often at the greatest risk, as the youngest recruits do not yet have incriminating or valuable information. Still, any salaried member faces a menace of penalty for leaving.

Unfortunately, most of what we know about recruitment is qualitative and subjective. We know of no systematic data on an adolescent's perceptions of what gang-life offers. How do they imagine gang life? What are their beliefs about the benefits and risks of gang life versus their alternatives, and are they correct? How do gangs manipulate these beliefs or choices? What would be the effect of reducing (or eliminating) structural barriers which prevent many marginalized youth to obtain a successful adulthood in the legal sector? We believe we need data and experimentation to answer these questions and design better anti-recruitment interventions.

Our approach draws on several social science traditions. In economics, traditional expected utility theory assumes people choose a career based on information they process in sophisticated ways (Becker, 1968). Of course, psychology and behavioral science have taught us that people tend not to seek out enough information, are vulnerable to manipulation, and can hold onto persistently erroneous beliefs (Tversky and Kahneman, 1992; Gilboa and Schmeidler, 2004; Benoit and Dubra, 2011).

For six years we have been conducting qualitative interviews with current gang members,

ex-members, survivors, and not-yet-involved children and adolescents—all as part of a larger set of projects to understand and combat these armed groups. Our interviews suggest that gangs deceive adolescents about their options inside and outside the gang. Thus, recruits undervalue their potential outside the armed group; they underestimate the risks of gang life (especially arrest and death); and they overestimate their likely success as criminals. Gangs foster these beliefs to exploit a group of (eventual) adult members with no other career options.

Such lack of aspirations and misperceptions is consistent with the literature on non-criminal decision-making in labor economics and psychology. For example, young people's choice of occupation is influenced by their reference group and surroundings. Students mostly rely on the information on earnings from workers surrounding them (Jensen, 2010). If students live in poor neighborhoods, they mainly see low-paid jobs and underestimate their returns to education. Experiments show that providing information on the returns to education increases enrolment. People are often unaware of their misperceptions. Many people follow a heuristic of what you see is all there is, considering information exclusively from what is in front of them, a form of "availability bias" (Enke, 2020). This bias worsens in more complex problems. When there is a high degree of risk involved, selection problems also worsen people's misperceptions. In risky project decisions, people tend to remember the successes and forget to account for the failures that are difficult to observe (Esponda and Vespa, 2018). There are parallels to the decision to join a gang. Youth in poor neighborhoods, with low mobility, may underestimate opportunities outside their reference group, including the value of education. Selection problems worsen the situation, as youth are likely to observe successful gang members, not those who are imprisoned or killed. Gangs may exploit these biases.

To the best of our knowledge, ours will be the first systematic, large-scale study of the decision to enter armed groups, including sources of deception and other even more coercive tactics (such as the use of drugs and addiction in adolescents).

The preliminary paper is structured in six sections including this introduction. Section 2 develops a life-cycle model of preferred career path for adolescents. Section 3 presents our survey design. Section 4 describes the sample selection process and reports descriptive statistics on about 15% of our target sample. Section 5 presents our plans for field experiments. Section 6 discusses ethics considerations.

2 A model of life-path preferences

In this model, an agent chooses the life-path option that gives him the greatest discounted life-time utility. He chooses from a feasible choice set subject to a budget constraint.

The objective of the exercise is to model the decision-making process that leads a kid to prefer a given life-path option in an unconstrained setting (we will call this their *aspirational life path*) and in a constraint setting where a budget constraint can rule out several paths because of structural barriers (we will call this their *expected life path*).

In the aspirational life path, the individual chooses from ($j \in J$), where J is the complete set of options. In the expected life path, the individual chooses from ($j \in \bar{J}$) subject to a budget constraint, where $\bar{J}$ is the set of options which present no financial barriers to the individual (the *feasible* set).

The choice we are primarily interested in understanding is the constrained decision. An individual i at time t is faced with the decision of choosing which life path he would like to follow ($j \in \bar{J}$), which gives him the highest discounted expected utility over his lifetime ending at time T subject to a budget constraint. We assume individuals have standard constant relative risk aversion preferences, governed by their coefficient of risk aversion, ϕ_i.

This should be understood as a hypothetical exercise aimed at understanding how youth project themselves into the future, what are the perceived benefits and costs of each life option, and which life paths they find feasible to pursue.

The expected utility of each career path will be determined by 4 key sets of factors. First, monetary rewards, denoted by $G_{it}^{j}(Y)$, which are beliefs about the distribution of earnings in life path j at time t. Second, non-monetary rewards, denoted by R_{it}^{jk}, which include beliefs about success securing a partner (R_{it}^{j1}), social status and reputation (R_{it}^{j2}), and the level of enjoyment of the day-to-day activities of the job (R_{it}^{j3}). Third, monetary costs, denoted by C^{j}, which include the direct costs of attending school (C^{j_b}) and university (C^{j_u}) as well as the opportunity cost of studying ($O_i^{j_b}$ or $O_i^{j_u}$). The direct monetary costs do not enter the life-time utility of the individuals, just determine whether these options are feasible for them. Options will be ruled out of the feasible set if the individual cannot afford the direct costs of education. $Y_f \leq C^{j_b} + C^{j_u}$ will rule out j_u and $Y_f \leq C^{j_b}$ will rule out j_b. Y_f is the family income available to study. Finally, non-monetary costs, denoted by B_{it}^{jk}, which include beliefs on family approval of the preferred career path (B_{it}^{j1}), anxiety or discomfort of executing the day-to-day activities of the job (B_{it}^{j2}), and E_i^j which includes the effort required to graduate from school ($E_i^{j_s}$) and university ($E_i^{j_u}$).

Each life path's utility will include η_j, which reflects the unobservable life path-specific factors that affect lifetime utility. The future benefits and costs will be discounted by a dis-

count factor δ_i and a present bias factor β_i. Finally, ϵ_j is a random term which is individual- and life path-specific, observable to individual i at the time of choosing their life path, but not to the econometrician.

The individual will be given a choice between J life options: legal life path without high school completion, denoted by j_s; legal life path with high school completion, denoted by j_b; legal life path with tertiary education, denoted by j_u; and illegal life path or gang entry, denoted by j_g.

For traceability, we will assume four different time periods. First, pre $t = 0$, when the individual takes the decision of the preferred path. Second, $t = 0$, where the individual will be studying if he chooses life-path j_b or j_u or working if he chooses life-path j_s or j_g. This period is the "school" period. Third, $t = 1$, where the individual will be studying if he chooses life-path j_u or working otherwise. This period is the "university" period. Finally, $t = 2$ to T, where the individual will be working in all life-path options.

The expected utility of agent i of choosing life-path s (legal life path without high school completion) is:

$$E_i(U_{ij_s}) = \theta\Big[\beta_i \sum_{t=0}^{T} \delta_i^t E\big(\frac{Y_{it}^{j_s(1-\phi_i)}}{1-\phi_i}\big)\Big] + \sum_{k=1}^{3} \alpha_k \Big[\beta_i \sum_{t=0}^{T} \delta_i^t \big(\frac{R_{it}^{j_s k(1-\phi_i)}}{1-\phi_i}\big)\Big] - \sum_{k=1}^{2} \sigma_k \Big[\beta_i \sum_{t=0}^{T} \delta_i^t \big(\frac{B_{it}^{j_s k(1-\phi_i)}}{1-\phi_i}\big)\Big] + \eta_{j_s} + \epsilon_{j_s}$$

The expected utility of agent i of choosing life-path b (legal life path with high school completion) is:

$$E_i(U_{ij_b}) = -\lambda\Big[E_i^{j_b} + O_i^{j_b}\Big] + \theta\Big[\beta_i \sum_{t=1}^{T} \delta_i^t E\big(\frac{Y_{it}^{j_b(1-\phi_i)}}{1-\phi_i}\big)\Big] + \sum_{k=1}^{3} \alpha_k \Big[\beta_i \sum_{t=1}^{T} \delta_i^t \big(\frac{R_{it}^{j_b k(1-\phi_i)}}{1-\phi_i}\big)\Big] - \sum_{k=1}^{2} \sigma_k \Big[\beta_i \sum_{t=1}^{T} \delta_i^t \big(\frac{B_{it}^{j_b k(1-\phi_i)}}{1-\phi_i}\big)\Big] + \eta_{j_b} + \epsilon_{j_b}$$

And the expected utility of agent i of choosing life-path u (legal life path with tertiary education) is:

$$E_i(U_{ij_u}) = -\lambda\Big[\Big(E_i^{j_b} + O_i^{j_b}\Big) + \beta_i\delta_i\Big(E_i^{j_u} + O_i^{j_u}\Big)\Big] + \theta\Big[\beta_i \sum_{t=2}^{T} \delta_i^t E\Big(\frac{Y_{it}^{j_g(1-\phi_i)}}{1-\phi_i}\Big)\Big] + \sum_{k=1}^{3} \alpha_k\Big[\beta_i \sum_{t=2}^{T} \delta_i^t\Big(\frac{R_{it}^{j_u k(1-\phi_i)}}{1-\phi_i}\Big)\Big] - \sum_{k=1}^{2} \sigma_k\Big[\beta_i \sum_{t=2}^{T} \delta_i^t\Big(\frac{B_{it}^{j_u k(1-\phi_i)}}{1-\phi_i}\Big)\Big] + \eta_{j_u} + \epsilon_{j_u}$$

Where, across the equations above, θ is the utility value of log income, σ_k is the utility value of the non-monetary rewards, α_k is the utility value of the non-monetary costs, and λ is the utility value of the costs of investing in education.

The expected utility the illegal path has two extra terms: P_i^{jail}, the probability of entering prison at t^* for x years and P_i^{hom}, the probability of suffering a homicide at time T^h. For simplicity, we assume the legal paths do not yield a significant risk of prison or homicide.

$$\begin{aligned}
E_i(U_{ij_g}) = (1-P_i^{hom})\Bigg((1-P_i^{jail})\Bigg[\theta\Big[\beta_i \sum_{t=0}^{T} \delta_i^t E\Big(\frac{Y_{it}^{j_s g(1-\phi_i)}}{1-\phi_i}\Big)\Big] + \sum_{k=1}^{3} \alpha_k\Big[\beta_i \sum_{t=0}^{T} \delta_i^t\Big(\frac{R_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big)\Big] - \sum_{k=1}^{2} \sigma_k\Big[\beta_i \sum_{t=0}^{T} \delta_i^t\Big(\frac{B_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big)\Big]\Bigg] + \\
(P_i^{jail})\Bigg[\theta\Big[\beta_i \sum_{t=0}^{t^*} \delta_i^t E\Big(\frac{Y_{it}^{j_s g(1-\phi_i)}}{1-\phi_i}\Big)) + \beta_i \sum_{t=t^*+x}^{T} \delta_i^t E\Big(\frac{Y_{it}^{j_s g(1-\phi_i)}}{1-\phi_i}\Big)\Big] + \\
\sum_{k=1}^{3} \alpha_k\Big[\beta_i \sum_{t=0}^{t^*} \delta_i^t\Big(\frac{R_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big) + \beta_i \sum_{t=t^*+x}^{T} \delta_i^t\Big(\frac{R_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big))\Big] - \\
\sum_{k=1}^{2} \sigma_k\Big[\beta_i \sum_{t=0}^{t^*} \delta_i^t\Big(\frac{B_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big)) + \beta_i \sum_{t=t^*+x}^{T} \delta_i^t\Big(\frac{B_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big))\Big]\Bigg]\Bigg) + \\
P_i^{hom}\Bigg((1-P_i^{jail})\Bigg[\theta\Big[\beta_i \sum_{t=0}^{T^h} \delta_i^t E\Big(\frac{Y_{it}^{j_s g(1-\phi_i)}}{1-\phi_i}\Big)\Big] + \sum_{k=1}^{4} \alpha_k\Big[\beta_i \sum_{t=0}^{T^h} \delta_i^t\Big(\frac{R_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big)\Big] - \sum_{k=1}^{2} \sigma_k\Big[\beta_i \sum_{t=0}^{T^h} \delta_i^t\Big(\frac{B_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big)\Big]\Bigg] + \\
(P_i^{jail})\Bigg[\theta\Big[\beta_i \sum_{t=0}^{t^*} \delta_i^t E\Big(\frac{Y_{it}^{j_s g(1-\phi_i)}}{1-\phi_i}\Big) + \beta_i \sum_{t=t^*+x}^{T^h} \delta_i^t E\Big(\frac{Y_{it}^{j_s g(1-\phi_i)}}{1-\phi_i}\Big)\Big] + \\
\sum_{k=1}^{3} \alpha_k\Big[\beta_i \sum_{t=0}^{t^*} \delta_i^t\Big(\frac{R_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big)) + \beta_i \sum_{t=t^*+x}^{T^h} \delta_i^t\Big(\frac{R_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big))\Big] - \\
\sum_{k=1}^{2} \sigma_k\Big[\beta_i \sum_{t=0}^{t^*} \delta_i^t\Big(\frac{B_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big)) + \beta_i \sum_{t=t^*+x}^{T^h} \delta_i^t\Big(\frac{B_{it}^{j_g k(1-\phi_i)}}{1-\phi_i}\Big))\Big]\Bigg]\Bigg) + \eta_{j_g} + \epsilon_{j_g}
\end{aligned}$$

Each agent will choose among options ($j \in \bar{J}$) to maximize the expected present value of lifetime rewards subject to a budget constraint $Y_f \geq C^{j_b} + C^{j_u}$:

$$max_{j \in \bar{J}} E_i(U_{ij})$$
$$s.t. Y_f \geq C^{j_b} + C^{j_u}$$

3 Survey design

To mitigate the identification problem concerning the separation of preferences, expectations, and structural barriers, we designed a survey instrument to gather data on subjective expectations of the main parameters of the model, as well as financial constraints.

In particular, we ask adolescents for data on (i) the subjective expectations of income conditional on each life path; (ii) the subjective expectations about non-pecuniary outcomes conditional on each life path (success of finding a partner, status or social recognition, enjoyment of day-to-day activities, anxiety or discomfort executing day-to-day activities, and approval of life path by family, as well as the probabilities of going to prison or being killed; (iii) the life path options which are within the budget constraint of the individual; (iv) the stated preferred life path with and without budget constraints; and (v) the risk aversion coefficient, the time discount coefficient and the present-bias coefficient.

Because of the sensitivity of the illegal life path and some of the parameters we measure, we ask about them indirectly. Throughout the survey, adolescents are asked how 10 students just like them (10 "clones") or 10 friends would behave or choose under different hypothetical scenarios. After extensive pretesting, detaching the question from the individual and asking for "clones" or friends, instead of the individual's choice, was the method that made respondents (and parents and school staff) more comfortable.

Measuring the parameters in a probabilistic way also allows respondents to express uncertainty about these future outcomes, allowing us to unmask potential heterogeneity in these beliefs. We explain each of the parameters we measure below.

Outcome measure The survey collects the adolescents' aspirational and expected career path choices. Specifically, we collect data on beliefs about the probability of following each of the four career paths in the complete set J: Legal life path without high school completion (j_s); legal life path with high school completion (j_b), legal life path with tertiary education (j_u), and illegal life path or gang entry (j_g). In the *complete set*, individuals can choose from the set of 4 options. To define their *feasible set*, we ask individuals two sets of questions to determine which life paths are financially feasible for them. Below we discuss how we measure budget constraints.

Expected earnings by life path We follow Dominitz and Manski (1997) and elicit subjective probabilistic income distributions under the different career paths $j \in J$ by getting information on points in their subjective cumulative distribution function. To adapt this method to our research setting, we follow Delavande et al. (2011); Brimble (Brimble) and adapt their suggestions to design an exercise appropriate to work with children and adolescents. We define five intervals based on the income distribution of Medellín ($I_n; n = \{1, ..., 5\}$). Each interval has a lower (Th_{n-1}) and upper (Th_n) thresholds. We ask each respondent to distribute 10 hypothetical individuals "just like them" in these intervals.

For the variable Y_{ij} we can write the intervals and their subjective probabilities P_n as:

$$P_n(Y_{ij} \in I_n) = P_n(Th_{n-1} < Y_{ij} \leq T_n) \text{ for } n = \{1, 2, 3, 4, 5\}$$

Therefore, for each individual i, we can calculate their perceived probability of being below a given income threshold (Th_k with $k = 1, 2, 3, 4, 5$) given each person's information set (φ) about which formation we remain agnostic:

$$F_{i,j,k} = P(Y_{ij} \leq Th_k|\varphi) \text{ for } k = \{1, 2, 3, 4, 5\}$$

We calculate the perceived probability of being below a certain threshold ($Y_{ij} \leq Th_n$) by adding the probabilities of being in the intervals $P_k(Y_{ij} \in I_k$ for $k = \{1, ..., n\}$ which give us the cumulative probability F_n:

$$F_n(Y_{ij} \leq Th_n) = \sum_{k=1}^{n} P_k(Y_{ij} \in I_k)$$

This gives us four points in each individual's career specific income cumulative distribution function. We assume the functional form of the income distribution to be log-normal, following McKenzie et al. (2013). Let $G(Y_{i,j}; \mu_{i,j}, \sigma^2_{i,j})$ denote the cumulative distribution function of a log-normal distribution, where $logY_{i,j} \sim N(\mu_{i,j}, \sigma^2_{i,j})$. For each respondent and each career path, we estimate $\mu_{i,j}$, $\sigma^2_{i,j}$ to solve the least squares problem:

$$min_{\mu_{i,j},\sigma_{i,j}} \sum_{k=1}^{4} \left[F_{i,j,k} - G(Y_{i,j,k}\,;\, \mu_{i,j}\sigma^2_{i,j})^2 \right]$$

From this fitted distribution, which is different for each individual and career path, we can obtain different moments of interests (mean, median, standard deviation).

Probability of jail or death in the criminal life path We collect information on P_i^{jail} and P_i^{hom}, which represent the probabilities of entering prison or being killed, conditional on following the illegal life path.

Expected non-monetary rewards by life path We collect information on: the success of finding a partner, social status or reputation, and enjoyment of day-to-day activities conditional on each life path. We conceptualized all these as binary outcomes, whether individuals will succeed in finding a partner, whether individuals will become someone recognized in their community, or whether individuals will like the day to day of their job. Following Delavande and Zafar (2019), we measure these binary outcomes as probabilistic expectations.

Expected non-monetary costs by life path We collect information on: the anxiety or discomfort of executing day-to-day activities, the effort needed to graduate from school or university, and family disapproval of the life path chosen, conditional on each life path. Similar to the non monetary rewards, we measure discomfort and family disapproval as probabilistic expectations directly elicited in the survey. We measure the effort needed on a scale of 1 to 10.

Budgets In the *complete set*, individuals can choose from the set of four life-path options. As we mention above, to define their *feasible set*, we ask individuals two sets of questions to determine which life paths are financially feasible for them. First, we ask individuals whether they believe they could pay (either with family or friends support or accessing a scholarship) for their studies if they decide to study until they finish grade 11, and whether they could pay for their studies if they decided to go and were accepted to university. If the individual report not being able to afford studying until grade 11, their financial feasible set ($\bar{J}_i^f$) will contain j_s, j_g, if the individual reports being able to afford until grade 11 but not university, their financial feasible set will contain j_s, j_g, j_b. Finally, if the individual reports being able to afford studying until grade 11 and university, their financial feasible set will contain j_s, j_g, j_b, j_u. The feasible set ($\bar{J}_i$) will include the life paths the individuals believe their families can afford and the ones they will succeed in, $\bar{J}_i = \bar{J}_i^{ps} \cap \bar{J}_i^f$.

Furthermore, we also collect information on direct costs and opportunity costs to pursue the different career paths. The direct costs of school (C^{jb}) and university (C^{ju}) are only included in the budget constraint. The opportunity costs (O_i^{jb}) and (O_i^{ju}) which represent the income individuals forgo if they continue in education past the mandatory age—grade 9 in our setting. Individuals incur in this opportunity cost if they finish high school or continue

into higher education. We model this opportunity cost as their expected income if they were to drop from education after finishing grade 9, ($E[Y_i^{js}]$).

Risk preferences We elicit the risk aversion parameters via a certainty equivalent game derived from the Global Preferences Survey (GPS) for Colombia (Falk et al., 2018, 2016). The exercise remains broadly similar except for the quantities, which we multiply by 10 to mimic "salary" style magnitudes.

The exercise fixes a lottery that pays $w1$ with probability $\frac{1}{2}$ and $w2$ with probability $\frac{1}{2}$, and asks the respondent to choose between this lottery and a certain payment of $w3$. The lottery payment remains fixed while certain payment changes. Each respondent answers five interdependent questions that offer different $w3$ based on their previous answers to zoom in around the individual's certainty equivalent.

We identify indifference levels between the lottery and the certain equivalent as the point where the individual changes from choosing the lottery to the risk-less payment. We calculate the risk aversion parameter assuming a constant relative risk aversion form for the utility function governed by ϕ_i, which is heterogeneous between agents:

$$U_i(x) = \frac{x^{(1-\phi_i)}}{1-\phi_i}$$

To back out ϕ_i, we look for ϕ_i which equates the expected utility from the lottery and the certainty equivalent. We also assume that the $U(0) = 0$.

$$\frac{1}{2}\frac{w1^{(1-\phi_i)}}{1-\phi_i} + \frac{1}{2}\frac{w2^{(1-\phi_i)}}{1-\phi_i} = \frac{w3^{(1-\phi_i)}}{1-\phi_i}$$

where $w1$, $w2$, and $w3$ are the hypothetical payments the individual can choose from in the certainty equivalent survey exercise.

Time preferences Following Patnaik et al. (2020), we jointly estimate the discounting and the curvature parameters from survey questions that elicit time preferences. We not only estimate the discount factor but also the present bias following Laibson (1997). We elicit time preferences via a "money earlier or latter" exercise, were we ask a series of five interdependent hypothetical binary choices between immediate and delayed (in one year) financial rewards following the Global Preferences Survey for Colombia (Falk et al., 2018, 2016). In the series of five questions, the immediate reward remains the same but the delayed reward changes based on previous answers to zoom in into the switching point between a smaller immediate reward and a bigger delayed reward. This exercise remains broadly similar

as in the Global Preferences Survey, same except for the quantities, which we multiply by 10 to mimic "salary" style magnitudes.

To be able to separate the time discount from the present bias, we ask the same series of five interdependent hypothetical binary choices but this time between a financial reward in one year and a financial reward in two years following. Individuals choose between two payment options $\{T1, x1\}$ and $\{T2, x2\}$ with $T1 < T2$, where we fix $x1$ and allow $x2$ to change. We set as the point of indifference the point where the individual switches from the earlier to the later payment.

We use the ϕ_i estimated and the individual's elicited time preferences to calculate δ_i and β_i. We follow Laibson (1997) and assume present-biased preferences. Therefore, the discounted present value utility is of the form:

$$U_t = E_t\Big[U(x_t) + \beta \sum_{t=t+1}^{T} \delta^t U(x_{t+1})\Big]$$

For an indifference pair $\{T1, x_1\}$ and $\{T2, x_2\}$, where $T1$=today and $T2$=one year from today and a given ϕ_i, we calculate the $\delta_i \times \beta_i$ as:

$$\frac{x_1^{1-\phi_i}}{1-\phi_i} = \beta_i \times \delta_i [\frac{x_2^{1-\phi_i}}{1-\phi_i}]$$

We then repeat the exercise for an indifference pair $\{T1, x1\}$ and $\{T2, x2\}$, where $T1$=one year from today and $T2$=two years from today, to calculate δ_i:

$$\beta_i \times \delta_i \frac{x_1^{1-\phi_i}}{1-\phi_i} = \beta_i \times \delta_i^2 \frac{x_2^{1-\phi_i}}{1-\phi_i} =$$
$$\frac{x_1^{1-\phi_i}}{1-\phi_i} = \delta_i \frac{x_2^{1-\phi_i}}{1-\phi_i}$$

Once we have the $\delta_i \times \beta_i$ and the δ_i parameter for each individual, we back out the β_i as:

$$\beta_i = \frac{\delta_i \times \beta_i}{\delta_i}$$

Note that for this exercise we are assuming individuals do not have any additional income other than the prize received in the hypothetical exercise.

4 Sample selection and descriptive statistics

4.1 Data sources and sample selection

Because education is nearly universal in Medellín until roughly grade 9—the legal minimum educational requirement nationwide, we are conducting the surveys at schools. To simplify the logistics, we focus on public schools only—which are broadly the ones at the highest risk of student recruitment. To select our analytical sample of students, we target schools located in areas of the city where adolescents face a higher risk of recruitment. To do this, we use three administrative datasets: school enrollment, school environment surveys, and arrest data. This data allow us to know which schools have a higher rate of alumni arrested for allegedly gang-related crimes, which schools have a higher perceived gang presence by school teachers and students, and which schools have a higher dropout rate.

School enrollment data This dataset includes all elementary, middle, and high-school students in Medellín who are enrolled in public educational institutions from 2006 to 2021. We organize this dataset at the student-year level. The dataset includes student national identification numbers, the school they are enrolled in, basic socioeconomic information on the student, and enrollment status. This data is from the Secretariat of Education of Medellín.

School environment survey This is a school-based survey where teachers and students describe the school environment. This survey includes a question on the perceived presence of gangs in or near the school and the level of violent conflicts happening in the school. The city runs this survey annually and includes information at the school-year level. This data is also from the Secretariat of Education of Medellín.

Police arrest data The arrest data includes the universe of arrests in Colombia from 2004 to 2019. This dataset describes each individual arrest, including the national identification number of the person arrested, the reason for the arrest, and the exact time, day, and location. This data is from the National Police of Colombia.

Sample selection We are targeting the universe of male students enrolled in grade 7 in the schools we deem at highest risk. To select the sample schools, we stratify them into four groups, in a 2×2 matrix featuring two dimensions. On one dimension we split schools above and below the median of gang presence using the school environment survey. On the other dimension we split schools above and below the median of school-level arrest rates,

which we produce by merging the school enrollment data with arrest data, allowing us to determine the rate of alumni arrested for allegedly a gang-related incident. We then sampled 97 schools, giving priority to the high-high-risk category (where we include all schools).

4.2 Descriptive statistics

Below we present early results from the original survey, which we launched in August 2021 and is currently running in the 97 schools in our sample. As we discussed above, we are currently surveying approximately 5,000 boys, representing two-thirds of the universe of male students enrolled in grade 7 in our sample of schools. We estimate losing about one-third of our sample because of the difficulties in getting parental consent.

Sample characteristics Table 1 presents the characteristics of the 750 students for whom we currently have available data. Students in our sample are on average 13 years old, most are Colombian born but almost 10% are from Venezuela. Most identify as *mestizo* or white, and some as afrocolombians, roughly reflecting the ethnic composition of Medellín. Almost 80% belong to low- or middle-low-income families and a few report having disabilities.

Stated choices As we discuss in section 2, we are interested in individual decisions on career paths across the following options: legal life path without high school completion, legal life path with high school completion, legal life path with tertiary education, and illegal life path or gang entry. In this section, we refer to these different paths as Group 1 (low-skilled path), Group 2 (medium-skilled path), Group 3 (high-skilled path), and Group 4 (gang path), respectively.

Figure 1 presents the distributions of adolescents' choices across three different sets of questions: Sub-Figure (a) presents the distributions of career paths with no constraints (all career paths are available), Sub-Figure (b) presents the distributions of career paths with one constraint (high-skilled career or group 3 is not available), and Sub-Figure (c) presents the distributions of career paths with two constraints (medium-skill and high-skill careers or groups 2 and 3 are not available). In each case, boys distribute ten *clones* We observe several interesting features. First, some boys, even when all options are available to them, reveal a positive probability of joining a gang. Second, as expected, this probability grows as boys have fewer options. Third, boys are relatively optimistic about their future, with the distributions on higher-skilled options slightly more to the right compared to other distributions.

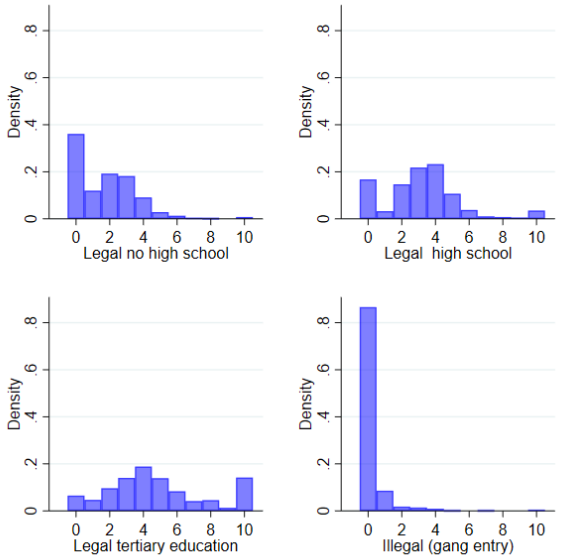
Table 1: Socioeconomic characteristics

	Obs (1)	Mean (2)	Std. Dev. (3)	Min (4)	Max (5)
A. Entry Survey					
Age	750	13,189	1,058	11	17
Country of birth					
Colombia	750	0,907	0,291	0	1
Venezuela	750	0,091	0,287	0	1
Other	750	0,003	0,052	0	1
Ethnic groups					
Afrocolombian	702	0,121	0,326	0	1
Palenquero	702	0,001	0,038	0	1
Raizal	702	0,003	0,053	0	1
Indigenous	702	0,026	0,158	0	1
Mulato	702	0,043	0,202	0	1
Mestizo	702	0,301	0,459	0	1
White	702	0,299	0,458	0	1
None	702	0,207	0,405	0	1
B. Administrative data					
Income					
Low	750	0,315	0,465	0	1
Mid-low	750	0,484	0,500	0	1
Mid-high	750	0,183	0,387	0	1
High	750	0,008	0,089	0	1
Highest	750	0,001	0,037	0	1
Disabilities	750	0,033	0,180	0	1

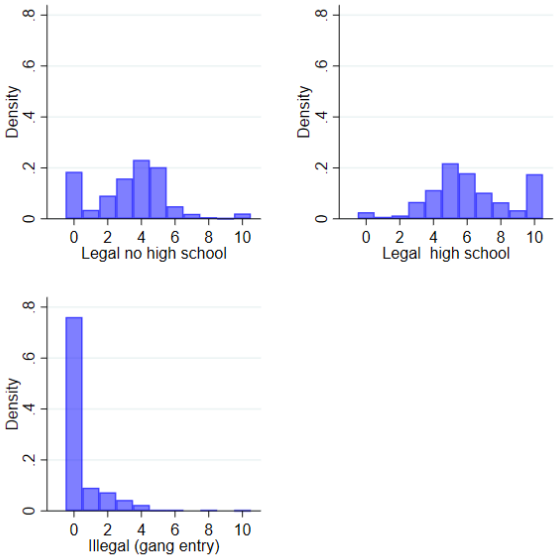
Notes: The table displays the socioeconomic characteristics of the 750 students surveyed to date.

Figure 1: Choice set between the 4 groups

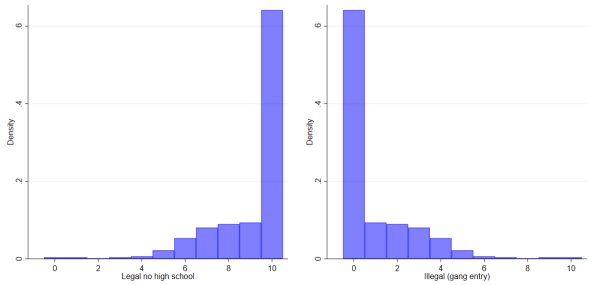
(a) *Choice set 4 groups (G1, G2, G3 and G4)*



(b) *Choice set 3 groups (G1, G2 and G4)*



(c) *Choice set 2 groups (G1 and G4)*



Notes: The figure presents the distributions of adolescents' choices across three different sets of questions: Sub-Figure (a) presents the distributions of career paths with no constraints (all career paths are available), Sub-Figure (b) presents the distributions of career paths with one constraint (high-skilled career or group 3 is not available), and Sub-Figure (c) presents the distributions of career paths with two constraints (medium-skill and high-skill careers or groups 2 and 3 are not available).

Table 2: Constrains

	Mean (1)	Median (2)	Std. Dev. (3)	% (4)	Feasible sets (5)
Dropout before finishing high school	0.108	0	0.311	10.80	G1 and Combo
Cannot afford higher education	0.085	0	0.279	8.53	G1, G2 and Combo
Know-how restricted	0.197	0	0.398	19.73	Full set
Socially distant	0.235	0	0.424	23.47	Full set
No constraints	0.529	1	0.499	52.93	Full set

Notes: The table displays describes descriptive statistics on the distribution of constraints as reported by students. We create dummy variables that takes the value 1 if the student reported each of these constraints. In column (5) we report the implied feasible set given the reported constraint

Structural constraints We focus on four types of constraints: dropping out before finishing high school—for instance because they need to contribute financially at their households, being unable to afford higher education, limitations on know-how restrictions, and lack of social connections to continue their education.

Table 2 describes the distribution of constraints as reported by students. We create a dummy variable that takes the value 1 if the student reported each of these constraints. In the last column we report the implied feasible set given the reported constraint. About half of the boys report no constraints and among the ones who report, most respond that they are either know-how restricted or feel they are socially distant which might limit their opportunities.

Table 3 describes the distribution of *clones* across different career paths, conditional on the constraints reported. We see that boys that report no constraints seem to reveal a lower probability of following the gang path. Furthermore, we see that there is broad consistency between the constraints and the career paths boys see more likely, as chances of reaching a high-skill occupational career are higher for those with no or less biding constraints. For reference

Expectations about future income As we describe above in section 3, we ask students to distribute 10 hypothetical individuals "just like them" across different income ranges for the different occupational paths. We then use this information to generate a fitted distribution for each individual and career path, and from that we obtain different moments of interest.

Table 4 reports descriptive statistics on the average monthly income reported by students across the different occupations for someone in their early twenties. Units are in COP$ millions, roughly equivalent to US$ 740 adjusting for purchasing power parity. As we expected,

Table 3: Distribution of clones for the three different choice sets and constraints

	Group 1 (1)	Group 2 (2)	Group 3 (3)	Group 4 (4)
Dropout before finishing high school	2.41	3.33	3.89	0.36
Can't afford higher education	2.21	3.56	3.95	0.27
Know-how restricted	1.67	3.35	4.71	0.27
Socially distant	1.79	3.25	4.57	0.39
No constraints	1.67	3.09	5.04	0.21

Notes: The table displays describes the mean distribution of clones across different career paths, conditional on the constraints reported.

Table 4: Expectations about future incomes

	Mean (1)	Std. dev. (2)	Min (3)	Max (4)
Group 1	1.337	0.782	0.001	4.510
Group 2	1.982	0.713	0.001	4.510
Group 3	3.238	0.817	0.842	4.510
Gang	1.544	1.363	0.001	4.510

Notes: The table reports descriptive statistics on the average monthly income reported by students across the different occupations for someone in their early twenties. Units are in COP$ millions, roughly equivalent to US$ 740 adjusting for purchasing power parity.

students report, on average, that there is a skill premium: the higher the skills the higher the expected income. Students locate the combo somewhere between the low- and mid-skilled careers.

Expectations about non monetary costs and benefits We focus on three types of non-monetary benefits: status or reputation, whether they will enjoy work, and success with partners. We also ask boys about two types of non-monetary costs: harassment, stress or nervousness, and whether they will disappoint their family. As we do with other questions, we ask boys to distribute their *clones* across occupations that will yield the different types of non-monetary benefits and non-monetary costs.

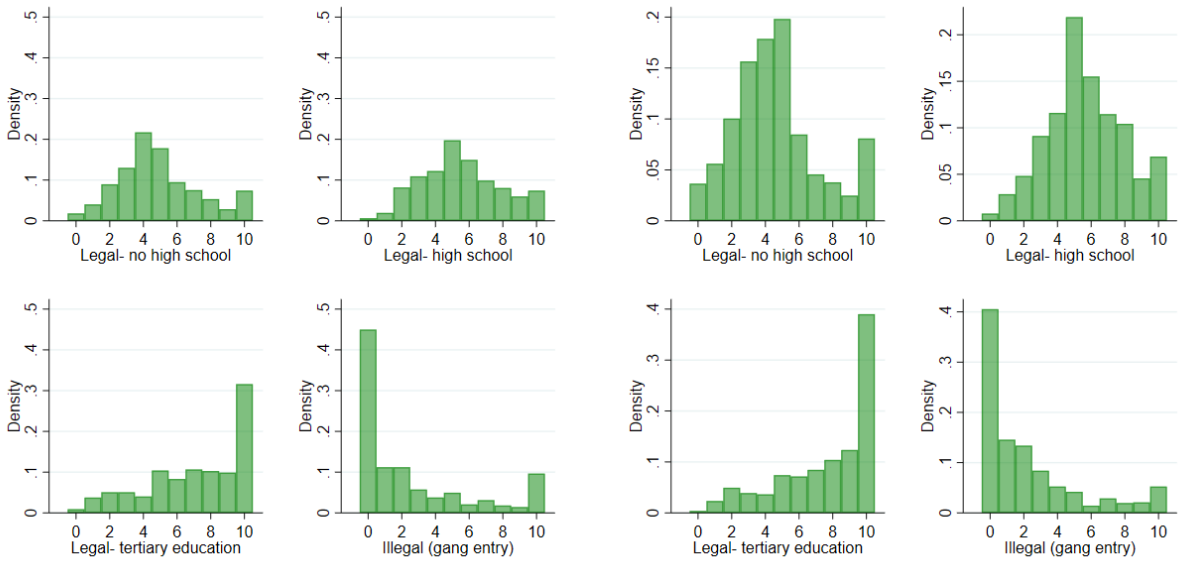
Figure 2 presents the distributions across occupations for the three different non-monetary benefits. We see that, in general, a not negligible share of students in our sample believe that the illegal path would render status and reputation, an enjoyable work, and success with partners. The best perceived path to do so seems to be a high-skill occupational path.

Figure 3 depicts the distributions across occupations for the three different non-monetary

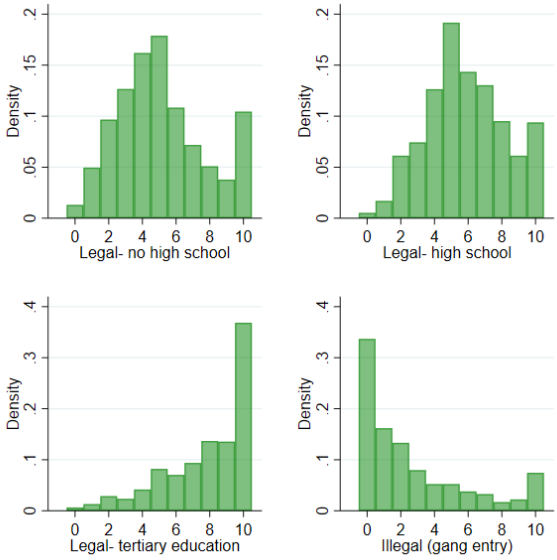
Figure 2: Non-monetary benefits

(a) *Status or Reputation*

(b) *Enjoy work*



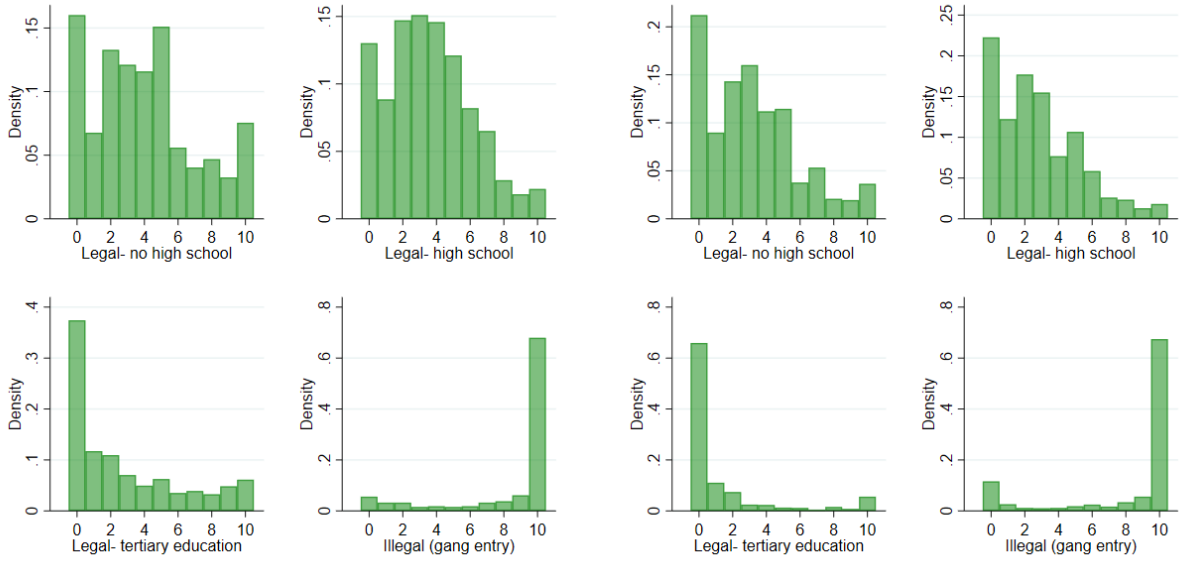
(c) *Success with partner*



Notes: The figure presents the distributions across occupations for the three different non-monetary benefits. Fub-Figure (a) presents the distributions of career paths according to status or reputation, Sub-Figure (b) presents the distributions of career paths for enjoying work, and Sub-Figure (c) presents the distributions of career paths according to the perceived success with women.

Figure 3: Non-monetary costs

(a) *Harassed, stressed or nervous* (b) *Disappointing the family*



Notes: The figure presents the distributions across occupations for two different non-monetary costs. Fub-Figure (a) presents the distributions of career paths when thinking about feelings of being harassed, stressed, or nervous, Sub-Figure (b) presents the same distributions considering the possibility of disappointing the family.

costs we focus on. Generally, we observe that the gang path is the more costly in terms of harassment or stress, or disappointing their families. Nonetheless, a share of students believe they would not disappoint their families with a life in the gang.

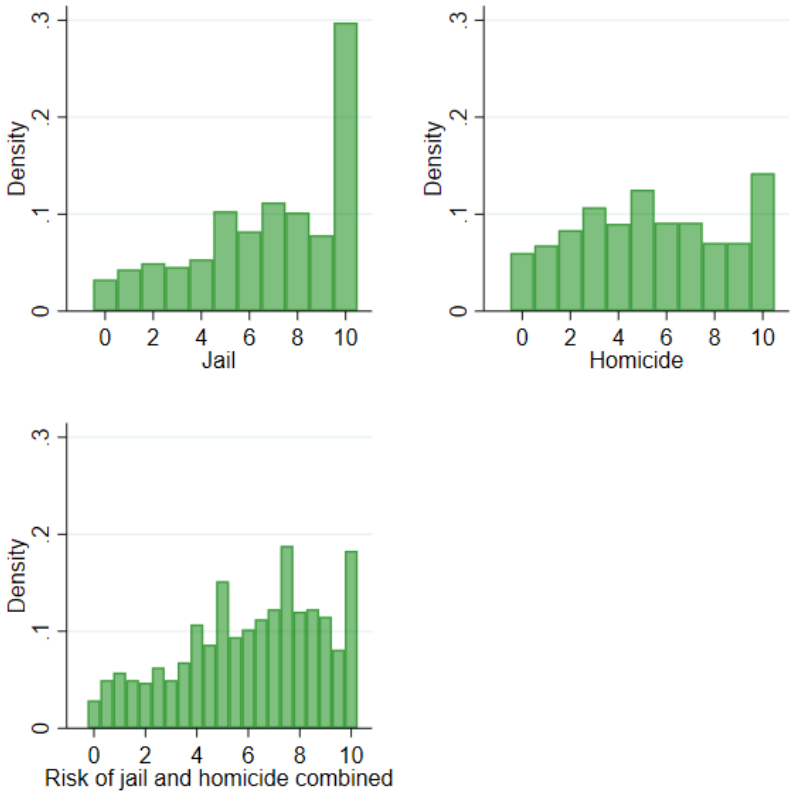
Probability of jail and homicide As we discuss above, we are asking students about their perceived probability of being jailed or murdered if they follow a criminal life path. Our pilots suggested the perceived probability of jail and homicide were too low for legal paths, hence we opted to keep this question only for the gang alternative. We ask students to answer how many, out of ten people, would be jailed or killed in their early twenties if they follow the illegal occupational path.

Figure 4 depicts the distributions for jail and homicide independently and for both combined. Broadly, boys perceive a high risk of both jail and homicide.

Risk and time preferences As we describe in section 3 in detail, we elicit risk and time preferences using games in our survey. We focus our analysis on our estimates for risk aversion, time discount and present bias jointly, and time discount independently.

Table 5 report the average coefficients for risk aversion, time discount and present bias

Figure 4: Probability of jail and homicide



Notes: The figure shows the distributions for jail and homicide for ten friends if they were part of the gang, either independently or for both risks combined.

Table 5: Time and risk preferences

	Mean (1)	Std. dev. (2)	Min (3)	Max (4)
ϕ	-0.261	2.376	-19.44	0.796
$\beta \times \delta$	0.682	0.297	0.000	3.476
δ	0.670	0.265	0.000	0.994

Notes: The table displays descriptive statistics for 3 parameters: risk aversion, time discount and present bias jointly, and time discount independently.

jointly, and the time discount independently. With an average risk parameter (ϕ) of roughly -0.26, we believe the students in our sample seem to be in the risk loving realm. On the other hand, our average estimates for time discount and present bias jointly ($\beta \times \delta$) and time discount independently (δ) are about 0.68 and 0.67, respectively. Compared to other student populations, it seems that our sample of students is more impatient and have dynamically inconsistent preferences, but their current self seems to be more patient than their future self (e.g., Patnaik et al., 2020). However, most other studies focus on a slightly older population of students, so we take these interpretations with caution.

Exploratory correlational analysis We now turn to examine the correlates of the recruitment risk, measured as the perceived probability of a gang life across the different potential constraints. As we discuss above, we are asking students to distribute their future *clones* under three different scenarios: no constraints (all career paths are available), one constraint (high-skilled career or group 3 is not available), and two constraints (medium-skill and high-skill careers or groups 2 and 3 are not available). We average the number of *clones* that would follow a gang path, standardize it with mean zero and standard deviation one, and use it as the outcome in our regressions.

Table 6 reports the results. Columns (1) through (10) report independent correlations and Column (11) present a regression of the outcome on all the dimensions we focus on in this report. Broadly, we see no surprises: more constraints associate with a higher risk, lower skill premiums associated with higher risk, higher perceived non-monetary rewards on the gang alternative associate with higher risk, lower perceived non-monetary costs on the gang alternative associate with higher risk, and more impatience associate with higher risk.

Table 6: Exploratory correlational analysis

	Gang recruitment risk index										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Number of restrictions	0.0912*										0.0599
	(0.0375)										(0.0338)
Mean income G1		0.133***									0.129***
		(0.0337)									(0.0382)
Mean income G2			0.00940								-0.00877
			(0.0376)								(0.0446)
Mean income G3				-0.113**							-0.104*
				(0.0404)							(0.0446)
Mean income G4					0.0719						0.0150
					(0.0390)						(0.0372)
Non-monetaty rewards						0.200***					0.192***
						(0.0420)					(0.0405)
Non-monetary cost							-0.106**				-0.0728
							(0.0388)				(0.0419)
Jail and homicide index								-0.0612			-0.0105
								(0.0404)			(0.0469)
Theta (Risk)									0.00313		-0.0521
									(0.0442)		(0.0315)
Beta-Delta (Time)										0.116***	0.135***
										(0.00162)	(0.00952)
Observations	750	750	750	750	750	750	750	750	750	750	750
R-squared	0.009	0.019	0.000	0.013	0.005	0.042	0.012	0.004	0.000	0.014	0.102

Notes: *** p<0.001, ** p<0.01, * p<0.05. The table displays an exploratory correlational analysis with Gang recruitment risk. Columns (1) to (10) report the coefficients and standard errors from ordinary least squares regressions of each individual covariate. Additionally, we included in column (11) a multivariable regression with all the previous variables

5 Intervention plans

5.1 Intervention ideas

We will use the in-person survey of adolescent males to identify the highest risk and implement and evaluate two interventions. We are aiming for a total experimental sample size of 1400 high-risk adolescent males—roughly 465 in three arms: a control arm and two treatment arms. We will select them from the sample of 7,000 youth in the baseline survey.

The 1,400 highest-risk males who consent to participate will be individually randomized within schools to one of the three arms. We have data on the classroom composition of all schools, as well as the social network of all surveyed adolescents (from the baseline survey)—enabling us to assess and account for spillovers within the classroom or network.

Both interventions will be conducted outside of school (offsite), both after school classes and on weekends, in small groups, over 2-3 months, during the summer and fall of 2023.

Treatment arm I: Increasing aspirations and expectations on legal career paths We are working with the Caja de Compensación Familiar de Antioquia (COMFAMA) and the Servicio Nacional de Aprendizaje (SENA) to design and implement an 8–10 weeks training and mentorship program—leveraging on the flexibility of curricula—that aims to raise adolescents' aspirations and expectations of success in legal occupations, counter any misinformation, and reduce the costs of understanding how to follow this path.

SENA is the largest public tertiary education provider in the country. During 2021, roughly 8.9 million people were trained by SENA in occupational careers or complementary courses. SENA has training centers throughout the country, with several major centers in Medellín. The largest center in Medellín trains about 70 thousand people per year. Occupational careers cover most disciplines, from robotics to accounting to gastronomy. Complementary courses (including orientation courses) are aimed at youth 13 years old or older and seek to either expose youth or update adults on all subjects linked to occupational training in the supply schedule. All courses, either occupational or complementary, are free.

COMFAMA is one of the largest of several private organizations in charge of administering mandatory payroll taxes made by Colombian employers to provide welfare programs and improve well-being. COMFAMA has more than 400,000 affiliated workers and offers a wide range of services in the areas of health, leisure, finance, education and culture, among others. It has multiple service centers across Medellín and different programs directly linked with our objectives.

The training and mentorship program consists of two main components. First, exposure to schooling, culture, and labor markets outside their neighborhood. Our preliminary

data suggests that a large number of at-risk youths seldom leave their neighborhood or are knowledgeable about opportunities beyond a narrow set of possible futures from their neighbors—which includes the gang life of other youth. The first component is intended to adjust adolescent males' aspirations in a general way, making them aware of other opportunities. In particular, partnering with the largest tertiary education provider in the country, we aim to orient adolescents in a set of short 4-8 hours courses on subjects such as robotics, cell phone applications and video games design, mechanics, electronics, and others that improve aspirations and expectations.

Second, mentorship. The final component consists of a series of short sessions with mentors—themselves successful young adults from disadvantaged contexts—who will focus on helping adolescent males note their true opportunities in legal occupations and train them and their families in how to navigate the complex application and enrollment processes. The mentors will be assigned to groups, to help organize and facilitate their participation in the activities. Mentors will also work individually with students and their families on a plan for completing school and pursuing post-secondary studies.

This program is based on SENA's and COMFAMA's existing capabilities and funding. SENA already has a set of short introductory courses aimed at adolescents exposing them to different fields. They will provide these courses under their own budget. In addition to the experimental evaluation, the main departure from normal operations is the recruitment of high-risk youth. SENA has a history of trying to engage high-risk youth and seems well-equipped to do so. This includes, for instance, victims of Colombia's 50-year-long conflict and former combatants, as well as adolescent and young adult participants in previous (non-rigorously evaluated) training and reintegration programs. These courses are the key part of the program described above. On the other hand, COMFAMA has different programs and services such as Expedición Cultura (*Culture Expedition*), which already works with high- and at-risk adolescents to know and experience new city contexts and realities to widen their aspirations and knowledge. We are currently working with COMFAMA on shaping our idea and adapting their existing programs to make their support fundable under their existing budget. Finally, we have relationships with current and former city officials that have been part of Ser Legal Sí Pega (Being legal pays)—a program that very closely resembles the mentorship component of the intervention we describe above. We will use these relationships, as well as the own capabilities of SENA and COMFAMA to provide the mentioned mentorship component.

Treatment arm II: Changing beliefs about the death and arrest probabilities of gang life and the possibility of changing career after gang membership Our

qualitative interviews and pilot data suggest adolescent males underestimate the actual probability of dying or being imprisoned for long periods if they follow a gang life. Similarly, our qualitative observations suggest that adolescents underestimate the probability of changing their career if they join a gang. Interviews with gang leaders suggest that playing up the benefits of joining and playing down the risks is a common strategy for recruitment. The full survey will give us population-based data on these beliefs.

A promising area of intervention, therefore, is changing beliefs about the death and arrest probabilities from this career path, via a light-touch behavioral or informational intervention. In this case, we will use our baseline survey to estimate the life-cycle structural model that we describe above, which will allow us to understand which of these factors—and to what extent—seem to play a more important role in the decision to join a gang—where the outcome will be their own perceived probability of joining a gang.

Once we identify the areas where adolescents hold wrong or biased beliefs, we will work with the school system and city agencies to develop specific informational interventions using professional videos as well as events and personal follow-ups by the school or municipal agency staff. We expect each student in this treatment arm to be engaged for 1–2 months, roughly 3-4 times a week, with videos no longer than 15 minutes—outside of schools. However, we anticipate there might be small changes once we run the life cycle model and develop the video content.

The basis for this idea is our long-term collaboration with the city, and in particular, our work with outreach staff from the Ser Legal Sí Pega program. This is a school-based crime prevention program that seeks to reduce the justifications for crime and violence, the culture of illegality, and the idea of some known gang leaders as heroes. The program involves workshops and talks for students, where they learn about the costs of crime—including talking to ex-convicts, talks for teachers and parents about what it means to be an adolescent and the role they can fulfill to promote legality, and prison visits where students experience the negative consequences of crime by talking and sharing with inmates. We have worked with the Ser Legal Sí Pega team for more than five years to conduct prison interviews and to develop information campaign ideas that could trigger changes in adolescents' beliefs about gang life.

Pilot stage Before implementing the full-scale interventions we describe above, we aim to pilot them through several small-scale interventions. The objective of these is to assess the logistical and operational feasibility of implementing the two treatment arms, the reception that they will have among our target population, and their theoretical foundations. We plan to start running these small-scale interventions during the second quarter of 2023.

5.2 Outcomes

Short-term behavioral measure In the weeks following the interventions, school and municipal agencies will present members of the experimental sample with opportunities to participate in other programs—job applications, work-related training, or information sessions on technical and vocational schools. These will be real opportunities and will not be recognized as associated with the intervention (to reduce response bias, such as experimenter demand). We will assess differences in take-up across treatment arms.

Survey-based outcomes We plan to return to interview the adolescents and their families to survey them on various outcomes, including gang involvement. We are currently piloting approaches to assessing gang involvement directly in ways that minimize reporting error. These include, but are not restricted to, building community or family networks, using administrative data on arrests related to specific types of crimes, or using machine learning techniques to predict gang involvement (when observed) with other observable data such as antisocial behaviors at school or contact with the criminal justice system.

Administrative data outcomes Each student's ID number is also their national ID. Our access to administrative data will allow us to link students to: (i) anti-social behaviors such as school disciplinary problems; (ii) drop-out rates; (iii) involvement with the youth criminal justice system; (iv) enrollment into tertiary education; (v) contact with the criminal justice system as adults; and (vi) economic outcomes such as participation in the labor market and salaries over their lifetime. To ensure access to this third set of outcomes, we have recently signed data-sharing agreements that will allow us to link this data with rich administrative data on long-run outcomes for all adolescents in the school system.

6 Ethics

These data and analysis raise several ethical considerations, which we have spent three years working on before launching the full-scale research. All of the activities described in this proposal have been reviewed and approved by the institutional review boards of the University of Chicago, Princeton University, and Universidad EAFIT in Medellín. We have also secured support from the Ministry of Justice, the Secretary of Education of Medellín, and the Secretary of Security of Medellín. We have formal legal counsel for this project—and our larger set of organized crime studies—to ensure we are complying with all legal requirements in Colombia, especially child consent, and the protection of our data.

Our survey was developed in consultation with anti-recruitment programs in Medellín; child and youth NGO workers; school principals and teachers; survivors; and our legal team. Moreover, these instruments were developed after years of interviews with current members, survivors, parents, practitioners, and experts. We piloted it extensively over approximately two years.

We also have a referral system for youth at risk of self-harm. This referral system has been developed together with our partner the Secretariat of Education to ensure all students who show warning signs of being under mental health distress are referred to the program Entorno Escolar Protector—a program of the Secretariat of Education for preventing and mitigating violence towards students and ensuring their human rights are honored. This program includes one psychologist in each school.

Another consideration is whether interviewing or intervening with children already connected to gangs poses risks, and how to mitigate them. One of the principal reasons for working with a group of mainly 13 and 14-year-olds is precisely to mitigate this risk. Combos very rarely have formal and salaried members this age. At most, children of this age are socially connected to the gang, possibly working part-time as carriers or lookouts, and the majority of children in this role will never become formal members. (After all, gangs seldom grow larger than 35 young men, and there are severe limits to recruitment.) Hence, a child who associates with a gang in this informal fashion, but who pursues a different path (or answers a survey) is highly unlikely to attract attention.

It is true that formal and salaried gang members in late adolescence or early adulthood could face more difficulties, which is one reason we do not work with them. Nonetheless, it is generally safe for formal gang members to associate with programs. The city has run programs for actively involved youth for decades, including with many of our partners. Some of our staff come from working in these programs and have extensive experience with interviewing and working with children that are much more formally engaged with the gang. So long as the participant is not a relatively high-ranking member of the gang, with information that would make them an asset to the police or rival gangs, participating in city or NGO surveys and programs does not usually carry a menace of penalty.

Finally, our field team is recruited, hired, and managed via Innovations for Poverty Action Colombia and is provided (as in all our larger set of organized crime studies) with specific training to understand the gang ecosystem in Medellín, security in the field training (which we have developed after years of work collecting data on organized crime groups), and professional vicarious trauma training.

References

Baird, A. (2018). Becoming the baddest: Masculine trajectories of gang violence in medellín. *Estudios Socio-Jurídicos 20*(2), 9–48.

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pp. 13–68. Springer.

Benoit, J.-P. and J. Dubra (2011). Apparent overconfidence. *Econometrica 79*(5), 1591–1625.

Blattman, C., G. Duncan, B. Lessing, and S. Tobón (2021). Gang rule: Understanding and countering criminal governance. Technical report, National Bureau of Economic Research.

Bourgois, P. et al. (2003). *In search of respect: Selling crack in El Barrio*. Number 10. Cambridge university press.

Brimble, P. Obtaining measures from subjective probabilistic expectations.

Butti, E. *We Are the Nobodies: Adolescents' firs steps into crime in Colombia.* Book manuscript. Under review.

Carvalho, L. S. and R. R. Soares (2016). Living on the edge: Youth entry, career and exit in drug-selling gangs. *Journal of Economic Behavior & Organization 121*, 77–98.

Delavande, A., X. Giné, and D. McKenzie (2011). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of development economics 94*(2), 151–163.

Delavande, A. and B. Zafar (2019). University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints. *Journal of Political Economy 127*(5), 2343–2393.

Dominitz, J. and C. F. Manski (1997). Using expectations data to study subjective income expectations. *Journal of the American statistical Association 92*(439), 855–867.

Enke, B. (2020). What you see is all there is. *CESifo Working Paper*.

Esponda, I. and E. Vespa (2018). Endogenous sample selection: A laboratory study. *Quantitative Economics 9*(1), 183–216.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics 133*(4), 1645–1692.

Falk, A., A. Becker, T. J. Dohmen, D. Huffman, and U. Sunde (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Netspar discussion paper*.

Gilboa, I. and D. Schmeidler (2004). Maxmin expected utility with non-unique prior. In *Uncertainty in economic theory*, pp. 141–151. Routledge.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics 125*(2), 515–548.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics 112*(2), 443–478.

McKenzie, D., J. Gibson, and S. Stillman (2013). A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad? *Journal of Development Economics 102*, 116–127. Migration and Development.

Melnikov, N., C. Schmidt-Padilla, and M. M. Sviatschi (2020). Gangs, labor mobility and development. Technical report, National Bureau of Economic Research.

Patnaik, A., J. Venator, M. Wiswall, and B. Zafar (2020). The role of heterogeneous risk preferences, discount rates, and earnings expectations in college major choice. *Journal of Econometrics*.

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty 5*(4), 297–323.