Title
Three Applied Economic Studies in Collaboration with Government Policymakers

Author
Garz, Seth Aaron Levine

Publication Date
2016

Peer reviewed|Thesis/dissertation

Three Applied Economic Studies in Collaboration with Government Policymakers

by

Seth Aaron Levine Garz

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy
in Agricultural and Resource Economics

in the Graduate Division of the
University of California, Berkeley

Committee in charge:

Professor Alain de Janvry, Chair
Professor Michael L. Anderson
Professor Paul J. Gertler

Summer 2016

Abstract

Three Applied Economic Studies in
Collaboration with Government Policymakers

by

Seth Aaron Levine Garz

Doctor of Philosophy in Agricultural and Resource Economics
University of California, Berkeley

This manuscript features three independent applications of contemporary techniques in applied economics conducted in collaboration with three different governments. In the first essay, I present findings from an experimental intervention conducted in collaboration with the government of the Dominican Republic in which the government varies whether front-line volunteer community workers are recruited through a public advertisement or through local staff referrals. I find that governments likely face tradeoffs in selecting optimal recruiting mechanisms as publicly recruited candidates demonstrate superior observable characteristics to referred candidates across a wide variety of indicators except for the key indicator of cognitive skills. I also find public candidates are more likely to accept job offers and attend trainings conditional on being hired. To my knowledge, this is the first study to rigorously test the impacts of an open public recruitment process versus a private targeted recruitment process.

In the second essay, my co-author and I evaluate impacts of a World Bank funded road improvement and employment generation intervention in Nicaragua. We employ a difference-in-difference research design, matching proprietary road building data from the government of Nicaragua's Ministry of Transportation with three rounds of a publicly available household survey. We find strong evidence that the World Bank's Fourth Roads Rehabilitation and Maintenance Project fulfilled its primary goal of improving road infrastructure and suggestive evidence of select economic and social impacts. Notably, we do not observe impacts on the likelihood of employment or incidence of poverty.

In the third essay, I collaborate with the City of San Francisco's Office of the Treasurer to investigate mechanisms that potentially mediate household decisions to save for distant future expenditures with particular relevance for education-oriented savings. Employing a randomized experiment in the natural setting of the Kindergarten-to-College school district-wide college savings program, I vary the messages of postcard savings reminders with information about either the availability of college financial aid or the increasing cost of tuition. My results suggest that savings reminders may overcome inattention to lumpy future expenditures, but only among those who intended to save. I find no evidence that manipulating the salience of college cost affects savings behavior.

Dedication

To my children and grandchildren,

that absolute poverty may not still be around to demand your attention when it is your turn.

Contents

Acknowledgements

First and foremost, I would like to acknowledge the members of my dissertation committee. Thanks to Alain de Janvry as my committee Chair and academic advisor for your encouragement, optimism, and wisdom, which was always available at a moment's notice, and for introducing me to the world of development economics as my teacher. Thanks to Paul Gertler for your mentorship: for showing me how to do real research, engage in authentic open collaboration, and design academic work to answer real world problems. Thanks to Michael Anderson for your lucid instruction and ongoing counsel on causal inference and research design; when attending your Applied Econometrics class I began to see the world with new eyes and became determined to pursue the Ph.D. in ARE.

Thanks to my classmates Sylvan Herskowitz and Kenneth Lee for the collective angsting, casual intellectualizing, and irreverent humor throughout this journey. Thanks to my officemates Yang Xie, Daniel Tregeagle, and Marieke Kleemans for your capacity to overlook molding coffee cups, to indulge frantic technical questions, and to take seriously my intellectual meandering.

Thanks to my teachers and unofficial advisors Elisabeth Sadoulet, Jeremy Magruder, and Fred Finan. Special thanks to Peter Berck for reminding me that with an infant and a first year Ph.D. course load, I should have been satisfied just to still be alive, and for always being around for guidance when doubt crept in. Thanks to David Roland-Holst for employing me as a Reader in Chinese Economy and an instructor at the MDP boot camp for so many years. Thanks to Carmen Karahalios for always giving me the administrative inside scoop. Thanks to Señora Marjorie Spaeth for sending me to Mexico in 9th grade and catalyzing my interest in the foreign.

Thanks to my collaborators on the project that is the focus of Chapter One, including Hector Medina, Ezequiel Volquez, José Antonio Pellerano, Yaneth Basora, Ricardo Cuesta, Elianny Medina, Sandro Parodi, Maricha Freidman, as well as the ProSoli, Gabinete, and CEGA staff. Special thanks to Justo Mirabal. Thanks to my Chapter Two co-author Elizaveta Perova and collaborators among the staff of MTI in Nicaragua. Thanks to my collaborators, advisors, and commenters on the project that is the focus of Chapter Three, including SF Treasurer José Ciscneros, staff of the SF Office of Financial Empowerment, Anee Brar, Minah Jung, Leif Nelson, Sofia Villas-Boas, Aluma Dumbo, and Lei Cheng.

Thanks to the founders and maintainers of the East Bay Regional Parks, the fire trails of which often rejuvenated me during steep afternoon runs, and to Alexis Madrigal for joining me up those hills. Thanks to the architects of the Morill Act, the builders of UC Berkeley, and the California taxpayers that continue to support the University of California whose vision and work has afforded me the privilege of studying and teaching at a public university that is a model of diversity and intellectual effervescence to which the world can aspire.

Thanks to Pat for being real. Thanks to Joshy for your intentionality.

Thanks to my great-grandparents, my grandfathers, Uncle Doug, Grandma Rae, and Grandma Tillie for leading us on our passage from the shtetl to the ivory tower. Thanks to Mom and Dad for every opportunity in the world. Thanks to my sister Jess for sharing how I understand. Thanks to my Sherwood Family for the love and support.

Thanks to Olea and Eshel for true joy.

Thanks to Julie.

Optimal Recruitment Design for Frontline Public Servants: Public Advertisements vs. Private Referrals

Seth Garz

August 9, 2016

Abstract

In designing an optimal recruitment process, managers must decide to what extent they should invest in open public searches or privately targeted referral-based searches. Numerous studies document the widespread use of referrals in the labor market, but existing evidence comparing the performance of referred to non-referred workers suggests contradictory results. Furthermore, these studies generally rely on associations in data that are likely subject to important selection biases and often fail to define the counterfactual to referral-based recruiting in a way that meaningfully reflects the recruitment decisions routinely faced by managers. In this study, I collaborate directly with the government of the Dominican Republic to experimentally vary whether front-line volunteer community workers are recruited through a public advertisement or through local staff referrals. I find that the government faces tradeoffs in selecting among recruitment methods with publicly recruited candidates revealing superior observable characteristics across a wide variety of important indicators, but referred candidates demonstrating unambiguously superior cognitive skills. I also find publicly recruited candidates are more likely to accept job offers and attend trainings conditional on being hired. To my knowledge this is the first study to rigorously test the impacts of an open public search recruitment process versus a private targeted search recruitment process. Worker selection is particularly important in the public sector given government's limited means of motivating workers with traditional incentives and the large scale of public front-line workers in security, education, health, and social protection. My findings suggest a number of future research topics that could collectively have important policy relevance for improving government recruitment efforts.

1 Introduction

A large body of literature documents that workers both in developed and developing countries frequently find jobs through referrals. A handful of studies also demonstrate that firms frequently encourage referrals as a recruitment method. Researchers have empirically validated specific mechanisms by which firms can leverage the network attachments between referring employees and referral candidates to improve worker screening, monitoring, matching, satisfaction, and effectiveness generally.[1] However, to conclude that firms benefit more from using referrals than from using other recruitment mechanisms would misinterpret the existing evidence. Merely observing the frequency of referral-based job search and validating plausible benefits from a referral-focused recruitment process is insufficient to discount the potential benefits to firms of employing alternative recruitment mechanisms, such as open public searches, about which there is little evidence. This study seeks to contribute to the literature on the explanations for and consequences of different job recruitment designs by providing what is to my understanding the first rigorous experimental evidence comparing an open public search recruitment process to a privately targeted referral-based recruitment process.

Heath (2013) argues that the prevalence of referrals among hired employees is *prima facie* evidence for the inferiority of open public recruitment processes. However, this need not be true of the public sector even if we take revealed preference seriously to conclude that firms disciplined by the profit motive in competitive sectors of the economy conscientiously and rationally rely heavily on referral-based recruitment. While publicly advertising competitive processes for senior bureaucrats may be a regulatory requirement among many governments, the hiring of low level government staff and community volunteers that make up the front-line workers at the base of the public sector pyramid in developing and developed countries alike are almost certainly more likely to depend on hiring through references. In our study context in the Dominican Republic (DR), for example, roughly 12,000 volunteer community workers (CWs) administer home visits to welfare beneficiaries with the recruitment of the CWs entirely left to the discretion of local supervisors.

Front-line worker hiring choices in the public sector may be particularly distorted given the low cost of the workers and the difficulty of fully internalizing the productivity losses from these workers to an organization that is not subject to market discipline. Ineffective recruitment may be particularly damaging to the effectiveness of public agencies given the difficulties they face motivating workers to perform and their limited ability to use strong incentives.[2] The role of these workers in national and civil defense, education, health, and administering social protection services to the most vulnerable citizens may render sub-optimal recruitment strategies and hiring decisions particularly costly in terms of social

[1] Afridi et al. (2015) provides a review of recent theoretical and empirical work on social networks and labor productivity. As an example of *managers* reporting the frequency of referrals, as opposed to *workers* themselves, Dhillon et al. (2013) uses responses of managers from the 2006 World Bank Enterprise Survey from India to show workplace referrals account for between 40-65% of hires among a diverse of array of industries.

[2] See Chaudhury et al. (2006) for a review of absenteeism among public education and health workers and Dixit (2002) for review of theory of incentives in the public sector.

welfare. Given that the public sector constitutes the largest source of employment in virtually every national economy, it stands to reason that modest improvements in recruitment processes of public agencies could have important social and economic implications. Furthermore, investigating ways to improve the selection of front-line social protection workers, in particular, is consistent with broad interest among governments and multilateral funders in developing cost-effective regimes for layering social intermediation services on top of existing cash transfers programs.[3] Thus, rigorous evidence comparing the characteristics and performance of candidates recruited through referral versus alternative means may be particularly valuable for these front-line employment settings.

In this study, I collaborate with the agency responsible for administering the DR conditional cash transfer (CCT) program, *Progresando con Solidaridad* (ProSoli), and for managing the associated CW volunteers. We experimentally vary for which neighborhood CWs are recruited through an open public advertisement intervention versus a private targeted referral process from local government staff, which was designed to closely resemble the status quo recruitment process. By design, the interventions resulted in areas for which there are overlapping samples of candidates from both public advertisement and private referral recruitment processes. Comparison of these overlapping samples will constitute my primary means of identifying differences in the candidates recruited through either intervention to answer the following key research questions:

1. Are candidates privately targeted by government staff for low skilled public service jobs systematically different from those that would be attracted through a public recruitment process?

2. Are publicly or privately recruited candidates superior according to observable criteria that would *a priori* be used to screen candidates by a hiring employer?

3. Are characteristics of publicly recruited candidates unambiguously superior across all criteria or would a hiring organization expect to have to weigh tradeoffs in the character of the candidates recruited through different mechanisms?

4. Does public recruitment lead to higher rates of job acceptance and persistence?

5. Among privately referred candidates, are those that are higher ranked by referrers better or worse candidates according to observable characteristics?

Results from comparisons of candidates recruited through open public search and private targeted referrals suggest systematically superior characteristics for public candidates across a wide variety of indicators observable to the researcher and hiring agency, including characteristics pertaining to demographics, employment, economic status, civic and political involvement, and personality. Referred candidates do, however, out-perform public candidates on important cognitive characteristics, begging the question of whether referred candidates may have other unobservable characteristics that would lead to better future job performance. Tension between the superiority of referrals on the cognitive dimension and

[3] See Camacho et al. (2014) for a recent review of the role of social intermediation services in CCTs.

the superiority of public candidates on virtually all other observable characteristics suggests governments may face important tradeoffs in deciding among alternative recruitment mechanisms. Nevertheless, results from a small scale sub-experiment within the broader study suggest public candidates are also more likely to accept jobs and more likely to show up at subsequent trainings, casting doubt on the idea that privately targeted referrals constitute a better candidate recruitment and selection process. Additional results from comparison of referred candidates ranked higher and lower by the referrers suggests that there is little variation among privately targeted referrals.

In terms of operational efficiency and cost, results from the study suggest that the public search process was modestly successful, recruiting candidates to roughly half of the targeted areas intended at cost equivalent to only one to two months wages per position. By comparison, the private referral process only succeeded in covering 19% more of the same targeted areas. This modest operational success coupled with the apparent superiority of candidates recruited through the public search process across most observable characteristics begs the question of why these front-line public sector jobs rely on private referrals and offers a promising policy alternative to improving the quality and persistence of front-line community workers. Interpreted more broadly, these findings question the conventional wisdom that the prevalence of referrals in job markets coupled with empirical evidence explaining various plausible benefits of referral processes demonstrate the superiority of private targeted searches in comparison to well-defined and sufficiently resourced open public searches. Nevertheless, the results demonstrate that employers face tradeoffs and further investigation to better establish results consistent with those presented here is needed before applying these findings to more general policy recommendations. In future research, I intend to track the performance of the different types of candidates and their impacts on CCT beneficiaries to complement my current findings.

This study seeks to overcome important limitations in existing studies that attempt to compare the characteristics and performance of referred versus non-referred workers. First and foremost, numerous studies define "being referred" in binary terms through self-reports of job candidates, identifying who is referred without rigorously defining the counterfactual to being referred.[4] However, it is the counterfactual to referrals that is precisely what is of interest to public and private organizations when making the key management decision of how to optimally design an employee recruitment process. "Referral" in the sense that the term is currently used in many studies does not meaningfully correspond to a specific recruitment option that an organization may choose to pursue.

Whether or not to use or ignore peer referrals does not effectively capture the management decision an organization faces. More relevant to most organizations is the choice between whether to pursue some form of *open public search* process or to rely on the *private targeted search* processes of its employees. More specific recruitment options include: i) investing in a robust public search process, ii) relying on employees natural proclivities to pursue their own private search techniques, which often leverage their own social networks, iii) investing in incentivizing employees to pursue a private search process that is more compatible with the firm's recruitment priorities, or iv) letting nature run its course and seeing which candidates

[4] See Burks et al. (2015) and Munshi and Rosenzweig (2006) for examples.

walk in the front door, so to speak. The precise counterfactual recruitment process to pure referrals can certainly vary and may drastically alter the comparison of referred versus non-referred candidates. In the extreme, just accepting applications from people that happen to walk in the front door would be a form of untargeted public search, but is unlikely to yield many qualified candidates.

In Pallais and Sands (2015), the existing study that is closest in design to my study, the counterfactuals to referred workers are workers randomly invited to apply for piecemeal work from the pool of qualified temporary workers that had previous work experience with the hiring firm. This random recruiting approach is not completely unreasonable in the study's stylized setting of a private sector offshore business processing center that maintains a network of pre-screened temporary workers, but likely lacks external validity for more general recruitment settings and public sector settings, in particular, where there is rarely an existing pool of pre-screened workers from which to randomly select. Furthermore, the superior performance of referred workers in Pallais and Sands (2015) is not entirely unexpected given that there is no real open public pool from which to refer or randomly select new recruits, implying that non-referred workers may be systematically avoided by referrers by construct. Like in many other studies, referrals in Pallais and Sands (2015) come from peers and not managers who would likely apply a different private search screening process that may be more consistent with local hiring dynamics in low skilled job settings where mass candidate screening can be costly. Notably, the stylized design features of the Pallais and Sands (2015) study were intentional and enabled the authors' to focus on a variety of other compelling questions related to monitoring and collaboration.

In contrast to a highly stylized alternative to referral-based recruiting, the public search method in my study setting is the intervention the government deemed the most practical and obvious "counterfactual" innovation to the status quo approach of private search by local government staff. Specifically, the research team posted posters and handed out fliers in the exact neighborhoods from which the privately referred candidates were also recruited. This was possible because our setting requires candidates to live in close proximity to the welfare beneficiaries they are being selected to serve.

A handful of existing studies suggest that referred and non-referred candidates with similar observable characteristics at the time of recruitment have quite different levels of turnover and productivity, though the direction of the results in favor or against referrals is contradictory. Fafchamps and Moradi (2015) use associations in historical data in a low income country context on recruitment and desertion in Ghana's Colonial Army to suggest referrals reduce the quality of recruits with higher desertion and dismissal rates among referred soldiers. Importantly, the authors find that, compared to non-referrals, referred candidates actually had superior observed characteristics, specifically height and chest circumference, at the time of recruitment. In contrast, using data from nine large firms across three industries mostly in the United States, Burks et al. (2015) find that referred workers are associated with higher rates of job offer acceptance, similar if not higher productivity, and lower rates of turnover despite having similar characteristics at the time of hire. These studies suggest caution in using results from my study demonstrating systematically different observable characteristics among public versus referred candidates to predict future job persistence and performance. Nevertheless, these existing studies with their contradictory conclusions suffer from a second key limitation in the existing literature - that frequent positive associations between referrals and outcomes of interest to employers are not necessarily causal and may reflect sorting, survivor bias, and selection on characteristics not observable to the econometrician. The fact that 48% of the neighborhoods targeted by the public search intervention and 23% of neighborhoods targeted by the private targeted referral intervention in this study failed to yield any job applicants reflects the importance of considering selection and controlling for quite local geographic fixed effects in interpreting associations in the data.

That public searches can be costly may be the simplest candidate explanation for why organizations frequently prefer private search by employees over investing in robust public search recruitment processes. Relying on referrals through private employee-led recruitment efforts, however, is not without both private and public costs, as well. The field experiment by Beaman and Magruder (2012) reveals that imperfectly aligned incentives in the employee-employer principal-agent dynamic may yield sub-optimal referral behavior by employees without firm's investing in proper incentives. This is consistent with the common practice of firms offering financial inducements for referrals often conditional on referrals passing a screening process and persisting in a job.[5] In addition to creating cost for firms, various studies have suggested that referrals can be socially costly, exacerbating economic inequality and racial and gender bias in hiring (Calvo-Armengol and Jackson, 2004, 2007; Beaman et al., 2015). These social externalities of the recruitment process may even warrant policy solutions that mandate more open public searches, lending further importance to evidence presented in this study. Notably, the private referral process in this study only led to 19% greater coverage of targeted beneficiary clusters than the public search process, suggesting the public intervention could succeed in effectively recruiting candidates at only modest marginal cost to the government. This basic, but straightforward result may be the most actionable finding for the local policy context in which the government is regularly in need of new CW candidates.

This study shares methodological similarities with a variety of recent field studies from the developing world that have manipulated different aspects of the recruitment process to better understand the relationship between employee selection and performance. Ashraf et al. (2014) find that variations in motivation matter for the performance of low level government employees in the health sector in Zambia. Specifically, the authors find that local health workers recruited by advertisements that make career incentives salient perform better than those recruited through advertisements that make pro-social civic service incentives salient. This may provide cautionary evidence against positively interpreting my results that show publicly recruited candidates are significantly less motivated to volunteer for career or extrinsic goals. Dal Bo et al. (2013) find that contrary to the hypothesis that there is tension between extrinsic motivation and work in public service, offering high wages for public sector community development agent jobs in Mexico actually attracts applicants with superior personality and cognitive characteristics, and generally similar pro-social motivation. This suggests that the higher opportunity costs faced by public candidates in my study, in the form of more frequent employment and higher wages, may not inhibit public candidates'

[5]Footnote 1 in (Burks et al., 2015) suggests: "A leading online job site estimates according to their internal data that 69% of firms have a formal employee referral program."

willingness to persist in their volunteer jobs. In the Dal Bo et al. (2013) study, however, candidates do differ substantially from the candidates in my study in that they are of much higher cognitive ability, better educated, and much better compensated. Finally, evidence on front-line doctors in poor community public health clinics in Punjab, India from Callen et al. (2015) suggests positive personality traits and pro-social motivation are positively associated with work attendance, negatively associated with collusion with inspectors, and greatly amplify the effect of a randomized monitoring intervention. This evidence is promising for the results from my study, which suggest public candidates have greater pro-social behavior as indicated by their involvement in civic service and have superior personality traits, according to a few indicators.

2 Background

2.1 The Role and Selection of Volunteer Community Workers

The widespread use of volunteer community workers in the DR CCT program may provide a compelling model for other low and middle income countries looking to scale the delivery of social intermediation services to their poor populations at minimal cost. Prior to this study, however, the profile and recruitment of these volunteer CWs has not been examined in detail.

Approximately 12,000 volunteer CWs, called *Enlaces Familiares*, participate in the implementation of ProSoli, the DR CCT scheme. As new eligible households are enrolled in ProSoli, new CWs are recruited from the same local communities as the beneficiaries. Consistent with descriptive statistics from the study sample in Table 1, CWs are often ProSoli beneficiaries themselves. Importantly, there is no formal screening process for new CWs with substantial leeway given to local ProSoli staff in their recruitment. Formally, the only requirements are that CWs are of adult age and know how to read and write, though literacy is never tested and in many cases CWs are only semi-literate.

After being selected as new volunteers, CWs are assigned to a cluster of 30-60 CCT beneficiary households (*nucleo*) and are responsible for visiting each of their assigned households at least once monthly. CWs are briefly trained for one to two days in a curriculum of 12 structured lessons related to citizenship, health, wellness, employment, education, etc., which are intended to be delivered to their beneficiary households across their first 12 months of enrollment in ProSoli. Beyond delivering the content of the lessons, CWs are intended to be a general resource to beneficiary families, reminding beneficiaries of the responsibilities required to maintain their benefits, directing them to local health, educational, and vocational training resources, etc.. A household member must sign a form during each visit to validate that the home visits take place. The quantity and frequency of these visits are reported to ProSoli's central office, which independently audits the reports.

Importantly, CWs are not intended to be responsible for monitoring and policing compliance with CCT conditionalities. Rather, it is their supervisors (*Supervisores de Enlaces*) that are responsible for checking that pregnant women and children 0-5 years old have attended preventive health checkups with compliance for primary and secondary education requirements automatically checked in centralized ProSoli and Ministry of Education databases. Supervisors monitor compliance at Family Schools (*Escuelas Familiares*), which are required monthly convocations of beneficiaries that often include health related presentations from guest speakers. Whether or not CWs are able to provide impartial support to beneficiaries or, in practice, are delegated responsibility from the Supervisors to monitor their households' compliance remains unclear and may vary. For example, conversations with CWs suggests that CWs are even sometimes responsible for running the Family Schools.

Though CWs are volunteers in the sense that they are not paid a true wage and are only expected to work part-time, they may earn up to 1000 pesos a month (US$22) as a travel stipend, depending on the fraction of assigned households they visit monthly. Descriptive statistics in Table 1 show that this is equivalent to 8.5% of the mean monthly household income reported by candidates in the raw study sample. This is also approximately equivalent to 121% of the monthly cash transfer delivered to the 53.3% of the sample that are ProSoli beneficiaries, though anecdotally CWs in the field report payments are sometimes delayed and incomplete. In comparison, the Supervisors are expected to work full-time and receive 7000 pesos a month (US$156) plus 1000 pesos as a travel stipend depending on the fraction of households visited by the CWs under their supervision.

CWs are embedded in the same communities as the beneficiaries and are deeply familiar with ProSoli often with personal experience as beneficiaries themselves. ProSoli leaders often refer to CWs in glowing terms, reflecting their perceived dedication to and leadership in their communities. Other citizens are more skeptical of the impacts of the CWs and speculate that they volunteer for the sake of developing connections to the government that will lead to future employment for themselves or other family members. Regional managers often lament the difficulty they face in recruiting new CWs and anecdotally explain that they use private targeted searches for recruiting new CWs because it is so difficult to find new volunteers otherwise.

3 Study Design

This study was conducted in the context of a broader randomized controlled trial (RCT) designed to evaluate the impacts of social intermediation services provided by CWs on CCT recipients in the DR. CCT recipient households in two of the treatment arms of the broader RCT were selected to receive monthly home visits from volunteer CWs. The government exploited the need to recruit new volunteer CWs to test how a practical alternative to the status quo recruitment mechanisms would affect the quality and quantity of candidates attracted to the job. In the first treatment arm - the *Public Advertisement* arm - posters and leaflets advertising the opportunity were distributed throughout targeted neighborhoods by research staff and interested candidates called into a dedicated call center toll-free. In the second treatment arm - the *Private Referral* arm - volunteer community worker candidates were referred by local supervisors in a manner designed to mimic the status quo recruitment process.

3.1 Public Advertisement Recruitment Intervention

The public advertisement intervention to recruit new CWs was designed as an experimental yet practical alternative to the status quo referral system. In the new experimental intervention, research staff distributed posters and fliers with information about the volunteer opportunity throughout targeted neighborhoods. Figure 1 displays the graphic used on the flier, which includes the headline "Would you like to be a community leader?" Interested citizens could register their name and contact information with a dedicated call center setup for this study through the toll-free number listed on the advertisements. The intervention targeted 108 household clusters randomly assigned to the Public Advertisement arm of the broader RCT. 400 individuals contacted the call center and registered their personal information of which 132 were successfully contacted by research staff and subsequently participated in the screening survey. Interested candidates were reminded of program age and literacy requirements twice, first by call center operators and second during callbacks by research staff to schedule surveys, before being surveyed by research staff in person at survey sites in the local neighborhoods near their homes. The 1000 peso travel stipend for volunteers was not mentioned until the brief orientation prior to the screening survey.

This study employed a rather large and costly distribution effort, professionally printing ∼4500 colored posters, ∼9000 colored handouts, and ∼1000 black and white handouts, and hiring field staff to distribute them across locations in seven provinces over the course of 2-3 weeks. Nevertheless, it is notable that this process did not employee resources or technology that are otherwise out of reach for local ProSoli supervisors or regional managers. Local supervisors are well known throughout the local communities in which they work and readily accessible by mobile phone. As volunteer hiring opportunities arise, they could easily distribute handmade if not printed posters and fliers advertising volunteer opportunities and accept calls on their own mobile phones without incurring fees to communicate with interested applicants who they could screen according to their own preferred criteria. From the perspective of ProSoli regional managers, this public advertising approach could easily be implemented at scale, employing the legions of low-cost assistants and even the CWs themselves to distribute the materials. Furthermore, public advertisement is not a particularly creative intervention. For example, supervisors advertise changes in scheduling of local Family School meetings with hand written notes at local corner stores frequented by program beneficiaries and posters with public service announcements are common. In this context, the public advertisement intervention should not be viewed as employing an entirely novel or elaborate search technology, but rather constitutes an alternative recruitment regime that both local supervisors and regional program managers could, but for some reason choose not to, pursue.

3.2 Private Referral Intervention

The referral intervention to recruit the new CWs was designed to mimic the status quo process by which local ProSoli staff informally select new volunteers. ProSoli staff in the central office distributed the lists of new beneficiaries organized in new clusters to local staff in November 2015 with instructions that the local staff should recommend and rank four potential candidates for each cluster.[6] The names and contact information for the referred candidates was to be entered manually on a basic form, which the research staff then collected and digitized. Figure 2 displays the referral form that was distributed to regional offices. Volunteer referrals were requested for 107 clusters of beneficiary households in the Private Referral arm, 108 clusters in the Public Advertisement arm, and 107 clusters in a third arm for which student candidates from universities where also being recruited to work as salaried employees. Each of these household clusters had been randomly assigned to their respective experimental arms from 537 original clusters in the sample frame with balance verified such that the neighborhood or household characteristics of a given cluster should be orthogonal to the recruitment intervention applied to the given treatment arm.

After reviewing the initial referrals, it became apparent that the directions had been misinterpreted and many recommended candidates were already working as CWs at which time ProSoli redistributed the instructions in January 2016, re-emphasizing that candidates could not be existing CWs.[7] There was much higher compliance with the instructions for the second round. Nevertheless, many clusters had less than four recommended candidates and in some cases the candidates were not ranked.

Following the second round of recommendations, survey sites were selected as close to the location of the intended household clusters as possible and recommended candidates were invited to participate in the survey. 564 candidates were *formally* referred, including 6 candidates that had also called into the public advertisement intervention call center. These 6 candidates will be considered "referred" for the remainder of the paper as they do not reflect a marginal new population of candidates that would exclusively have been recruited through the public search. Research staff were able to reach many of the candidates in a phone callback to invite them to the survey directly and otherwise asked the recommender to share date and location information with the recommended candidate. During this callback, candidates were reminded of the age and literacy requirements and informed that the survey would be used for screening candidates for strictly volunteer positions without any mention of even the travel stipend provided to CWs. Table 2 reveals that 526 formally referred candidates attended the screening survey. In clusters for which recommendation forms were not received, research staff directly contacted the local supervisors and encouraged them to send referred candidates directly to the survey sites. Table 2 shows that 212 *informally* referred candidates recruited in this last-minute manner participated in the screening survey. These last-minute referrals were conducted for the sake of advancing with the study and ensuring sufficient coverage of clusters. This imperfect implementation reflects the real world research setting, rendering the results more externally valid for the actual government bureaucracies that must recruit personnel and implement programs at scale in diverse regions with varying levels of local capacity. Comparing Columns (1) and (2) in Table 3, validates that the average characteristics of the full sample of referred candidates, including the last-minute referrals, are not substantially different from the sub-sample of those that were formally recommended with a form. Nevertheless, in my principal analysis I will limit comparisons with public advertisement candidates to those referred candidates that were formally referred

[6] The lists of beneficiaries actually included both new and old beneficiaries that were combined to create clusters of beneficiaries of sufficient size to warrant hiring a new CW.

[7] Figure 3 provides a general project timeline.

with a form.[8]

The referral intervention was designed to mimic the status quo recruitment process as closely as possible. Nevertheless, some differences are worth considering for the sake of interpreting the subsequent analysis. For the study intervention, it was the central ProSoli office initiating recruitment of new CWs. Local staff were encouraged to recommend and rank more than one candidate and candidates were subject to a screening survey. Under normal circumstances, recruitment of CWs is all conducted locally with no formal screening mechanism. As a result, local staff may have felt they were being monitored and under greater pressure to refer better or at least different candidates than normal, e.g. candidates to whom they were not related, etc.. A variety of evidence suggests this was not, in fact, the case. The counts in the first row of Table 4 reflect that in 84.9% of the 458 cases for which the referrer complied with identifying their own identity and title, those identified as referrers were local CW Supervisors who received the formal referral directions from their local offices and were not necessarily aware the directions came from the central ProSoli office. Existing CWs and Field Supervisors were the second and third source of referrers, but formally only referred 28 and 27 candidates respectively. Thus, referrals can be thought of as broadly reflecting the private search process of CW Supervisors. The fact that many intended referrers did not comply with directions, either by recommending fewer than four candidates, failing to rank the candidates, or failing to fill out the recommendation forms, further suggests they did not take seriously any implicit threat of monitoring. Finally, the profile of the recommended candidates captured in summary statistics in Table 3 confirms informal characterization from anecdotes and field visits that CWs are largely beneficiaries themselves, mostly female homemakers, and often otherwise unemployed.

4 Data

4.1 Screening Survey

Surveys of CW candidates were conducted January 26-March 9, 2016 by trained enumerators using tablet computers. The average survey took 80 minutes to complete, including time to validate personal identification and consent forms. The candidates were convened in sites near their homes, such as schools, vocational training facilities, and local ProSoli offices. Surveys were often conducted simultaneously in the same room, though effort was made to disperse respondent-enumerator pairs to reduce background noise and maintain privacy.

Despite the fact that respondents were aware that they were applying for a job and many of the questions required self-reporting personal information, such as employment, income, etc., there are a variety of reasons to have confidence that the data are not systematically biased and more specifically that measurement error in responses by candidates recruited by referral versus public advertisement is not systematically different. The survey consent,

[8]An additional benefit of limiting principal analysis to formally referred candidates is the ability to assign candidates to the unique cluster indicated on the form. In comparison, those informally referred candidates that showed up to the screening interviews were not explicitly associated with any cluster and, given the geographic proximity of many of the clusters, were sometimes eligible to work in a number of different clusters.

which was read out loud by enumerators, emphasizes the importance of responding to the questions honestly, that participation is voluntary, that respondents can decline to answer questions or halt their participation in the survey at any time, and that their data will be kept confidential. Furthermore, other than literacy and age requirements, the respondents were not given any information about CW characteristics or selection criteria required or desired by ProSoli. For many questions, it is hard to imagine in which direction candidates would consider biasing their answers to game the selection process. For example, it seems equally plausible that having a source of regular income could help and hurt their chances of being hired.

Additionally, candidates recruited through referral and public advertisement were included in the same survey sessions and enumerators were not informed of the differences in interventions or which candidates were recruited through which channels. To avoid priming respondents, questions about how candidates became aware of the volunteer opportunity and by whom they may have been referred are only introduced in the latter half of the survey after the key personal information is already revealed. Finally, enumerators were assigned to candidates as they became available for the sake of efficiency and were not systematically assigned to candidates by gender, age, or any other characteristics.

The fact that candidates were aware that the survey could be used to screen their eligibility for the CW position suggests the information provided more closely resembles what an organization would be able to observe in a normal hiring process run, for example, by a human resources department. Of course, in the case of referrals, the source of the referral may be privy to unobservable information that is not revealed or is misrepresented to the interviewer. Whereas the referrer would likely use any private information about referred candidates to inform their own less formal CW selection process, general staff of the government would not have such insider information to use for screening public advertisement candidates.

4.2 Sample Description

CW candidates were recruited to work in neighborhoods that corresponded to 322 clusters of beneficiary households. These clusters were randomly assigned to one of three treatment arms in the broader RCT from an overall sample of 537 clusters. My main comparison of candidates from public versus private referral searches will focus exclusively within clusters in the treatment arm intended to be targeted for the public advertisement intervention. Nevertheless, it is worth comparing the full sample of CW candidates to the broader population to get a sense of how candidates may differ from their neighbors.

A handful of questions in the screening survey exactly or closely resemble questions asked in the 2010 National Census and National System for Beneficiary Selection (SIUBEN), a poverty census that is used to select ProSoli beneficiaries.[9] Table 5 draws on all three data sources to compare the sample of candidates, the sample of ProSoli beneficiaries these candidates are being recruited to serve, and the general population living in the same neighborhoods as the ProSoli beneficiary sample. The comparison should be interpreted with

[9] The *Sistema Nacional de Seleccion de Beneficiarios* (SIUBEN) is the data system used to screen eligibility for many of the DR's social protection services.

caution as the surveys were fielded at different times by different organizations. Comparing Column (3) to Column (2) suggests the CW candidates were more likely women and better educated that the ProSoli candidates they are recruited to serve. Nevertheless, they live in larger households and are less likely to identify their primary occupational status as being permanently or occasionally employed, mostly because they disproportionately identify as homemakers and students. Not surprisingly, the CW candidates appear to be less poor than the ProSoli beneficiaries with more household assets, lower likelihood of renting their home, and higher likelihood of health insurance coverage. Comparison of Columns (1) and (3) reveals that on average CW candidates are demographically different from their neighbors in the general population; although, it is not clear whether they are economically advantaged as their index of durable assets is similar five years after the census data was collected and they are less likely to rent their homes.

The study sample is not nationally representative, so interpreting the external validity of the results will require some context. Differences in social networks and economic activity in urban versus rural sectors are *a priori* the most likely source of any geographic heterogeneity one might expect in recruitment outcomes. For example, the density of population and greater abundance of piecemeal employment in urban neighborhoods may render eligible job candidates less visible or influence their demand for new part-time employment. Columns (1) and (2) in Table 5 reveal that the geographic area targeted in this study is >90% urban, suggesting results from the study are more indicative of recruitment dynamics in urban settings.[10] Despite the overwhelming urban concentration, Table 6 reveals that there is geographic variation by province across the sample of candidates recruited both through public advertisement and referral interventions. Though not designed to be representative, the sample roughly reflects the large concentration of the population in the country's two largest cities, with 45.0% and 19.0% of the CW candidates from the provinces of Santo Domingo and Santiago respectively, the vast majority of which are from the major urban agglomerations of Santo Domingo and Santiago cities.[11]

4.3 Identification

A candidate explanation for why public advertisements are not normally used for volunteer recruiting is that they would be ineffective. In fact, ProSoli leaders were initially skeptical that such advertisements would work. For this reason, in addition to distributing the public advertisements, the study also solicited ProSoli staff recommendations for candidates in the 108 clusters in the Public Advertisement arm. Thus, clusters in the Public Advertisement arm were intended to have overlapping sub-samples of candidates recruited through

[10] The urban concentration of the sample reflects the fact that the sample of ProSoli beneficiary clusters for the broader RCT, which frames this study, is neither a representative sample of the general population or the beneficiary population. Rather, the sample frame for the broader RCT was limited to those households which were eligible to be ProSoli beneficiaries according to their responses to the SIUBEN poverty census, but had not yet been enrolled as beneficiaries either because budget limitations that have limited program rollout or because they were only recently surveyed.

[11] The sample from Santo Domingo province include the capital National District and three candidates from just over the provincial border in the neighboring province of San Cristobal. "Other" refers to 14 candidates whose addresses were mistakenly not collected.

advertisement and through referrals by design. This detail of the recruitment's overlapping targeting will form the crux of my identification strategy.

The presence of overlapping sub-samples of candidates recruited through different mechanisms for the same positions in the same locations allows for valid estimation of the causal impact of employing public versus private search on the characteristics of candidates recruited into the hiring pool. Comparison of overlapping samples in the same clusters is, in some sense, the cleanest counterfactual comparison one can imagine and even better than randomization. The clarity of the estimation is particularly useful in my medium sample size context as I do not have to rely on balance from randomization to maintain valid causal inference as I would in estimating intention to treat (ITT) impacts when comparing candidates recruited to Public Advertisement and Referral arms. In this case, estimating an ITT would be dubious because, as reflected in Figure 4 and Figure 5, the public advertisements only succeeded in recruiting candidates to cover roughly half of the Public Advertisement arm clusters (56 of the 108 intended clusters) and references only succeeded in covering 77% of the Private Referral arm clusters (82 of 107 clusters). The modest rate of coverage of Public Advertisement arm clusters is an interesting finding in and of itself and is further discussed below. Despite the limited success of the public advertisement intervention, there are 72 clusters in the Public Advertisement arm covered by referred candidates, which include 53 clusters where recruited candidates from advertisement and referral interventions overlap. In my preferred specification, I will exclude clusters that only had informal last minute referrals and focus on the subset of 40 of these clusters in which there were formal referrals. In these 40 clusters I can directly compare candidates recruited through public advertisement and private formal referral search.

Why individuals in neighborhoods with some clusters responded to advertisements and in others did not is a question beyond the reach of our data. There is evidence, however, to reject the idea that variation in the response rate to public advertisements reflects inconsistent effort in the distribution process. For example, Figure 6 shows three separate clusters located less than two kilometers from one another in the same section of a major city. Each dot represents the location where a poster was posted or fliers were handed out. Despite significant distribution efforts in all three clusters in very close proximity, one of the clusters attracted 6 public candidates and the other two failed to yield any candidates.

4.4 Defining the Sample for Analysis

Given that the government does not have strict selection criteria, my choice of which observations to compare by intervention within each cluster is somewhat subjective and will depend on the nature of the research question. For example, in cases for which I have multiple candidates recruited by public advertisement for a given cluster, do I select the "best" candidate according to some subjective criteria or do I estimate the average of the candidates for the sake of comparison with the referred candidates? Conversely, among the referred candidates for a given cluster, some were informally referred to attend the screening survey last minute, some were referred formally with a paper form, and a subset of the formal referrals were ranked by the referrers. For comparison with public advertisement candidates, should I limit the analysis only to the top ranked formally referred candidates that may best reflect who would have been the referrer's first choice if the recruitment was entirely within their control as it normally is? Should I include all the formally referred candidates and omit those convened informally last minute who may have merely been convened out of convenience? Or should I include all the referred candidates? Who I elect to include will not only change the sample of candidates in a given cluster, but will also change the sample of clusters for which I have overlapping public advertisement and referral candidates. Comparison of results across alternative specifications will yield similiar findings. However, it is worth taking the sample construction seriously as it will influence how to interpret the findings.

The following list identifies three alternative ways of defining the sample and the associated primary research question implied by the sample definitions. Figure 5 illustrates the implications for sample size in terms of numbers of clusters and candidates for the alternative sample constructions.

1. **Compare all public to all referred.** How do applicants that self-select to volunteer in response to advertisements differ from the pool of candidates that local ProSoli staff *could attract to potentially volunteer*?

2. **Compare all public to all formally referred.** How do applicants that self-select to volunteer in response to advertisements differ from those potential volunteers local ProSoli staff *know well-enough to identify by name and personal identification number*?

3. **Compare all public to top-ranked formally referred.** How do applicants that respond to public advertisement differ from the *top candidate local ProSoli staff would recommend for the position*?

Self-reports by candidates about how they became aware of the CW opportunity will help inform whether or not informally referred candidates were selected substantially differently than formally referred candidates. Table 2 reveals that informal references were 5.7% less likely to believe that they had been referred by ProSoli staff and 4.8% less likely to know the ProSoli staff that referred them. This is equivalent to $\sim$11 of the 212 informally referred candidates. These differences are mostly explained by the higher rate by which informally referred candidates learned of the opportunity through a public advertisement, suggesting some ProSoli staff may have actually posted signs locally in their community to informally refer candidates. An additional concern about using the informally referred candidates is that they were not specifically ranked or assigned to a cluster of beneficiary households and in many cases could be assigned to multiple geographically proximate clusters, which would complicate analysis.

Whether or not to use all of the formally referred candidates or just those ranked highest will depend on the desired research question. Ultimately, however, the results discussed in greater depth in the next section reveal there is no consistent variation among observable characteristics of formally referred candidates by rank. Therefore, my preferred specification will use the subset of 40 clusters for which there were overlapping public candidates and formally referred candidates, labeled as "Ref-Form" in Figure 5 and described in Table 7.

5 Results

My principal findings suggest hiring firms and government agencies, in particular, confront real tradeoffs in choosing between public search and private referral recruitment processes. One tradeoff reflects tension between more experienced and dedicated public candidates and more cognitively skilled referral candidates. Public search yields job candidates that are superior to those recruited through private search across a wide variety of characteristics, such as education, labor market experience, civic engagement, and instrinsic motivation, that should be *a priori* important to the hiring government bureaucracy. However, private search candidates have unambiguously superior cognitive skills, performing 0.25 to 0.39 standard deviations higher on a Raven's test of cognitive skills even when controlling for key covariates. When selection in hiring is randomized between public versus private referred candidates, public candidates are 6.4% more likely to accept a job offer and 23.6% more likely to attend the subsequent training after accepting the offer with the magnitude of the impact increasing with basic controls.

A second tradeoff reflects tension between the ease of accessing larger pools of referrals versus the greater diversity of public candidates. The public advertisements attracted candidates to cover 19% fewer new beneficiary clusters than the intervention designed to mimic the status quo recruitment process. Whether this rate of coverage is relatively high or low is subjective, but it does suggest that there may be some locations for which public advertisements will consistently fail to attract candidates. Nevertheless, the potential pool of referral candidates has limitations, as well. My comparison of higher and lower ranked referred candidates demonstrates little variation across a wide variety of characteristics, suggesting the pool of referrals are relatively homogeneous.

A third tradeoff reflects the higher relative cost of the public search strategy. To estimate the marginal costs of the public intervention, I exclude the time cost of human resources required to distribute advertisements and answer calls to the call center as the government has a captive staff of field workers for the distribution work and a dedicated call center, which was actually used in this study. Focusing just on the hard costs of printing the advertisements, the cost of the public advertisement intervention was modest at approximately $17 per public candidate or $39 per beneficiary cluster, roughly equivalent to one to two months of a CW's travel stipend.[12] In comparison, a normal referral recruitment process is basically free. Whether or not the public intervention proves to be cost-effective will depend on the relative impacts of the hired public candidates versus the hired referral candidates.

The specific results are presented as follows. Tables 8-15 present the principal results comparing the means of candidates recruited through public and private searches across 8 families of characteristics. The distributions of select continuous and categorical variables are plotted separately to validate results from the comparison of means. Basic p-values are calculated using cluster robust variance estimator (CRVE) and observations are weighted by intervention type within cluster to ensure clusters are equally weighted despite the imbalance of observations across clusters. I also employ two approaches to correct for possible overrejection of the null hypothesis (that there is no difference between public and referred

[12] In estimating cost per beneficiary cluster, I use the 53 beneficiary clusters that were eligible to be covered by the recruited public candidates in the denominator.

types) due to multiple inference across 112 variables.[13] The main implication of the multiple inference adjustments is to moderate the unadjusted result that public and referred candidates vary significantly according to select personality traits observed through psychological tests.

The first multiple inference adjustment controls for the familywise error rate. For each family of outcomes I aggregate, normalize, and weight the variables indicated in Column "index" using the method recommended in Anderson (2008) to construct a single family index.[14] The differences in mean values between types for the family indices are consistent with patterns among individual variables in sign, but are not readily interpretable in magnitude. The family index p-values, which are estimated with the CRVE estimator, generally corroborate inference from individual outcomes with a couple exceptions in the political connectedness and personality families in which select variables exhibit significant differences, but the family indices do not.

The second multiple inference adjustment controls for false discovery rates. For each individual variable, I adjust the CRVE p-values to sharpened q-values using the technique and code from Anderson (2008), which is based on Benjamini et al. (2006). By design, the pattern in the sharpened q-values is consistent with the basic CRVE p-values, though in some cases where differences were significant using p-values, the probability of falsely rejecting the null is outside of conventional confidence levels using q-values.

Table 16 demonstrates that superior performance by referred candidates on a key indicator of cognitive ability is robust to the inclusion of numerous controls, including indicators for the amount of time taken in the survey and seriousness with which the candidate responded. Table 18 cherry-picks key variables that exhibit meaningful differences in comparisons between public and private referred candidates to demonstrate that referred candidates vary little according to the rank assigned by their referrer. Table 19 compares public candidates to top ranked candidates in the Ref-P1 sample that is limited to the 31 clusters with overlap between public and top ranked candidates as a robustness check to results from the preferred Ref-Form sample that includes all 40 clusters with overlap between public and all formally referred candidates. Finally, Table 17 analyzes the results of a small hiring sub-experiment to demonstrate that the use of public search has causal impacts on increasing the likelihood of recruited candidates accepting CW job offers and attending subsequent job trainings.

5.1 Demographics

Demographic characteristics are displayed in Table 8 and immediately reveal that the alternative recruitment efforts attracted quite different people. 57% of privately referred candidates are ProSoli beneficiaries, suggesting that CW Supervisors may look to the families they

[13] I do not include Table 12's time use indicators in adjustments for multiple inference as they require rather subjective decisions to normalize, are estimated in a different regression framework with unique controls, reflect latent variables captured in the other variable families, and are intended as a compliment to the core comparisons of means.

[14] I exclude variables for which there are many missing values (e.g. salary at prior job), for which there is no strong prior belief about which values are better or worse (e.g. household size), or which are indices of or autocorrelated with other variables by design (e.g. the Big 5 index aggregation of underlying five personality traits and alternative ways of classifying employment status).

themselves are responsible for monitoring to serve as CWs. Whether this is because they know them well, want to provide these people a constructive opportunity, or feel that they could easily persuade and/or monitor these individuals who would be working essentially as their assistants will be an important question for future research. In comparison, public candidates are 20% less likely to be ProSoli beneficiaries. They are also 16% more likely to be men, reflecting the low rate of referred male candidates and corroborating evidence that network referrals can augment gender inequalities in labor markets (Beaman et al., 2015). Public candidates are also older on average, which may explain some of the other differences in marriage rate, number of children, and likelihood of self-identifying as household head. Figure 7 exhibits the full distribution of ages, which seem to reflect a larger supply of public candidates in their 50-60's and a larger supply of privately referred candidates in their early 20's. How these demographic differences might impact their job performance is an interesting question for future research as energy and enthusiasm of younger candidates eager for opportunities to improve their position in the community and labor market would be balanced against the potential wisdom and authority of older individuals who may be more patient and convincing in motivating behavior change among their assigned beneficiary families. Figure 8 compares distributions of school grades completed, which significantly differ in favor of public candidates who have over a year of college education on average, according to Table 8.

5.2 Labor Market

The robust set of employment variables in Table 9 further evidences the relative virtues of public candidates who, not only are older and more educated, but also are more likely to be currently employed with greater likelihood of permanent work at higher salaries. As expected, the labor questions in the survey elicited different responses depending on how they were asked. Only 13% of public candidates responding that their primary occupational status was that of homemaker, well below the private referred candidates, but 82% self-identify as homemakers when asked directly in a followup question, a rate slightly higher but statistically similar to referred candidates. Notably, many men identified as homemakers as well. Figures 9 and 10 display the distributions of monthly salary from i) the last reported job and ii) income over the last 7 days in log form conditional on having income. Differences favor public candidates having higher salaries and are more statistically significant when comparing distributions than when comparing means and clustering errors. For the 34% of public candidates that earned income in the week prior to the survey, the income differences compared to referred candidates are substantial and equivalent to roughly 8.6 times the expected travel stipend provided to CW volunteers.

Whether having higher wages is a positive or negative attribute is an open question. Higher wages may indicate greater general aptitude on observable and unobservable dimensions. Employers would have to weigh this against higher opportunity costs and greater demands on one's time that may distract volunteers from performing the poorly compensated duties. The substantial variation in salaries for both public and referred candidates reveals that opportunity costs vary widely and does not lend itself to a clear salary target were the government interested in motivating workers with greater monetary incentives. Interestingly, there are approximately the same number of earners in the households of public and referred candidates, however, public candidates are far more likely to be their household's primary earner. This suggests public candidates may have more bargaining power to spend their time as they please within their households, but may also be more depended on to work as much as possible.

5.3 Wealth, Food Security, Assets, and Finance

Looking more broadly at household level income, assets, and general economic well-being, Table 10 is consistent with the theme that public candidates are better off. However, differences in monthly equivalent rental value of the home, likelihood of renting, and ownership of basic household durables and assets are not statistically significant at conventional levels in a comparison of means or when comparing the full distributions in Figure 11. Household income among public candidates displays greater variation both at upper and lower tails of the distribution in Figure 12 and is higher for public candidates in the comparison of means, but not significantly so at conventional levels after adjusting for false discovery rates with sharpened q-values. The more general observation that there seems to be greater variation in many outcomes for public candidates reflects a result that we explore more later: that there seems to be little variation in referred candidates. The greater variation for public candidates may yield superior characteristics on average. But, greater uncertainty may also be a major drawback for an employer recruiting with a public search. Notably, Table 10 shows that referred candidates were 9.0% more likely than public candidates to respond above the median on a Likert scale to the question "How frequently have you gone hungry without eating?" though answers to similar food security questions about skipping meals and medication were virtually identical. Consistent with having higher incomes and more permanent jobs, public candidates are 14% more likely to report any savings, 20% more likely to have a bank account, and 26% more likely to have taken out a loan over the last 12 months. To the extent CWs may encourage their ProSoli beneficiary families to save and participate in the formal financial system, these differences in experience with personal finance may further suggest the benefit of pursuing candidates through public search.

5.4 Civic Engagement and Leadership

Among candidates being recruited for community volunteer work with poor vulnerable households, one would expect the government to value prior involvement in the civic affairs of the community, leadership experience, and potentially even local political involvement. Table 11 reveals findings across these characteristics again demonstrating the relative virtue of public candidates. Referral candidates were 23% less likely to be members of the neighborhood council, 12% less likely to respond above the median in terms of frequency of attendance at community meetings, 24% less likely to have worked on the 2012 Presidential campaign, and 16% less likely to be planning on working on the 2016 campaign, which was being held only a few months after the screening survey was fielded. Maybe most notably, public candidates were 22% and 20% more likely to answer above the median Likert scale response to questions about the extent to which they consider themselves an effective community leader currently or expect to be a leader within five years, respectively.

Given that the vast majority of the referrers of candidates were local CW supervisors and Field Supervisors that know their neighborhoods intimately, it is hard to imagine that they do not know many of the local neighborhood council members that applied for CW jobs through the public search. Whether or not they conscientiously avoid referring these local leaders to work as CWs out of competition for local influence, because they think the position would provide greater benefit to less empowered individuals, or for others reasons provides an interesting topic for further inquiry.

5.5 Time Use

Table 12 describes how candidates spent their time across a number of different activities the day prior to being surveyed. The estimates further validate results from demographic, labor, and civic engagement tables, suggesting public candidates spend more time working and volunteering and less time doing household chores and participating in educational activities. Table 12 reports the mean for the pooled sample of both public and referral candidates and the coefficient on a dummy variable for being a public candidate from a linear regression controlling for the day of the week the interview was conducted and the sum of all the hours accounted for by the candidate. On average public candidates worked 72 minutes more and volunteered 17 minutes more than referred candidates.

5.6 Political Connectedness

To the extent CW positions are desirable as a means of gaining local influence, demonstrating leadership, networking with the government, etc., one might expect that they would be provided as rewards to more politically connected individuals. However, Table 13 reveals the opposite. It is public candidates who more frequently could name and claim to have had at least one conversation with the full array of political representatives, from neighborhood association members to senators, despite the fact that they were no more likely to be related to these elected officials. This outcome is consistent with their greater personal involvement in civic and political life and seems to rule out the use of CW positions as rewards for political patronage. These results should be interpretted with some caution, however, as adjustments for multiple inference render the statistical signifcance of the differences in means less convincing. Table 4 presents some crude evidence of positive associations between the seniority of the referrer and the political connectedness of the referred candidate, though there is insufficient variation in the source of the referral by seniority to examine this association in detail.

5.7 Motivations and Personality

Why, afterall, do people even want these poorly paid volunteer positions? Table 14 illustrates answers to questions about the salience of a variety of potential intrinsic and extrinsic motivations. Comparisons between the candidates are consistent with prior observations that public candidates are less in need of work and are more civically engaged and intrinsically motivated. These motivations translate into higher scores in Table 15 on all of the Big 5 personality characteristics but Agreeableness, though differences are only marginally significant when combined into a single Big 5 index and virtually disappear when controlling for familywise error rates and false detection rates in the multiple inference adjustments.[15] The association of these personality characteristics with real world behavior in prior work suggests these differences could be meaningful, particularly in the context of performing hands-on community service (John, 1990; Heckman, 2011; Callen et al., 2015). Other tests of personality, including Dictator Games and assessments of Stress, Expressivity, and Aspirations, failed to exhibit much variation even within candidate types likely reflecting the limitations of using these tests in a hiring process.[16] Public candidates did express greater creativity in a test that counted the number of answers they gave when asked to list all of the uses of a concrete block that they could think of in 45 seconds. Thus, public candidates appear to be more intrinsically motivated, to reveal greater capacity for creative thinking, and to be at least equal if not superior to referral candidates on Big 5 personality traits - all characteristics one would generally consider to be important for front-line community workers specifically and for employees in service oriented sectors more broadly.

5.8 Cognitive Ability

The two indicators of cognitive ability in Table 15 constitute the main characteristics upon which privately referred candidates seem to outperform public candidates. In Table 15's comparison of means and the comparison of distributions in Figure 13, referred candidates unambiguously score higher on an 11 point Raven's matrices test of non-verbal abstract reasoning.[17] Results from OLS regressions in Table 16 confirm that the significant difference in performance on the Raven's test stands up to converting the raw scores into z-scores and controlling for a variety of covariates, including cluster fixed effects. To ensure that the Raven's score does not confound ability and concentration, Column (4) in Table 16 controls for the amount of time spent completing the Raven's module and the survey enumerator's subjective assessment of the candidate's level of seriousness in taking the test, revealing referred candidates still perform a third of a standard deviation better when including all these controls and cluster fixed effects. This result is statistically significant at conventional levels despite the limited statistical power of a 191 observation sample. The second indicator of cognitive ability is less compelling, but worth mentioning as it confirms the pattern in the Raven's. A two question test of working memory executive function was embedded in the survey to test respondents' ability to recall basic facts late in the survey from a written passage they were made to read early in the survey. Referred candidates were more likely to

[15]The Big 5 index is constructed in the standard way as an average of the z-scores of each of the five sub-dimensions, reverse-coding neuroticism. See John (1990) for further detail.

[16]8 question stress scale derived from Peacock and Wong (1990). Expressivity scale from Gross and John (1997). Aspirations scale from Kasser and Ryan (1996).

[17]Respondents completed Raven's Progressive Matrices Set I, which includes 12 questions. The first question was used as practice such that the maximum score is 11 in this study. The Raven's test is one of the most commonly used assessments of cognitive ability.

correctly recall the answers to the questions despite public candidates having higher scores on reading comprehension questions asked immediately after reading the written passage. Collectively, these two tests suggest robust differences in cognitive ability.

Although public candidates appear superior on virtually every family of characteristics, cognitive ability constitutes a key dimension upon which employers often seek to hire workers and is measured in a way that cannot be manipulated by the respondent. In lieu of time consuming screening tests such as those administered in this study, cognitive ability and intelligence more broadly are not easily observable to employers and may reflect a key dimension on which referrers with private information about a prospective candidate can outperform a public search. Confronted with deciding among alternative recruitment mechanisms, employers may have to weigh the benefits of cognitive ability versus other important characteristics. The extent to which cognitive ability is associated with future job performance among CWs is a key question for followup investigation.

5.9 Behavioral Outcomes: Job Offer Acceptance

Although data on job performance is not yet available, I can use job acceptance and subsequent participation in job training to assess the extent to which public candidates might be more motivated and persistent in working as a CW. In fact, among all candidates offered jobs, 93.0% of public candidates accepted the job versus 86.3% of referred candidates and 76.7% of public candidates offered the job attended the subsequent job training versus 61.1% of referred candidates. This raw association in the data may, however, reflect selection bias given that public candidates were not recruited to all of the clusters originally intended and arithmetic bias in over-weighting clusters in which candidates reject offers.

A key feature of the actual screening and selection process allows me to make stronger causal claims about the positive impacts of the public recruitment process on candidate acceptance and persistence. Among 48 of the clusters where public and private referred sub-samples overlapped, the government randomized whether to hire from the public or referred candidate sample. Furthermore, candidates were randomly selected for job offers from among the pool of candidates within a given cluster from whichever of the public versus referral types had been randomly selected for use in that cluster. In this context, the acceptance rates of the candidates offered jobs should be orthogonal to the characteristics of the clusters. Therefore, comparisons should be free of the selection bias driving incomplete coverage of the public advertisement clusters, for example. To be clear, acceptance rates will still endogenously reflect the characteristics of types of people attracted through public versus referral mechanisms. But this is precisely the point: to see how the different types of people attracted to the job through different recruitment mechanisms respond to offers. The sample here is only 48 people of which three did not accept the offer (two of 20 referral and one of 28 public) and 15 did not show up for the subsequent training (9 or 20 referral and 6 of 28 public); so, regression results should be interpreted with caution. Nevertheless, the experimental estimates in Table 17 confirm the pattern in the raw data that public candidates are more likely to accept the job offer and to attend the training conditional on receiving the job offer (the intention to treat). On average, the public advertisement intervention recruits candidates that are 6.4% more likely to accept job offers and 23.6%

more likely to attend subsequent job trainings, which rises to 8.5% and 30.8% respectively with the use of basic controls for which a hiring agency could conceivably screen. These small sample estimates are not significant for acceptance rates, but are significant at 90% confidence for ITT estimates of participation in training. Among organizations for which employee turnover is costly and for which there may be limited capacity to dynamically recruit and screen new job candidates, these findings may demonstrate a key benefit of public search recruitment. From the perspective of ProSoli program effects, persistence may be a key mediator of household impacts as CWs learn on-the-job and develop tight relationships with their beneficiary families over time.

5.10 Lack of Variation Among Referred Candidates by Ranking

As mentioned previously, formally referred candidates appear to exhibit less variation on a number of key characteristics than public candidates. Nevertheless, to the extent ProSoli staff know potential candidates well and are motivated to target their most preferred candidates for the CW positions they themselves will have to monitor and supervise, we might expect the top ranked CW candidates to have systematically superior characteristics to the lesser ranked candidates. I can exploit the fact that the formal referral process required ranking of candidates to test whether top candidates are in fact observably different than lower ranked candidates. Table 18 compares top ranked candidates to a pooled sample of lesser ranked candidates on a cherry-picked sample of variables that exhibited substantial variation between public and referred candidates. There is no discernible pattern of differences. This lack of variation by rank may reflect the lack of variation in the population networked to the referring ProSoli staff. For example, ProSoli beneficiaries are well known to the ProSoli staff, more likely to be referred, and by definition come from a narrower sample frame than the general population in the same neighborhoods. Alternatively, this might reflect ProSoli staff actively targeting certain more narrowly defined types of people for inclusion as CWs, being poorly motivated to look far for potential candidates, or being unable to effectively discern differences among potential candidates. Given that CW supervisors are relatively low level ProSoli staff themselves living in poor neighborhoods often with little higher education, this pattern may reflect a finding from the Beaman and Magruder (2012) field experiment that, unlike higher skilled workers, the lowest skilled workers were unable to refer high skilled individuals even when faced with strong incentives to do so.

6 Discussion

In a natural setting in collaboration with a major government agency across multiple provinces of a developing country, this study employs a novel research design to rigorously test the benefits of an oft-overlooked counterfactual to referral-based recruitment. This counterfactual open public search presents a clear and viable recruitment method, similar to options available to the management of virtually any organization.

Results from the study suggest that employers may face tradeoffs in deciding among recruitment strategies. In my study context, candidates recruited through public advertisement appear to have superior characteristics across education, labor market experience, socio-economic status, civic engagement, leadership, political connectedness, and intrinsic motivation categories. They also behaviorally demonstrate higher rates of acceptance and persistence in the early stages of the job. These virtues must be weighed against the higher cognitive ability of referred candidates and the greater coverage and lower marginal cost of referral-based recruiting strategies. Nevertheless, the apparent superiority of the public search candidates across many dimensions here begs the question of why these front-line public sector jobs rely on private referrals and offers a promising policy alternative to improving the quality and persistence of front-line community workers. Interpreted more broadly, these findings question the conventional wisdom that the prevalence of referrals in job markets coupled with empirical evidence explaining various plausible benefits of referral processes sufficiently demonstrate the superiority of private targeted searches in comparison to well-defined and sufficiently resourced open public searches.

The patterns revealed in comparisons of public and private referral candidates introduce a number of new questions deserving further explorations. Better understanding the mechanisms that led to the patterns of recruitment, the motivations of referrers, the explanation for the lack of variation among ranked referrals, the importance of specific personal characteristics in mediating job performance, and the benefits of the job opportunity to the different types of candidates offer promising topics for future research.

References

Afridi, F., Dhillon, A., and Sharma, S. (2015). Social Networks and Labour productivity: A survey of recent theory and evidence. CAGE Online Working Paper Series 243, Competitive Advantage in the Global Economy (CAGE).

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484):1481–1495.

Ashraf, N., Bandiera, O., and Lee, S. S. (2014). Do-Gooders and Go-Getters: Career Incentives, Selection, and Performance in Public Service Delivery. *Mimeo*.

Beaman, L. and Magruder, J. (2012). Who Gets the Job Referral? Evidence from a Social Networks Experiment. *American Economic Review*, 102(7):3574–93.

Beaman, L., Magruder, J., and Keleher, N. (2015). Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi. *Mimeo*.

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.

Burks, S. V., Cowgill, B., Hoffman, M., and Housman, M. (2015). The Value of Hiring through Employee Referrals. *The Quarterly Journal of Economics*.

Callen, M., Gulzar, S., Hasanain, A., Khan, Y., and Rezaee, A. (2015). Personalities and Public Sector Performance: Evidence from a Health Experiment in Pakistan. Working Paper 21180, National Bureau of Economic Research.

Calvo-Armengol, A. and Jackson, M. O. (2004). The Effects of Social Networks on Employment and Inequality. *American Economic Review*, 94(3):426–454.

Calvo-Armengol, A. and Jackson, M. O. (2007). Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory*, 132(1):27 – 46.

Camacho, A., Cunningham, W., Rigolini, J., and Silva, V. (2014). Addressing access and behavioral constraints through social intermediation services : a review of Chile Solidario and Red Unidos. *World Bank Policy Research Working Paper No. WPS 7136*.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., and Rogers, F. H. (2006). Missing in Action: Teacher and Health Worker Absence in Developing Countries. *Journal of Economic Perspectives*, 20(1):91–116.

Dal Bo, E., Finan, F., and Rossi, M. A. (2013). Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service. *The Quarterly Journal of Economics*, 128(3):1169–1218.

Dhillon, A., Iversen, V., and Torsvik, G. (2013). Employee Referral, Social Proximity And Worker Discipline: Theory And Evidence From India. Working Papers in Economics 04/13, University of Bergen, Department of Economics.

Dixit, A. (2002). Incentives and Organizations in the Public Sector: An Interpretative Review. *Journal of Human Resources*, 37(4):696–727.

Fafchamps, M. and Moradi, A. (2015). Referral and Job Performance: Evidence from the Ghana Colonial Army. *Economic Development and Cultural Change*, 63(4):715 – 751.

Gross, J. J. and John, O. P. (1997). Revealing feelings: Facets of emotional expressivity in self-reports, peer ratings, and behavior. *Journal of Personality and Social Psychology*, 72(2):435–448.

Heath, R. (2013). Why do Firms Hire Using Referrals? Evidence from Bangladeshi Garment Factories. *Mimeo*.

Heckman, J. J. (2011). Integrating Personality Psychology into Economics. Working Paper 17378, National Bureau of Economic Research.

John, O. (1990). The “Big Five” factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In *Handbook of personality: Theory and research*, pages 66–100. Guilford Press., New York.

Kasser, T. and Ryan, R. M. (1996). Further examining the american dream: Differential correlates of intrinsic and extrinsic goals. *Personality and Social Psychology Bulletin*, 22(3):280–287.

Munshi, K. and Rosenzweig, M. (2006). Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy. *American Economic Review*, 96(4):1225–1252.

Pallais, A. and Sands, E. G. (2015). Why the referential treatment: Evidence from field experiments on referrals. Working Paper 21357, National Bureau of Economic Research.

Peacock, E. J. and Wong, P. T. P. (1990). The stress appraisal measure (sam): A multidimensional approach to cognitive appraisal. *Stress Medicine*, 6(3):227–236.

Table 1: Summary Statistics for Raw Candidate Sample

Statistic	N	Mean	St. Dev.	Min	Max
Male	865	0.24	0.43	0	1
Beneficiary	865	0.53	0.50	0	1
Age	860	31.49	11.16	17.52	72.03
Education	865	12.01	2.95	1.00	19.00
Hhld Head	865	0.27	0.44	0	1
Insured	864	0.72	0.45	0	1
Permanent Work	865	0.18	0.39	0	1
Homemaker	865	0.74	0.44	0	1
Last Job Salary	758	7,153.95	6,556.93	150	85,000
Hhld Income	781	11,723.09	9,154.24	0	70,000
Food Insecurity	865	0.37	0.48	0	1
Neighbor Committee	864	0.20	0.40	0	1
Involved Comm Org	865	0.40	0.49	0	1
Campaign 2016	865	0.25	0.43	0	1
Leader	865	0.61	0.49	0	1
Know City Manager	863	0.41	0.49	0	1
Motive Fut Job	862	4.32	1.16	1	5
Motive Gov Connection	861	4.32	1.12	1	5
Motive Bored	857	4.28	1.21	1	5
Ravens	865	3.92	2.45	0	11
Ravens z-score	865	0.00	1.00	−1.60	2.89
Exec Function	865	1.03	0.76	0	2
Reading Ability	865	1.74	0.89	0	3
Big5 Index	865	−0.00	0.70	−4.10	1.44
Creativity	864	5.25	2.17	1	16

Note: Raw candidate sample includes public candidates and referral candidates.

Table 2: Knowledge of Referrer or Alt. Source of Information about CW Job

	Public	Ref-Formal	Ref-Informal	Count
Referred by Unknown Prosoli Staff	4.7%	2.3%	1.4%	21
Referred by Known Prosoli Staff	11.8%	83.1%	78.3%	618
Informed by a Friend or Relative	12.6%	12.0%	12.3%	105
Informed by Public Advertisement	70.1%	1.7%	5.2%	109
Informed at ProSoli Info Center	0.0%	0.2%	0.5%	1
Informed by Member of Benef. Cluster	0.0%	0.2%	0.5%	2
Informed by Supervisor of Family School	0.0%	0.4%	0.5%	3
Informed by ProSoli Other Staff	0.0%	0.2%	1.4%	4
Informed by Pastor or Member of Church	0.8%	0.2%	0.0%	2
Total Count	127	526	212	865

Note: "Referred Unknown" in the first row indicates the CW candidate affirmed that they had specifically been referred by someone who works for ProSoli, but they did not know who it was that referred them. "Referred Known" indicates they did claim to know who referred them whether or not the person they named was in fact the source of the referral. When candidates responded that they had not been referred, they were asked "From whom did you first hear of the opportunity to be a ProSoli volunteer?" The options in Table 2 that begin with "Informed" indicated their choice in response to this question. "ProSoli Info Center" refers to vocational training centers associated with ProSoli or local offices where ProSoli provides information and administrative assistance to beneficiaries. "Member Benef. Cluster" refers to the clusters to which each beneficiary household is assigned and to which a given CW provides services, often ranging from 30-60 households. "Supervisor of Family School" refers to the CW Supervisor who runs the monthly Family Schools to which beneficiaries must attend. A given CW Supervisor normally manages 6-10 CWs.

Table 3: Comparing Alternative Constructions of Referred Sub-Sample

	Ref-All	Ref-Form	Ref-Rank	Ref-P1
Male	0.22	0.22	0.21	0.22
Beneficiary	0.57	0.59	0.60	0.59
Age	30.54	30.36	31.05	31.64
Education	11.91	11.94	11.83	12.08
Hhld Head	0.24	0.23	0.23	0.24
Insured	0.72	0.73	0.74	0.71
Permanent Work	0.17	0.15	0.15	0.13
Homemaker	0.74	0.73	0.76	0.77
Last Job Salary	6663.76	6445.68	6846.27	6693.20
Hhld Income	11220.99	10957.49	11196.00	11088.08
Food Insecurity	0.38	0.37	0.37	0.37
Neighbor Committee	0.17	0.18	0.18	0.20
Involved Comm Org	0.37	0.37	0.36	0.38
Campaign 2016	0.23	0.23	0.22	0.23
Leader	0.58	0.57	0.57	0.54
Know City Manager	0.38	0.40	0.38	0.32
Motive Fut Job	4.43	4.44	4.45	4.39
Motive Gov Connection	4.40	4.45	4.45	4.39
Motive Bored	4.32	4.31	4.32	4.30
Ravens	3.96	3.94	3.99	4.00
Ravens z-score	0.02	0.01	0.03	0.03
Exec Function	1.07	1.04	1.02	1.08
Reading Ability	1.71	1.69	1.68	1.68
Big5 Index	-0.02	-0.03	-0.03	0.01
Creativity	5.14	5.09	5.12	5.39
N	738	526	454	167

Note: Table 3 compares characteristics across samples constructed according to different methods. "Ref-All" includes all referred candidates, excluding 14 candidates for whom address information was mistakenly omitted during the survey. "Ref-Form" is a subset of "Ref-All" that includes only those candidates formally referred using a referral form, which required knowledge of candidate name, ID number, telephone number, reading ability, and address, and excludes informally referred candidates. Though referrals were intended to all be formally reported on referrals forms, referrals in many clusters were informally invited to the screening survey last minute without the use of the forms. "Ref-Rank" is a subset of "Ref-Form" and includes all referred candidates ranked when formally referred. Though referrals were intended to rank their referred candidates, not all referrers complied. "Ref-P1" is a subset of "Ref-Rank" and only includes those candidates ranked top priority by referrers for a given cluster, excluding all those that are not ranked or are ranked lower than top-priority. Reported variables are selected from among the full set of variables to reflect indicators that differ most meaningfully in the comparison of public to referred candidates in later tables. All values reported are means unless otherwise indicated.

Table 4: Civic and Political Characteristics by Source of Private Referral

	CW	CW Sup	Field Sup	Reg Manager	Prov.Assist
N	28	389	27	6	8
% of total	6.1%	84.9%	5.8%	1.3%	1.7%
Time Political	0.00	0.04	0.15	0.37	0.00
Neighbor Committee	0.17	0.11	0.17	0.33	0.00
Petition	0.33	0.07	0.30	0.56	0.38
Campaign 2012	0.17	0.32	0.19	0.30	0.12
Campaign 2016	0.17	0.25	0.21	0.33	0.12
Leader	0.67	0.50	0.58	0.59	0.38
Know City Manager	0.50	0.18	0.38	0.48	0.62
Spoken w City Manager	0.67	0.07	0.26	0.44	0.62

Note: All values reported are means unless otherwise indicated. The columns correspond to 5 alterative sources of referral reported in increasing level of seniority left to right according to position in the ProSoli bureaucratic hierarchy. CW refers to existing community workers as opposed to candidates for new CW positions. Reported variables are selected from among the full set of variables to reflect the most meaningful indicators of involvement in civic and political life.

Table 5: Comparing Candidates to Neighbors in General and Beneficiary Pop.

	(1)	(2)	(3)
	General (2010)	Hhld Sample (2012)	CW Candidates (2016)
Urban	0.94	0.92	–
Male	0.49	0.43	0.24
Edu	8.89	9.98	12.01
Hhld Size	3.59	4.02	4.55
Single	0.21	–	0.45
Married	0.18	–	0.11
Separated	0.19	–	0.08
Permanent Work	–	0.25	0.18
Occasionally Work	–	0.14	0.11
Temp Work	–	0.06	0.07
Unemployed	–	0.16	0.17
Other Work	–	0.39	0.47
Homemaker	0.44	0.30	0.74
Student	0.16	0.07	0.47
Asset Index Short	3.76	2.56	3.83
Renter	0.43	0.37	0.31
Insured	–	0.19	0.72
N	1,331,931	31,234	865

Note: All values reported are means. Column (3) shows the raw unweighted means for the full sample of interviewed CW candidates. Column (2) estimates include all the ProSoli beneficiaries in clusters for which candidates were effectively recruited, excluding the ∼35% of individuals that were surveyed prior to 2010. The majority of the SIUBEN poverty census data were fielded in a first major survey in 2005-2006 and then a second wave in 2011-2012, which were intended to survey the entire population of targeted neighborhoods. However, the second wave failed to reach all of the targeted population, which is why some individuals had been last surveyed prior to 2010. Additionally, the SIUBEN actively seeks to include individuals that were omitted on a rolling basis, such that some of those included in Column (2) were surveyed before or after the second major wave of SIUBEN surveys in 2010 or 2013-2015. CW candidates were required to be adults, so the sample in Column (2) is further limited to 18-75 year olds, the approximate range of CW candidates. Column (1) estimates use the full population of the 2010 National Census, restricting observations to 18-75 year old individuals living in the same neighborhoods as individuals included in the ProSoli beneficiary sample used to calculate Column (2). Neighborhoods are defined by the codes for administrative units called "*barrio/paraje*" used in the 2010 national census and the second major wave of the SIUBEN.

Table 6: Candidate Type by Province

Province	Public	Reference	Total	Percent
Azua	4	21	25	2.9%
Barahona	3	79	82	9.5%
Bahoruco	16	97	113	13.1%
San Juan	16	62	78	9.0%
Santiago	8	156	164	19.0%
Santo Domingo	80	309	389	45.0%
Other	0	14	14	1.6%
Total	127	738	865	100%

Note: The broader RCT that defined the sample of household to be served by CW candidates formally targets 7 provinces, including the National District (*Distrito Nacional*) of the city of Santo Domingo, which is formally considered a province according to statistical agencies, but reported as part of Santo Domingo in Table 6. Several CW candidates from San Cristobal, the province just adjacent and within commuting distance to Santo Domingo city, were surveyed and are included in the Santo Domingo category. 14 observations are categorized as "Other" because their addresses were mistakenly omitted from the survey. The "Reference" column includes both formally and informally referred candidates.

Table 7: Sum. Stats for Preferred Sample (Formally Ref & Overlapping Public)

Statistic	N	Mean	St. Dev.	Min	Max
Male	195	0.27	0.44	0	1
Beneficiary	195	0.48	0.50	0	1
Age	193	33.46	11.38	17.95	72.03
Education	195	12.23	2.91	3.00	19.00
Hhld Head	195	0.30	0.46	0	1
Insured	195	0.76	0.43	0	1
Permanent Work	195	0.17	0.38	0	1
Homemaker	195	0.79	0.41	0	1
Last Job Salary	178	8,388.33	8,682.84	500	85,000
Hhld Income	178	12,729.11	9,801.49	7	70,000
Food Insecurity	195	0.34	0.47	0	1
Neighbor Committee	194	0.27	0.45	0	1
Involved Comm Org	195	0.39	0.49	0	1
Campaign 2016	195	0.25	0.43	0	1
Leader	195	0.68	0.47	0	1
Know City Manager	195	0.46	0.50	0	1
Motive Fut Job	194	4.17	1.29	1	5
Motive Gov Connection	194	4.20	1.25	1	5
Motive Bored	193	4.30	1.29	1	5
Ravens	195	3.82	2.43	0	11
Ravens z-score	195	−0.04	0.99	−1.60	2.89
Exec Function	195	0.94	0.79	0	2
Reading Ability	195	1.79	0.92	0	3
Big5 Index	195	0.11	0.63	−1.83	1.24
Creativity	195	5.65	2.31	1	14

Note: The preferred sample, elsewhere referred to as "Ref-Form," includes 195 observations in the 40 clusters for which there are overlapping samples of public candidates and formally referred candidates. The Ravens z-score is calculated based on the full raw sample of candidates.

Table 8: Demographic

	public	reference	dif	pval	qval	index
Demographic Fam Index	1.23	0.88	0.35	0.00	–	-
Beneficiary	0.36	0.57	-0.20	0.04	0.14	Y
Male	0.33	0.17	0.16	0.05	0.14	Y
Age	37.89	31.59	6.30	0.00	0.01	Y
Education	13.12	11.89	1.23	0.01	0.07	Y
Hhld Size	4.05	4.59	-0.54	0.06	0.15	-
Children	2.56	2.13	0.42	0.13	0.22	-
Single	0.32	0.44	-0.12	0.15	0.23	-
Married	0.28	0.09	0.19	0.01	0.06	Y
Separated	0.09	0.06	0.02	0.55	0.47	-
Open Relationship	0.32	0.41	-0.09	0.31	0.35	-
Hhld Head	0.48	0.23	0.25	0.00	0.02	Y
Insured	0.80	0.78	0.02	0.76	0.52	Y

Note: Comparison of public versus formally referred candidates in the 40 Public Advertisement arm clusters in which there is overlap. Column "pval" refers to p-values estimated using CRVE without correcting for multiple inference issues that may arise when comparing candidate types across outcomes. Column "qval" adjusts p-values to control for False Discovery Rates in using the sharpened q-value method from Anderson (2008) based on Benjamini et al. (2006). Column "index" indicates whether or not a variable was used in the construction of the family index in the first row, which is used to control for the Familywise Error Rate. Variables included in family indices are coded so better outcomes are larger, normalized, aggregated, and weighted according to the procedure described in Anderson (2008). All estimation, including the Family Index, uses cluster-by-type proportional weights so clusters are equally weighted.

Table 9: Labor Market

	public	reference	dif	pval	qval	index
Labor Fam Index	1.08	0.94	0.14	0.07	–	-
Employed	0.58	0.29	0.29	0.00	0.02	Y
Student	0.34	0.44	-0.10	0.18	0.25	-
Homemaker	0.82	0.77	0.05	0.54	0.47	Y
Permanent Work	0.29	0.11	0.18	0.01	0.04	Y
Occasional Work	0.22	0.11	0.11	0.09	0.19	-
Temp Work	0.07	0.07	0.00	0.98	0.64	-
Unemployed	0.18	0.15	0.03	0.62	0.49	-
Homemaker Occup	0.13	0.32	-0.19	0.01	0.07	-
Student Occup	0.07	0.23	-0.17	0.00	0.03	-
Ever worked	0.95	0.89	0.05	0.20	0.26	Y
Last Job Perm	0.76	0.54	0.22	0.00	0.03	-
Last Job Temp	0.13	0.32	-0.19	0.00	0.02	-
Last Job Ocass	0.11	0.14	-0.03	0.59	0.47	-
Time Since Job	396.41	474.44	-78.02	0.69	0.51	-
Last Job Salary	11421.80	6880.90	4540.89	0.05	0.14	-
Last Wk Income	2559.53	549.16	2010.37	0.06	0.15	Y
Worked last wk	0.34	0.25	0.08	0.30	0.35	Y
Prime worker	0.52	0.18	0.33	0.00	0.01	Y
No. Earners	1.47	1.42	0.05	0.68	0.51	Y

Note: Comparison of public versus formally referred candidates in the 40 Public Advertisement arm clusters in which there is overlap. Column "pval" refers to p-values estimated using CRVE without correcting for multiple inference issues that may arise when comparing candidate types across outcomes. Column "qval" adjusts p-values to control for False Discovery Rates in using the sharpened q-value method from Anderson (2008) based on Benjamini et al. (2006). Column "index" indicates whether or not a variable was used in the construction of the family index in the first row, which is used to control for the Familywise Error Rate. Variables included in family indices are coded so better outcomes are larger, normalized, aggregated, and weighted according to the procedure described in Anderson (2008). All estimation, including the Family Index, uses cluster-by-type proportional weights so clusters are equally weighted.

Table 10: Wealth, Food Insecurity, Assets, Finance

	public	reference	dif	pval	qval	index
Wealth Fam Index	1.12	0.97	0.15	0.05	–	-
Hhld Income	15824.36	11354.65	4469.71	0.05	0.14	-
Income Increasing	0.29	0.33	-0.03	0.66	0.50	-
Income Decreasing	0.39	0.33	0.06	0.49	0.43	-
Remitances	0.69	0.82	-0.14	0.05	0.14	-
Depend on Remit	0.20	0.13	0.07	0.25	0.32	-
Food Insecurity	0.29	0.38	-0.09	0.20	0.26	Y
Skipped Meal	0.11	0.13	-0.02	0.77	0.52	-
Skipped Meds	0.50	0.44	0.05	0.58	0.47	-
Renter	0.40	0.37	0.03	0.77	0.52	-
Home value	3451.07	3303.99	147.08	0.67	0.51	-
Asset Index	5.33	5.38	-0.05	0.83	0.56	Y
Asset Index Short	3.93	3.92	0.01	0.97	0.64	-
Dirt Road	0.29	0.38	-0.09	0.28	0.33	Y
Have Any Savings	0.30	0.17	0.14	0.06	0.15	Y
Bank Account	0.62	0.42	0.20	0.01	0.05	Y
Loan Last 12 mo	0.45	0.19	0.26	0.00	0.02	Y
Purchase Credit	0.57	0.63	-0.06	0.47	0.41	Y

Note: Comparison of public versus formally referred candidates in the 40 Public Advertisement arm clusters in which there is overlap. Column "pval" refers to p-values estimated using CRVE without correcting for multiple inference issues that may arise when comparing candidate types across outcomes. Column "qval" adjusts p-values to control for False Discovery Rates in using the sharpened q-value method from Anderson (2008) based on Benjamini et al. (2006). Column "index" indicates whether or not a variable was used in the construction of the family index in the first row, which is used to control for the Familywise Error Rate. Variables included in family indices are coded so better outcomes are larger, normalized, aggregated, and weighted according to the procedure described in Anderson (2008). All estimation, including the Family Index, uses cluster-by-type proportional weights so clusters are equally weighted.

Table 11: Civic Engagement and Leadership

	public	reference	dif	pval	qval	index
Civic Fam Index	1.08	0.92	0.16	0.03	–	-
Neighbor Committee	0.41	0.18	0.23	0.00	0.01	Y
Attend Neigh Commit	0.66	0.53	0.12	0.08	0.18	Y
Attend City Council	0.34	0.33	0.01	0.89	0.58	Y
Petition	0.36	0.26	0.09	0.22	0.29	Y
Involved Relig Org	0.46	0.48	-0.03	0.78	0.53	Y
Involved School Org	0.13	0.07	0.05	0.22	0.29	Y
Involved Comm Org	0.51	0.29	0.22	0.02	0.08	Y
Involved Political Org	0.35	0.25	0.10	0.12	0.22	Y
Involved Womens Org	0.26	0.23	0.03	0.69	0.51	Y
Strike	0.08	0.06	0.02	0.62	0.49	Y
Registered Vote	0.97	0.95	0.03	0.44	0.41	Y
Voted in 2012	0.92	0.69	0.22	0.00	0.01	-
Voted in 2016	0.96	0.98	-0.02	0.59	0.47	Y
Campaign 2012	0.38	0.14	0.24	0.00	0.01	-
Campaign 2016	0.31	0.15	0.16	0.01	0.07	Y
Leader	0.84	0.62	0.22	0.00	0.02	Y
Future Leader	0.73	0.53	0.20	0.02	0.08	Y

Note: Comparison of public versus formally referred candidates in the 40 Public Advertisement arm clusters in which there is overlap. Column "pval" refers to p-values estimated using CRVE without correcting for multiple inference issues that may arise when comparing candidate types across outcomes. Column "qval" adjusts p-values to control for False Discovery Rates in using the sharpened q-value method from Anderson (2008) based on Benjamini et al. (2006). Column "index" indicates whether or not a variable was used in the construction of the family index in the first row, which is used to control for the Familywise Error Rate. Variables included in family indices are coded so better outcomes are larger, normalized, aggregated, and weighted according to the procedure described in Anderson (2008). All estimation, including the Family Index, uses cluster-by-type proportional weights so clusters are equally weighted.

Table 12: Daily Time Use

	Mean All	Beta Public	pval
Time Caring Others	3.08	-0.05	0.91
Time Hhld Chores	2.87	-0.53	0.19
Time Work	1.78	1.20	0.02
Time in Recreation	1.39	-0.29	0.20
Time in Edu	0.81	-0.51	0.16
Time Volunteer	0.46	0.29	0.10
Time Searching Wk	0.39	0.04	0.78
Time Political	0.22	0.07	0.61

Note: Table 12 reflects comparison of daily time use among public vs. referred candidates reported in response to questions phrased as follows: "[Yesterday] What amount of time, in hours, do you estimate you spent....?" The values in the column "Mean All" reflect the weighted mean values of variables with probability weights reflecting the inverse frequency of public or referred type candidates by cluster designed to equally weight each type-by-cluster cell in the estimation. The column "Beta Public" reports coefficients on a dummy variable for public candidates from OLS regressions with time spent on a given activity as the dependent variable and controls for a) the day of the week the respondent was surveyed and b) the sum of all of the hours the respondent reported across all activities. P-values were estimated using the cluster robust variance estimator (CRVE). Coefficients are similar in different specifications with cluster fixed effects not reported here, though standard errors are wider.

Table 13: Political Connectedness

	public	reference	dif	pval	qval	index
Politics Fam Index	0.99	0.91	0.08	0.17	–	-
Know Neigh Commit	0.62	0.46	0.16	0.16	0.23	Y
Know City Manager	0.55	0.35	0.20	0.04	0.13	Y
Know Mayor	0.94	0.88	0.06	0.13	0.22	Y
Know Congressman	0.58	0.40	0.18	0.07	0.16	Y
Know Senator	0.42	0.36	0.06	0.42	0.41	Y
Spoken w Neigh Commit	0.67	0.49	0.18	0.06	0.15	Y
Spoken w City Manager	0.43	0.29	0.14	0.14	0.23	Y
Spoken w Mayor	0.48	0.36	0.11	0.15	0.23	Y
Spoken w Congressman	0.43	0.17	0.26	0.00	0.03	Y
Spoken w Senator	0.19	0.12	0.08	0.15	0.23	Y
Related to Neigh Commit	0.10	0.07	0.03	0.68	0.51	Y
Related to City Manager	0.02	0.04	-0.02	0.43	0.41	Y
Related to Mayor	0.00	0.01	-0.01	0.32	0.36	Y
Related to Congressman	0.01	0.01	-0.01	0.32	0.36	Y
Related to Senator	0.00	0.01	-0.01	0.15	0.23	Y

Note: Comparison of public versus formally referred candidates in the 40 Public Advertisement arm clusters in which there is overlap. Column "pval" refers to p-values estimated using CRVE without correcting for multiple inference issues that may arise when comparing candidate types across outcomes. Column "qval" adjusts p-values to control for False Discovery Rates in using the sharpened q-value method from Anderson (2008) based on Benjamini et al. (2006). Column "index" indicates whether or not a variable was used in the construction of the family index in the first row, which is used to control for the Familywise Error Rate. Variables included in family indices are coded so better outcomes are larger, normalized, aggregated, and weighted according to the procedure described in Anderson (2008). All estimation, including the Family Index, uses cluster-by-type proportional weights so clusters are equally weighted.

Table 14: Motivation for Applying for CW Position

	public	reference	dif	pval	qval	index
Motivation Fam Index	1.08	0.95	0.13	0.02	–	-
Motive Fut Job	3.62	4.53	-0.91	0.00	0.01	Y
Motive Income	3.12	3.75	-0.63	0.03	0.10	Y
Motive Gov Connection	3.74	4.42	-0.68	0.00	0.03	Y
Motive Respect	3.93	4.28	-0.35	0.16	0.23	Y
Motive Support VP	4.15	4.40	-0.25	0.21	0.28	Y
Motive Bored	4.05	4.46	-0.41	0.03	0.11	Y
Motive Ex to Kids	4.88	4.88	0.00	0.98	0.64	Y
Motive ProSocial	4.93	4.92	0.02	0.65	0.50	Y
Motive Fut Work	0.89	0.97	-0.08	0.07	0.16	-
Motive Demonstrate Effectiveness	4.69	4.75	-0.05	0.64	0.50	-
Motive Signal to Employers	4.63	4.67	-0.04	0.77	0.52	-
Motive Fut Promotion	4.33	4.64	-0.31	0.03	0.11	-
Motive Benef Gov Connections	4.52	4.71	-0.18	0.16	0.23	-

Note: Comparison of public versus formally referred candidates in the 40 Public Advertisement arm clusters in which there is overlap. Column "pval" refers to p-values estimated using CRVE without correcting for multiple inference issues that may arise when comparing candidate types across outcomes. Column "qval" adjusts p-values to control for False Discovery Rates in using the sharpened q-value method from Anderson (2008) based on Benjamini et al. (2006). Column "index" indicates whether or not a variable was used in the construction of the family index in the first row, which is used to control for the Familywise Error Rate. Variables included in family indices are coded so better outcomes are larger, normalized, aggregated, and weighted according to the procedure described in Anderson (2008). All estimation, including the Family Index, uses cluster-by-type proportional weights so clusters are equally weighted.

Table 15: Cognitive and Personality

	public	reference	dif	pval	qval	index
Cognitive Fam Index	0.88	1.14	-0.26	0.02	–	-
Ravens	3.39	4.31	-0.92	0.02	0.09	Y
Ravens z-score	-0.22	0.16	-0.37	0.02	0.09	-
Exec Function	0.88	1.01	-0.13	0.27	0.33	Y
Reading Ability	1.92	1.74	0.19	0.18	0.25	-
Personality Fam Index	1.03	1.00	0.03	0.69	–	-
Big5 Index	0.24	0.07	0.17	0.13	0.22	-
Agreeableness	-0.01	0.19	-0.20	0.26	0.32	Y
Conscientiousness	0.36	0.10	0.26	0.05	0.14	Y
Expressiveness	0.29	-0.00	0.29	0.11	0.20	Y
Neuroticism	0.29	0.05	0.23	0.13	0.22	Y
Openness	0.26	-0.01	0.27	0.10	0.20	Y
Dictator game	0.46	0.44	0.03	0.55	0.47	Y
Stress	2.97	2.91	0.06	0.37	0.39	Y
Negative Expressivity	3.08	3.36	-0.28	0.06	0.15	Y
Positive Expressivity	5.87	5.69	0.18	0.26	0.32	Y
Expressivity Strength	4.39	4.78	-0.38	0.11	0.20	Y
Expressivity	4.45	4.61	-0.16	0.26	0.32	-
Intrinsic Aspirations	6.62	6.58	0.04	0.52	0.46	Y
Extrinsic Aspirations	4.94	4.94	-0.00	1.00	0.64	Y
Creativity	6.24	5.34	0.90	0.02	0.08	Y

Note: Comparison of public versus formally referred candidates in the 40 Public Advertisement arm clusters in which there is overlap. Column "pval" refers to p-values estimated using CRVE without correcting for multiple inference issues that may arise when comparing candidate types across outcomes. Column "qval" adjusts p-values to control for False Discovery Rates in using the sharpened q-value method from Anderson (2008) based on Benjamini et al. (2006). Column "index" indicates whether or not a variable was used in the construction of the family index in the first row, which is used to control for the Familywise Error Rate. Variables included in family indices are coded so better outcomes are larger, normalized, aggregated, and weighted according to the procedure described in Anderson (2008). All estimation, including the Family Index, uses cluster-by-type proportional weights so clusters are equally weighted.

Table 16: Ravens Test z-score with Controls

	Dependent variable:			
	Ravens z-score			
	(1)	(2)	(3)	(4)
Public	-0.374^{***}	-0.388^{**}	-0.251^{*}	-0.327^{**}
	(0.141)	(0.154)	(0.148)	(0.141)
Male		-0.101	-0.037	0.055
		(0.166)	(0.175)	(0.162)
Age		-0.007	-0.011	-0.014^{**}
		(0.007)	(0.007)	(0.007)
Education		0.036	0.022	-0.006
		(0.025)	(0.030)	(0.029)
ProSoli Benef		-0.156	-0.069	-0.230
		(0.146)	(0.160)	(0.150)
Time Raven				0.099^{***}
				(0.022)
Grade Raven				0.175^{**}
				(0.073)
Cluster FE?	N	N	Y	Y
Observations	195	193	193	191
R^2	0.035	0.064	0.364	0.468
		*p<0.1; **p<0.05; ***p<0.01		

Table 17: Impact of Public Recruitment on Job Acceptance and Training Attendance

	Accept Offer		Trained	
	(1)	(2)	(3)	(4)
Public	0.064	0.085	0.236*	0.308*
	(0.072)	(0.082)	(0.134)	(0.157)
Male		−0.033		−0.019
		(0.092)		(0.177)
Education		−0.024		−0.006
		(0.016)		(0.030)
ProSoli Benef		−0.105		−0.089
		(0.077)		(0.149)
Ravens		0.002		0.032
		(0.015)		(0.028)
Observations	48	48	48	48
R^2	0.017	0.111	0.063	0.101
		*p<0.1; **p<0.05; ***p<0.01		

Note: Table 17 compares the impact of having candidates accept job offers and participate in subsequent job trainings conditional on having been offered a job. The sample of candidates is limited to candidates offered jobs in 48 of the 53 clusters that had overlapping samples from public and referred candidates in which it was deemed there were qualified candidates of both types such that the type of candidate assigned to the cluster could be randomized. Only the first candidates offered the job in each cluster are included. Columns (1) and (2) reflect the coefficients from OLS regressions with a dummy variable for accepting the job offer with and without controls for a gender dummy, education level in grades, a dummy for being a ProSoli beneficiary, and the raw Ravens score from 0-11. Columns (3) and (4) reflect whether or not candidates attended the job training conditional on being offered the job (IIT impacts).

Table 18: Comparing References by Referrer Ranking

	Mean P1	Mean P2,P3,P4	dif	pval
Male	0.21	0.21	-0.00	0.90
Beneficiary	0.61	0.61	-0.00	0.95
Age	31.23	30.28	0.95	0.32
Education	12.19	11.65	0.54	0.07
Hhld Head	0.22	0.22	0.00	0.92
Insured	0.70	0.75	-0.05	0.26
Permanent Work	0.13	0.15	-0.02	0.65
Homemaker	0.77	0.73	0.04	0.32
Last Job Salary	6212.23	6830.69	-618.46	0.24
Hhld Income	11038.71	11099.58	-60.87	0.95
Food Insecurity	0.37	0.36	0.01	0.81
Neighbor Committee	0.21	0.16	0.04	0.24
Involved Comm Org	0.38	0.34	0.04	0.41
Campaign 2016	0.22	0.21	0.01	0.84
Leader	0.53	0.57	-0.04	0.42
Know City Manager	0.34	0.41	-0.08	0.08
Motive Fut Job	4.42	4.49	-0.07	0.51
Motive Gov Connection	4.38	4.51	-0.13	0.18
Motive Bored	4.29	4.33	-0.04	0.71
Ravens	4.03	3.95	0.08	0.74
Ravens z-score	0.04	0.01	0.03	0.74
Exec Function	1.09	0.97	0.11	0.15
Reading Ability	1.69	1.67	0.02	0.79
Big5 Index	0.01	-0.08	0.09	0.24
Creativity	5.32	4.91	0.41	0.05
N	167	287	–	–

Note: Table 18 reflects comparison of top ranked to lesser ranked candidates among the full sample of candidates that were formally referred and ranked by referrer. The values in the column “Mean P1” and “Mean P2,P3,P4” reflect the mean values for top ranked candidates and the pooled sample of all candidates that were not top ranked, respectively. The column “dif” reports differences between the first two columns of means. Column “pval” refers to p-values estimated using CRVE using the dependent variable of choice and a dummy for being the top ranked candidate. Observations are not weighted.

Table 19: Comparison of Public to Top Ranked Referrals in Ref-P1 Sample

	Public	Ref P1	dif	pval
Male	0.37	0.18	0.18	0.09
Beneficiary	0.35	0.62	-0.27	0.01
Age	37.98	32.02	5.97	0.01
Education	13.51	12.09	1.42	0.05
Hhld Head	0.51	0.18	0.33	0.00
Insured	0.77	0.73	0.04	0.72
Permanent Work	0.28	0.12	0.16	0.07
Homemaker	0.77	0.85	-0.08	0.44
Last Job Salary	12130.63	6309.62	5821.01	0.04
Hhld Income	17208.83	11128.64	6080.18	0.04
Food Insecurity	0.26	0.33	-0.07	0.46
Neighbor Committee	0.40	0.25	0.15	0.08
Involved Comm Org	0.51	0.30	0.21	0.09
Campaign 2016	0.33	0.12	0.22	0.03
Leader	0.83	0.62	0.21	0.06
Know City Manager	0.57	0.25	0.32	0.01
Motive Fut Job	3.65	4.50	-0.85	0.00
Motive Gov Connection	3.75	4.17	-0.41	0.20
Motive Bored	3.89	4.50	-0.61	0.03
Ravens	3.58	4.38	-0.80	0.17
Ravens z-score	-0.14	0.19	-0.33	0.17
Exec Function	0.92	1.03	-0.11	0.55
Reading Ability	1.90	1.60	0.30	0.13
Big5 Index	0.17	0.09	0.08	0.62
Creativity	6.07	5.08	0.99	0.03

Note: Comparison of public versus top ranked formally referred candidates in the 31 Public Advertisement arm clusters in which there is overlap. P-values estimated using CRVE.

Figure 1: Design for the Public Advertisement Intervention Posters and Fliers

Figure 2: Form Distributed to Collect Formal Referrals

FORMULARIO DE REFERENCIAS PARA ENLACE COMUNITARIO

Nombre del referente:	
Cédula:	
Teléfono:	
Cargo:	
Regional:	
Provincia:	

RECOMENDACIONES DE ENLACE COMUNITARIO

Código de núcleo nuevo: ___________________________

Ranking	**Nombre**	**Cédula**	**Teléfono**	**Dirección/Sector**	**Sabe leer y escribir**
1					
2					
3					
4					

Figure 3: Study Timeline

November 2015	*January 2016*	*January-March 2016*	*April-June 2016*	*June-July 2016*
Referral 1 1st round of referrals requested from ProSoli staff	**Referral 2** 2nd round of referrals requested from ProSoli staff	**Survey** Screening Survey	**Hiring Offers** Hiring Offers	**Training** Training

Figure 4: Cluster Coverage within Public and Private Referral Arms by Type

		Public Advertise Arm Clusters Covered	Percent of Intended	Referral Arm Clusters Covered	Percent of Intended
Public Candidates only covering:		3	3%	0	0%
Reference Candidates only covering:		20	19%	82	77%
Overlap Public & Reference:	**Formal Reference:**	**40**	**37%**	0	0%
	Informal Reference:	**13**	**12%**	0	0%
Total Covered:		76	70%	82	77%
Total Intended Covered:		108		107	

Figure 5: Relationship between Alternative Samples

		(1) REF-ALL: All Referrals	(2) REF-FORM: Formal Referral	(3) REF-RANK: Ranked Referrals	(4) REF-P1: Priority 1 Ranked Referrals
Reference Candidates (multiple treatment arms)	N:	738	526	454	167
	Clusters:	183	161	141	131
Public Candidates (Public Advert Arm only)	N:	127	127	127	127
	Clusters:	56	56	56	56
Overlap	Referral N:	NA	**84**	75	34
	Public N:	NA	**111**	94	67
	Clusters:	53	**40**	36	31

Figure 6: Map of Public Advertisement Distribution across Three Nearby Clusters

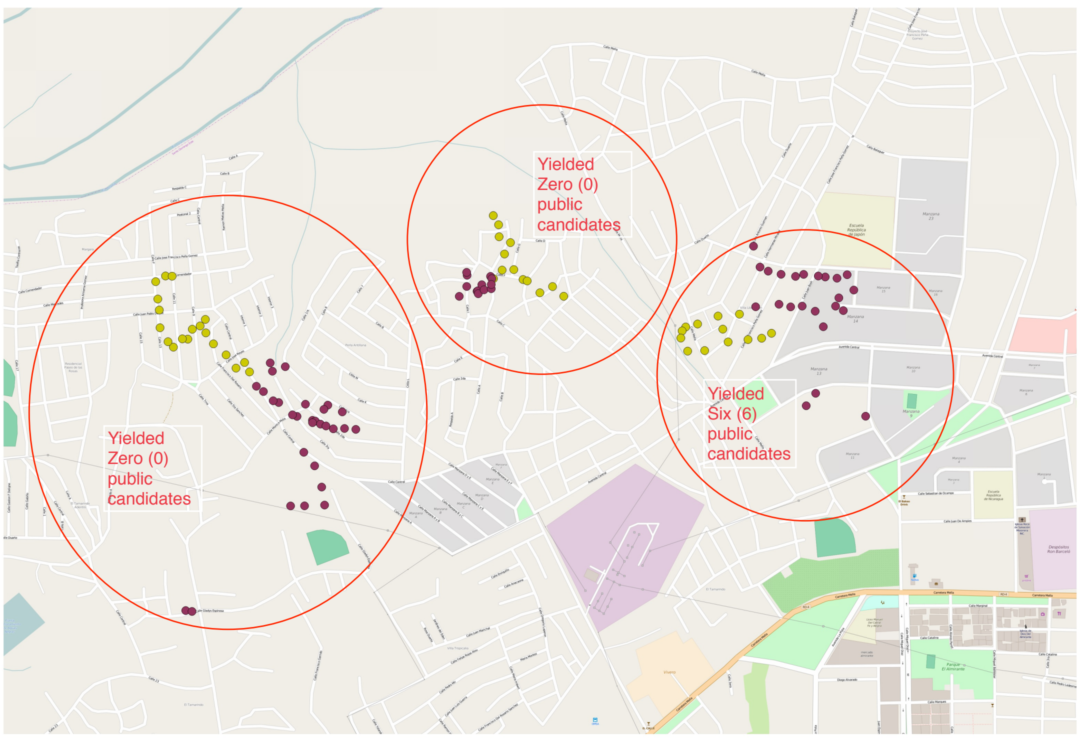


Figure 7: Candidate Age by Type

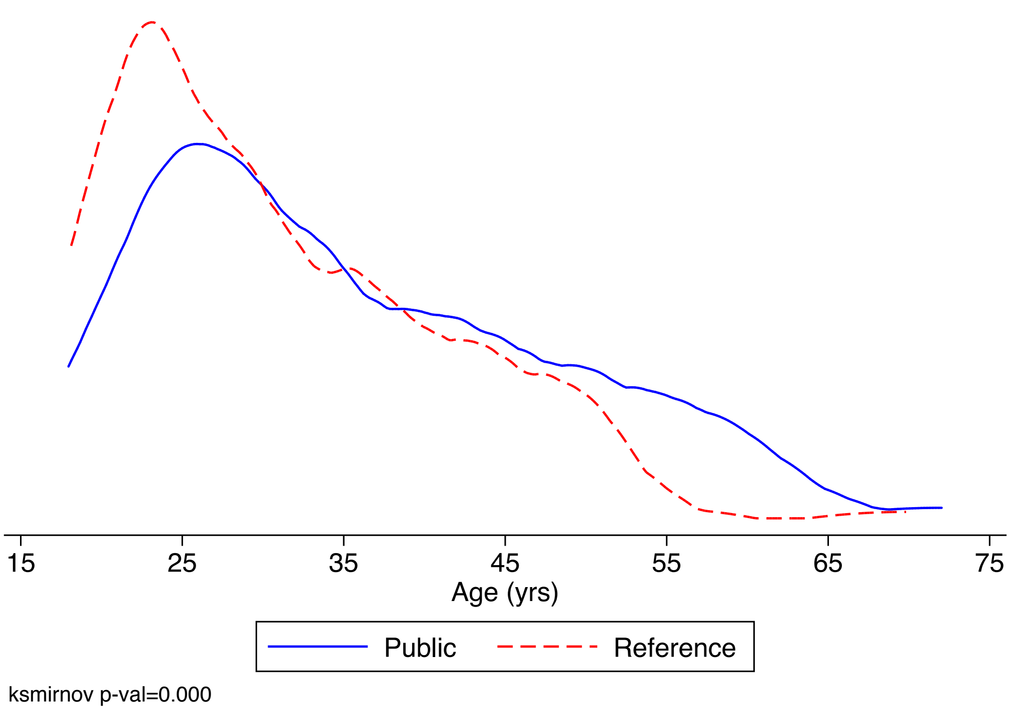


Figure 8: Candidate Education (Grades Completed) by Type

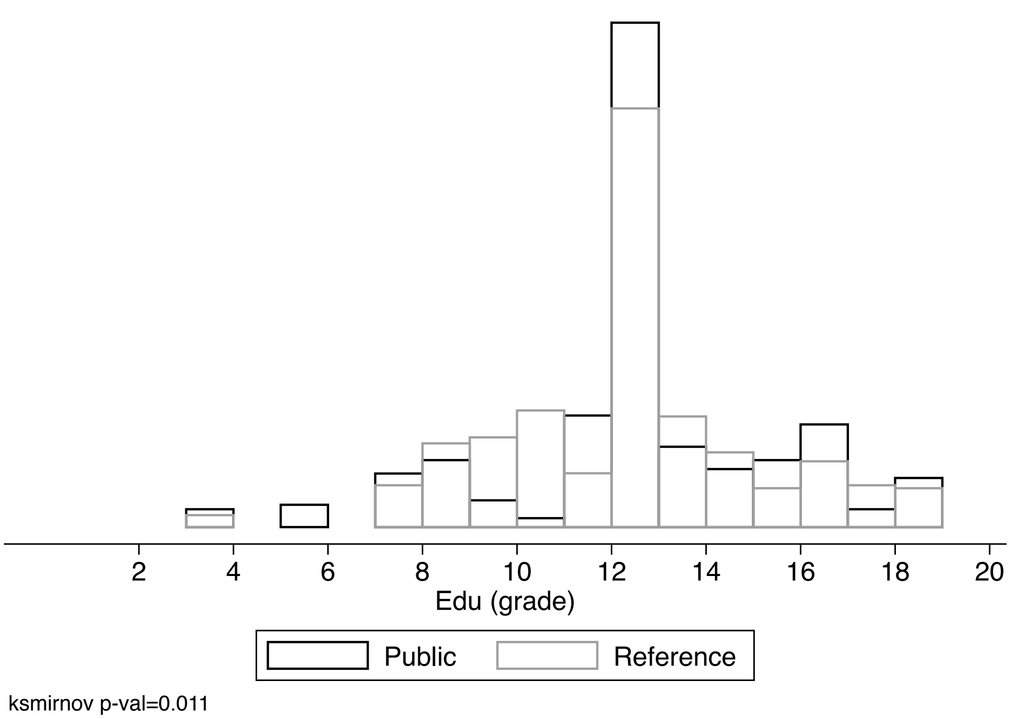


Figure 9: Candidate Monthly Salary at Most Recent Job by Type (log pesos)

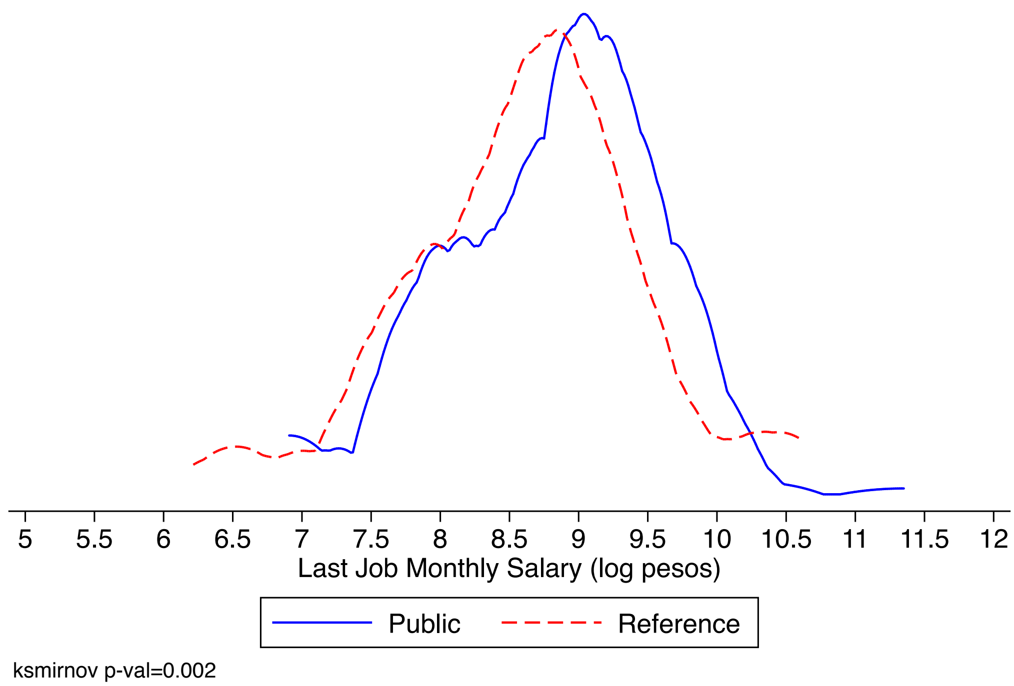


Figure 10: Candidate Individual Income Over the Last 7 Days by Type (log pesos)

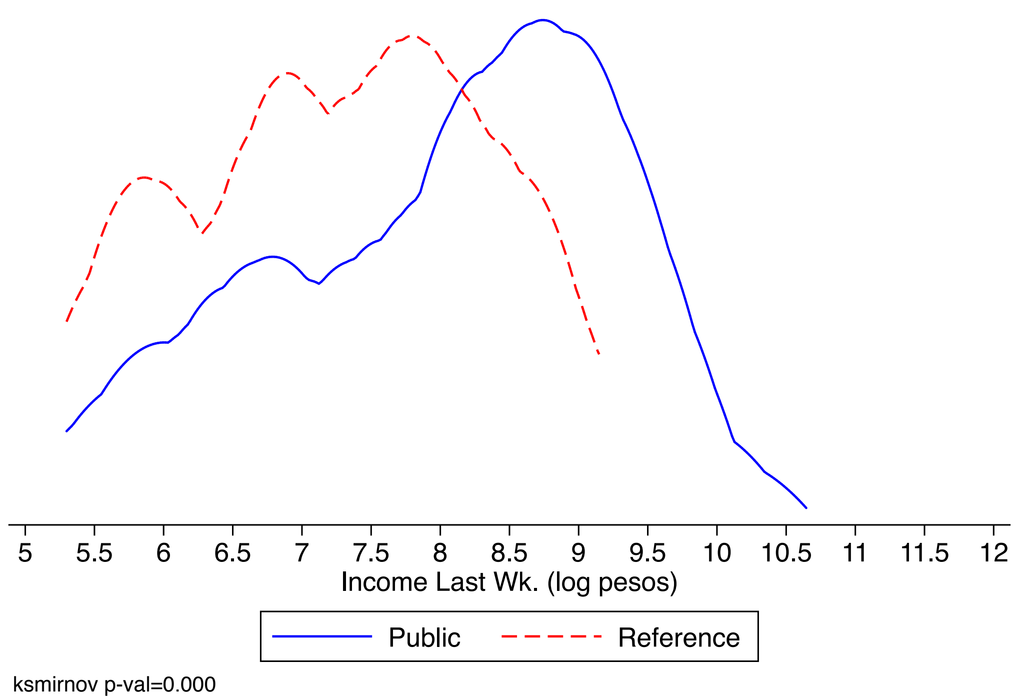


Figure 11: Candidate Household Durable Asset Index by Type

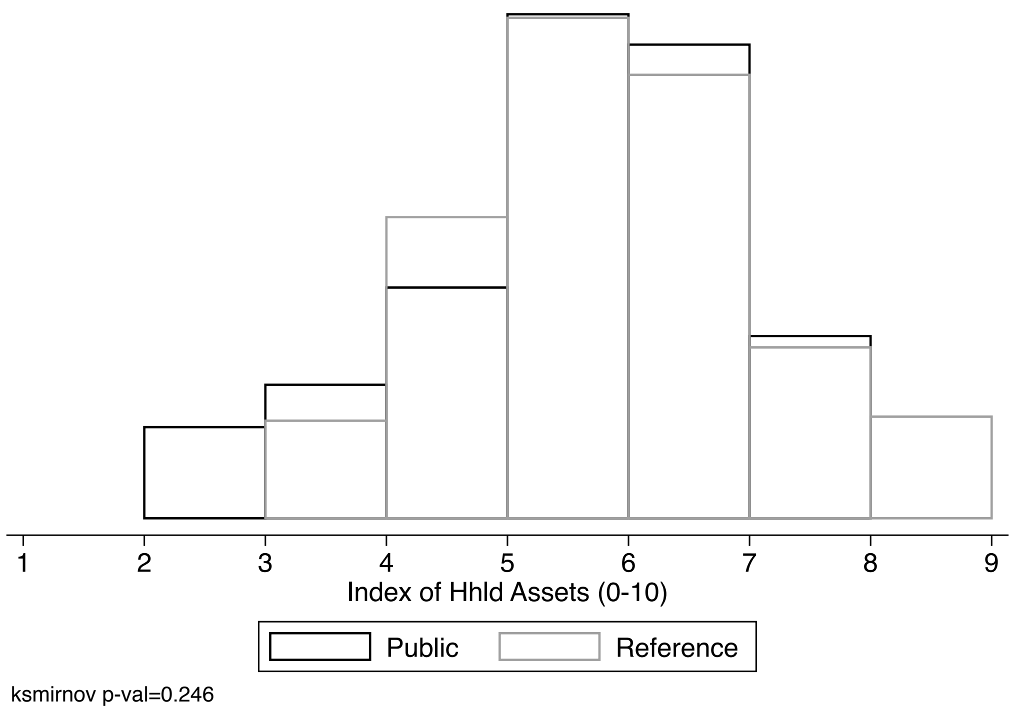


Figure 12: Candidate Household Income Last Month

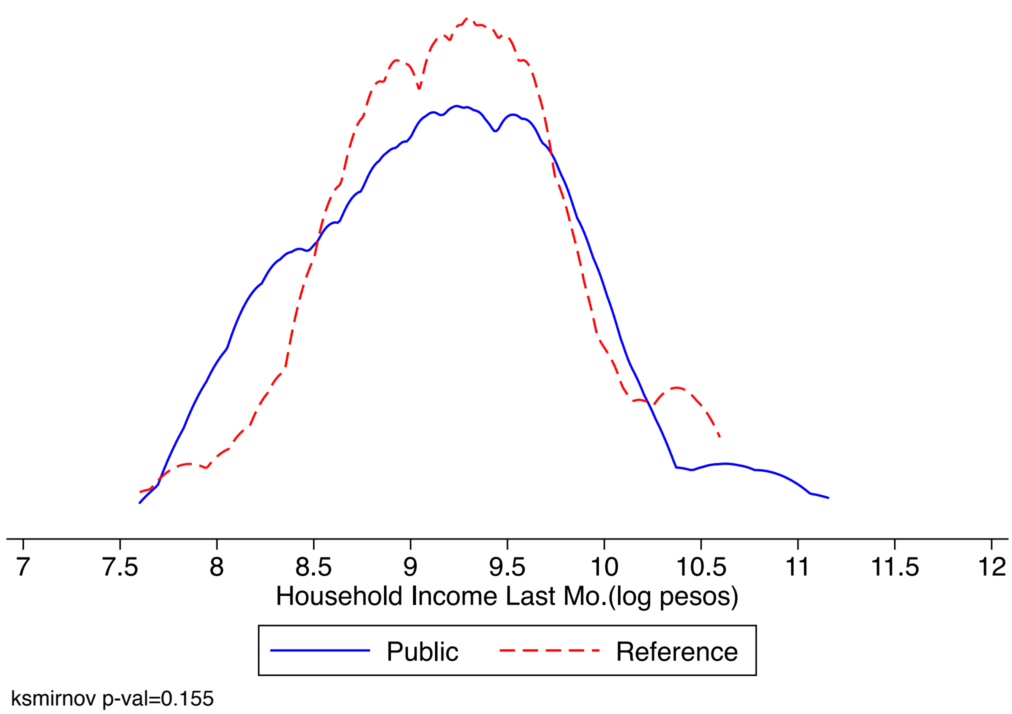
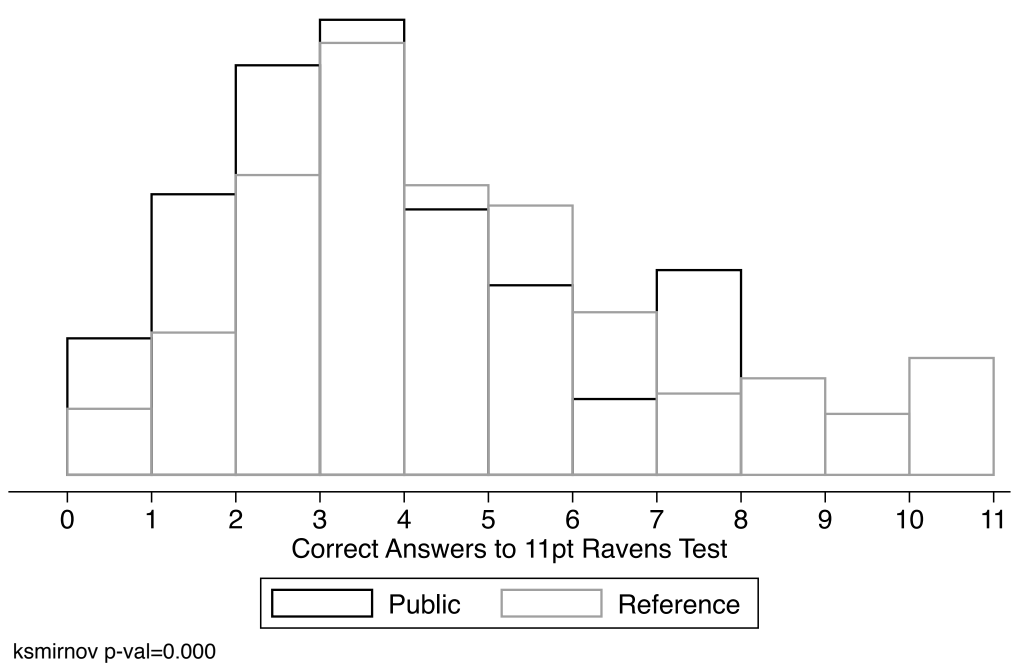


Figure 13: Candidate Scores on Raven's Cognitive Skills Test by Type



7 Appendix: Variable Dictionary

Short Description	**Long Description**
Single	Relationship status is single
Married	Relationship status is married
Separated	Relationship status is separated
Open Relationship	Relationship status is open relationship
Widow	Relationship status is widow/er
Age	Age in years
Male	Gender is male
Education	Highest grade of education completed. Started college but didn't complete coded as 12.5
Student	Self-identifies as a student when as directly "are you a student?"
Homemaker	Self-identifies as a homemaker when as directly "are you a homemaker
Children	Number of children
Hhld Size	Size of household
No. Earners	Number of earners in household
Beneficiary	ProSoli Beneficiary
Hhld Head	Considers self household head
Employed	Currently employed according to either general response about occupational status or specific response about income generating work in the last 7 days
Permanent Work	Primary occupational status is permanently works
Occasional Work	Primary occupational status is occassionally works
Temp Work	Primary occupational status is temporarily work
Unemployed	Primary occupational status is unemployed
Homemaker Occup	Primary occupational status is homemaker
Student Occup	Primary occupational status is student
Ever worked	Have ever worked
Last Job Perm	Prior job was a permanent job
Last Job Temp	Prior job was a temporary job
Last Job Ocass	Prior job was an occassional job
Last Job Salary	Prior job monthly salary in pesos
Time Since Job	Time since employed in last job in days
2nd Last Job Salary	Second to last job monthly salary in pesos
Worked last wk	Worked last week
Last Wk Income	Personal income earned in last 7 days

Short Description	Long Description
Asset Index	Index of durable assets including stove, refrigerator, tv, satellite tv, washer, motorcycle, car, air conditioner, computer, mobile phone
Asset Index Short	Index of durable assets including stove, refrigerator, tv, washer, car, air conditioner, computer
Dirt Road	Road to home is not paved
Have Any Savings	Have any savings
Bank Account	Have a bank account
Loan Last 12 mo	Have obtained a loan in last 12 months
Purchase Credit	Have purchased food on credit in last 12 months
Hhld Income	Last month household income
Income Increasing	Household income increased over the past 12 months
Income Decreasing	Household income decreased over the past 12 months
Remitances	Household receives remitances
Depend on Remit	Household depends on remittances somewhat or a lot
Renter	Pay rent to live in home
Rental Value	Monthly rental cost
Home value	Monthly rental cost or rental value of home if were to rent
Insured	Insured in the public health insurance
Food Security	Above median response on 4 category question about frequency of missing out on eating
Skipped Meal	Above median response on 4 category question about frequency of missing meals
Skipped Meds	Above median response on 4 category question about frequency of missing medication for reasons of cost
Prime worker	Primary worker in the household
Want More Wk	Interested in workering more to augment income
Job search	Has looked for additional work in the past 4 weeks
Time Work	Hours in last day working
Time Searching Wk	Hours in last day searching for work
Time in Edu	Hours in last day studying or attending school
Time Hhld Chores	Hours in last day doing household chores
Time Caring Others	Hours in last day caring for others such as children or elderly
Time Volunteer	Hours in last day volunteering
Time Political	Hours in last day involved in political activity
Time in Recreation	Hours in last day spent in recreation
Time Total	Hours in last day reported dedicated to any activity

Short Description	Long Description
Neighbor Committee	Member of the neighborhood committee
Attend Neigh Commit	Attended the neighborhood committee in the past 12 months
Attend City Council	Attended the city council in the past 12 months
Petition	Involved in any sort of petition in the last 12 months
Involved Relig Org	Above median response on 4 category question about frequency of attending a religious organization
Involved School Org	Above median response on 4 category question about frequency of attending a organization associated with a school or college
Involved Comm Org	Above median response on 4 category question about frequency of attending a community organization
Involved Political Org	Above median response on 4 category question about frequency of attending political organization
Involved Womens Org	Above median response on 4 category question about frequency of attending a women's organization
Strike	Participated in strike in the last 12 months
Registered Vote	Registered to Vote
Voted in 2012	Voted in 2012 presidential campaign
Voted in 2016	Planned to vote in 2016 presidential campaign
Campaign 2012	Worked for 2012 presidential campaign
Campaign 2016	Planned to work for 2016 presidential campaign
Leader	Consider self a local leader currently
Future Leader	Expect self to be a local leader within next 5 years
Know Neigh Commit	Know a member of their Neighborhood Committee
Know City Manager	Know City Manager
Know Mayor	Know City Mayor
Know Congressman	Know congressman
Know Senator	Know Senator
Spoken w Neigh Commit	Have ever spoken w a member of their Neighborhood Committee
Spoken w City Manager	Have ever spoken w City Manager
Spoken w Mayor	Have ever spoken w City Mayor
Spoken w Congressman	Have ever spoken w Congressman
Spoken w Senator	Have ever spoken w Senator
Related to Neigh Commit	Are related to a member of Neighborhood Committee
Related to City Manager	Are related to City Manager
Related to Mayor	Are related to City Mayor
Related to Congressman	Are related to Congressman
Related to Senator	Are related to Senator

Short Description	Long Description
Motive Fut Job	Above median motivation to work to obtain a job in the future
Motive Income	Above median motivation to work to make money
Motive Gov Connection	Above median motivation to work to forge connections with the government
Motive Respect	Above median motivation to work to earn the respect of neighbors
Motive Support VP	Above median motivation to work to support the Vice President
Motive Bored	Above median motivation to work to reduce boredom
Motive Ex to Kids	Above median motivation to work to be a positive image for children
Motive ProSocial	Above median motivation to work to improve lives of families in community
Motive Fut Work	Above median motivation to work because expect volunteering will lead to future work
Motive Demonstrate Effectiveness	Above median belief job will lead to future work because provides evidence that I am an effective employee
Motive Signal to Employers	Above median belief job will lead to future work because sends employers a positive signal
Motive Fut Promotion	Above median belief job will lead to future work because I can be promoted by the government
Motive Benef Gov Connections	Above median belief job will lead to future work because government connections will lead to work in other sectors
Reading Ability	Score on reading comprehension test (0-3). Reading excerpt derived and edited from community worker manual.
Exec Function	Score on memory of reading comprehension excerpt (0-2)
Ravens	Ravens Progressive Matrices Set I cognitive skills test raw score from 11 total potential points
Ravens-z	Ravens cognitive skills matrices test z-score
Big5 Index	Big5 z-score from Big Five Inventory (BFI) developed by John (1990).
Agreeableness	Agreeableness scale of Big5
Conscientiousness	Conscientiousness scale of Big5
Extroversion	Extroversion scale of Big5
Neuroticism	Neuroticism scale of Big5 reverse coded
Openess	Openess scale of Big5
Dictator game	Average percentage given across 3 trials of dictator game with 200, 150, 100 pesos
Stress	8 question stress scale derived from Stress Appraisal Measure (SAM) from Peacock and Wong (1990)
Negative Expressivity	Negative Expressivity scale from James and John (1997)
Positive Expressivity	Positive Expressivity scale from James and John (1997)
Expressivity Strength	Expressivity Strength scale from James and John (1997)
Expressivity	Expressivity combined scale
Intrinsic Aspirations	Index of Instrinsic Aspirations from Kasser and Ryan (1996)
Extrinsic Aspirations	Index of Extrinisic Aspirations from Kasser and Ryan (1996)
Creativity	Creativity scale that counts list of alternative uses of a concrete block mentioned in 45 seconds

Evidence on the Impacts of Short-term Employment Generation Projects to Improve Road Infrastructure in Rural Nicaragua

Seth Garz and Elizaveta Perova

April 2014

Abstract

This paper seeks to evaluate impacts of a World Bank funded road improvement and employment generation intervention in Nicaragua. Despite data limitations and the short time interval between project implementation and observed outcomes, we find strong evidence that the World Bank's Fourth Roads Rehabilitation and Maintenance Project (WB4) fulfilled its primary goal of improving road infrastructure and suggestive evidence of select economic and social impacts. In terms of road improvements, we find that WB4 projects are associated with a 16.4% increase in the likelihood that community access roads were paved. WB4 projects are also associated with a 7.4% increase in the likelihood that working age individuals worked most recently as laborers. In locations where projects concluded prior to the observed outcomes, projects are associated with an 8.5% decline in respiratory illness for men. Projects are also associated with a slight 2.0% statistically significant increase in the likelihood that women of childbearing age will access a medical clinic for primary care or medicine. Notably, we do not observe impacts on likelihood of employment or poverty. We interpret this collective evidence as suggesting that WB4 projects do have economic and social impacts on local communities, but that more precise data collection specifically tailored to the intervention with a longer evaluation time horizon may be required to acutely observe impacts.

1 Introduction

This paper seeks to evaluate impacts of the Fourth Roads Rehabilitation and Maintenance Project funded by the World Bank (WB4) to improve rural roads and provide short-term employment ("productive inclusion") implemented by Nicaragua's Ministry of Transport and Infrastructure (MTI) in 2006-2010. The project development objectives noted in the *Project Appraisal Document* of relevance to this evaluation include i) improving access of the rural population to markets and social and administrative services through improvements to road infrastructure and ii) supporting the generation of short-term employment opportunities for the rural population (World Bank, 2011). Beyond evaluating specific project goals, however, this analysis may be viewed within the context of the broader literature on the development impacts of road building, public works targeting, and anti-poverty public works employment generation programs ("workfare").

There are a number of reasons why road improvements may stimulate development, but be underprovided by communities and governments. Roads are generally non-excludable non-rival public goods in that individuals cannot be excluded from using roads and any given individual's use does not reduce the ability of others to use the road. Organizing to build roads presents a classic collective action problem that is traditionally solved by government operating on behalf of community residents. In the low-income country context, however, governments may lack the bureaucratic or technical capacity to organize and implement road improvements. Furthermore, political dynamics, local rent-seeking, and unintended capture of project funds and materials in areas with poor government transparency, where citizens are poorly educated, etc. may stymie even the best designed government road improvement projects.

When adequately provided and maintained, roads may operate through a number of key channels to promote economic and social development. Better roads may improve access to markets for rural residents, improving residents' access to off-farm employment, reducing their commodity prices through greater competition among suppliers, and broadening sources of demand for their goods and services. Casaburi et. al (2013) argue that road building in rural Sierra Leone reduced search frictions in commodity markets, driving down the cost of local food staples. Mu and van de Walle (2011) find that rural road building in Vietnam induces local market development even in particularly poor areas that may otherwise be systematically constrained from developing market institutions. Road improvements may also improve education outcomes. Better roads may encourage school attendance by making often distant rural schools more accessible with all the requisite benefits for cost, safety, etc.. Additionally, economic development due to road improvements may stimulate demand for education as labor market opportunities broaden with greater access to markets. Health benefits may accrue through improved access to health clinics and medicine supply, particularly for maternal and child health, and through reduction of respiratory illness aggravated by ambient dust from traffic on unpaved roads.

Given the multiplicity of channels through which road improvements may contribute to development, it should come as no surprise that theory is ambiguous about how short-term impacts should compare to long-term impacts. It may take time for short-term employment and income impacts to translate into greater demand for education, for example. Alternatively, immediate increases in attendance may decline if they are not rewarded in the labor market over the long-term. Short-term impacts may snowball if people escape poverty traps and enter a virtuous development cycle. Constructive investments of short-term increases in income or short-term substitution away from agriculture towards employment in service occupations, for example, may exhibit increasing returns to scale as rural residents overcome credit constraints or make fixed investments in small businesses. Alternatively, short-term impacts may taper as newly accessible market reach new competitive equilibria. Farmers may initially sell items for better prices in newly accessible markets until new supply drives down such prices. Local entrepreneurship may briefly boom until outside competition arrives. Notably, two recent studies that compare short-term to long-term effects of road improvement projects in Vietnam and Bangladesh consider outcomes after approximately four years "short-term" and outcomes after approximately six and eight years respectively "long-term" (Khandker and Koolwal, 2009; Mu and van de Walle, 2011). In comparison, the targeted regions in our data had been exposed to road improvement projects between six days and 3.5 years.

In addition to variation in impacts of road improvements over time, variation by other characteristics, such as wealth/poverty and gender, are of acute interest. Mu and van de Walle (2011) take this heterogeneity seriously in contemplating whether road improvements should target the poorest communities or the less extreme poor. They argue that the poorest communities may have basic attributes, such as mountainous geography, etc., that constrain the market development road improvements are intended to catalyze whereas less poor communities may already have key market institutions necessary for further economic development. The authors find the largest impacts on the poorest communities while admitting that characteristics of such communities, such as illiteracy and presence of ethnic minorities historically the subject of discrimination, do constrain potential impact.

The cross-sectional nature of our data is not well suited to explore heterogeneity by wealth; however, we are able to look for heterogeneous impacts by time invariant characteristics, such as gender. Consistent with Mu and van de Walle (2011) and Khandker and Koolwal (2011), for example, we find short-term impacts on education for primary school age girls, but not boys. The robustness of this finding across studies, including our own, is impressive and supports a variety of explanations that warrant further study. Households may expect that girls will benefit most in the labor market with gains from education, may be less dependent on girls for on-farm labor, or may have otherwise placed insufficient value on girls education to warrant the cost of transportation. The characteristics of new employment opportunities may be key. Research has shown that female-focused job recruitment for skilled call center jobs in India increased investment in girls but not boys education among rural households despite the fact that job opportunities were, in principal, available to both genders (Jensen, 2009). Alternative research has demonstrated that expansion of low-skilled manufacturing in Mexico actually decreased school attendance among both boys and girls, though differentially in different cohorts depending on the gender preference among new manufacturing employers at any given time (Atkin, 2009). Further investigation to explain precisely why roads impact matriculation in primary school among girls, in particular, across geographies as diverse as Bangladesh, Vietnam, and Nicaragua could help target key potential policy levers for improving girls education, empowerment, and wellbeing.

A variety of studies have found road improvements decrease poverty, which is consistent with evidence on market creation, changes in occupation choice, increases in schooling, and health impacts (Khandker, 2009; Gibson, 2003; Dercon, 2009). Somewhat less intuitive than these poverty reduction channels, credit constraints may be relaxed as a result of road building. For example, a randomized street paving experiment conducted in Mexico found that road improvements increased the value of adjacent property. This property value increase fueled use of collateralized credit and subsequently led to increased consumption of durable goods albeit in an urban context in which mortgage lenders may be more responsive to changes in neighborhood public goods than in the countryside (Gonzalez-Novarro, 2012).

The short-term employment impacts of our labor-intensive intervention may be a more likely source of liquidity for households facing credit constraints than collateralized debt. In the case of the WB4 intervention, the workfare programs are designed to be temporary. Such interventions are unlikely to function as true insurance against negative income shocks as is the case of unemployment insurance in higher income countries or the rural work guarantee in India. Nevertheless, they may provide positive income shocks that propel households out of poverty traps. There is an extensive literature on public works projects as safety nets, which is thoroughly surveyed in Subbarao, et. al (2013) and we will have little to say in our data about the income effects of our intervention.

We do not observe impacts on likelihood of employment or poverty, however, WB4 projects are also associated with a 7.4% increase in the likelihood that working age individuals worked most recently as laborers. We also find that WB4 projects are associated with a 16.4% increase in the likelihood that community access roads were paved. In locations where projects concluded prior to the observed outcomes, projects are associated with an 8.5% decline in respiratory illness for men. Projects are also associated with a slight 2.0% statistically significant increase in the likelihood that women of childbearing age will access a medical clinic for primary care or medicine. We interpret this collective evidence as suggesting that WB4 projects do have economic and social impacts on local communities, but that more precise data collection specifically tailored to the intervention with a longer evaluation time horizon may be required to acutely observe impacts.

2 Description of the Intervention

WB4 road improvements targeted existing roads selected from the pre-existing *Sector wide Medium Term Expenditure Framework* (MGSMP), the document that guided MTI funding priorities.[1] The project selection process is worth discussing in some detail as it is key in guiding our research design and econometric methods. In estimating the causal impacts of an intervention, one must be conscious of the reasons why some regions were targeted to receive interventions and others overlooked. To construct a valid counterfactual "control" group to which one can compare the targeted "treated" communities, it is important to control for the characteristics upon which communities were chosen for the intervention. In lieu of a randomized experiment or quasi-random natural experiment, researchers often explicitly control for project selection rules. A variety of rigorously identified work has investigated impacts of large transport infrastructure (e.g. Faber (2012) and Banerjee, Duflo and Qian (2012)). To our knowledge, however, only work by Casaburi et. al (2013), focused on rural Sierra Leone

[1] The document is known as the *Marco de Gasto Sectorial de Mediano Plazo* (MGSMP).

[2] The actual Spanish version of the question reads: "Cual es el principal via de acceso para

and Gonzalez-Navarro et. al (2012), focused on urban Mexico, claim to exploit randomization or quasi-random natural experiments in research related to smaller road improvement interventions. However, these works in particular, which are focused on rural West Africa and urban Mexico, may not be directly comparable to rural Nicaragua where the interventions we examine occur.

In the case of WB4, community targeting for the interventions considered a wide variety of characteristics for which we do not have appropriate data and gave MTI substantial discretion even after considering regional characteristics. Our analytical challenge of controlling for "selection bias" / "project placement bias" is not uncommon and haunts prior road improvement evaluations. The first stage at which we may be concerned that selection decisions may undermine the comparability of targeted vs. non-targeted communities is in the assignment of priority roads to a short-list in MTI's MGSMP planning document. We do not have much information on how roads made it into this document. A second stage of selection was in deciding which roads among those in the MGSMP document should be targeted for interventions with limited WB4 funds. At this stage, MTI first assigned scores to regions according to five criteria (productivity, population, road connectivity, poverty, and natural disaster vulnerability) and then selected a final list among the short-listed regions according to demand and taking account of equitable geographic distribution. For this reason, as explained in detail below, our preferred sample frame discards observations from departments (an administrative unit approximately the equivalent of U.S. states and below the region level) in which there were no WB4 projects.

Solely accounting for this fixed rule of prioritizing certain regions may be insufficient to effectively create a valid counterfactual group given the discretion provided to MTI within the short-list of roads. In lieu of pure fixed selection rules, researchers often attempt to match non-treated communities to treated communities according to pre-intervention characteristics. Unfortunately we do not have confidence that our data are representative at the *municipio* level, the most precise geographic level in our data, which we use to assign administrative data on interventions to survey data. Municipios are the administrative unit approximately equivalent to U.S. metropolitan regions or counties that collectively cover the entirety of Nicaragua. Collapsing household data into municipio aggregates to compare treated and non-treated municipios is, therefore, a potentially dubious exercise. Thus, we pursue a research design that allows for time invariant systematic differences between treated and control units, but requires that treated and control units exhibit parallel trends in their outcomes of interest prior to the intervention – a claim we review and describe in detail below.

The small size and geographically varied character of the road improvement projects within the WB4 intervention portfolio, are substantially different from traditional road improvement development projects that often include long stretches of asphalt road and bridge construction. In some sense, these projects more closely resembled traditional employment generating workfare projects. In fact, unlike other road improvement projects in Nicaragua funded by other development agencies during the same period, WB4 projects were designed as both road improvement and short-term employment generating activities. Specifically, WB4 projects focused on employing low-skilled unemployed rural residents organized in community construction teams known by their Spanish acronym MCAs (*Módulos Comunitarios de Adoquinado* or "community cobblestone groups," in translation) to improve short stretches of roads mostly by laying cobblestones.

MTI contracted directly with local mayors who were then in charge of recruiting the appropriate MCA managers and engineers according to MTI's precise guidelines. MTI official guidelines required recruiting laborers of adult age from local unemployment rolls from communities near project sites. The primary work of MCAs was to improve roadways by laying cobblestones ("*adoquines")* – a particularly low skill and labor intensive road construction technology that has demonstrated physical and cost advantages over other road surfacing technologies. MCA projects often briefly employed professional contractors who used heavy equipment to grade roadways at the beginning of MCA projects, for example, before MCAs could lay *adoquines*. MTI required specific accounting and procurement processes of MCAs and MTI staff claim that project materials were closely monitored to prevent leakage.

The process of recruiting laborers to participate in MCAs, introduces a third layer of potential selection bias into our sample that we do not adequately observe and that warrants further investigation. Despite MTIs rigorous guidelines, we do not have data to verify who actually worked on MCA projects and how they were selected. Even within the guidelines, local officials had substantial leeway to recruit MCA participants, suggesting there was likely non-random selection in project employment. To the extent political affiliation, familial ties to local leaders, indigenous ethnicity, and other characteristics unobservable in our data differentially influenced recruitment of MCA laborers and effects of MCA employment are heterogeneous along those characteristics of selection, we may be concerned about our ability to identify causal effects. Even if local leaders abided by the rule to recruit from the unemployment rolls, to the extent determinants of unemployment vary across municipios, we may be concerned about causal interpretations. For these reasons, future data gathering efforts to identify individual participation into MCAs would allow for more precise and nuanced analysis that could potentially shed light on trends and patterns on selection. Given that delegating organizing public works projects to local leaders is likely a common practice around the world, understanding patterns of selection in local recruitment in this setting may be useful both for improving targeting of the most needy households within the Nicaraguan context and improving design of similar interventions elsewhere.

Figures 1 & 2 depict maps of both WB4 and other road and bridge improvement projects implemented during 2005-2009 and identify WB4 targeted municipios. 40 projects were commissioned under WB4 targeting 30 of the 153 different Nicaraguan municipios. 23 projects commenced prior to the "endline" 2009 EMNV survey and projects in five targeted municipios captured in the EMNV concluded prior to the endline. To foreshadow the relevance of project commencement versus completion, we will find that impacts on primary schooling, in particular, appear to depend on whether or not projects were ongoing or had concluded by the time of the final survey. Figure 3 provides the distribution of the maximum "exposure" to the road building intervention among treated municipios in terms of number of days from project start to the endline, length of road improved, and budget. On the household level, exposure time ranges from 6 to 1243 days with a median of 305 days, suggesting the majority of households from treated communities in our data had only been exposed to the interventions for a relatively short period of time - well within a year of the final EMNV survey date. The lengths of the road segments by municipio reflect the sum length of all projects within a municipio that commenced before the endline survey and range from 1 km to 41 km with a median length of 19 km. The average project cost is 32 million Nicaraguan cordobas.

In addition to timing, length, and cost, projects also vary in the type of labor employed. Of the 40 total commissioned projects, 27 projects targeting 25 municipios employed rural residents in MCAs and 13 projects targeting nine municipios exclusively employed professional contractors (Empresas). According to MTI and World Bank staff, professional contractors were retained for the more technically challenging projects and often brought with them professional laborers from outside of the targeted communities. The map in Figure 1 depicts municipios targeted by MCA projects, Empresa projects, or both, and Table 1 summarizes how MCA vs. Empresa projects match to our municipio units in the data. For example, of the 25 unique municipios targeted by WB4 MCA projects, only 20 show up in the raw EMNV data and only 15 are represented in the 2001-2005-2009 municipio panel we construct. Our preferred "Base" sample specification, which drops all control municipios in administrative departments (similar to U.S. states) that do not have WB4 projects, includes 15 municipios targeted by MCA projects and 8 municipios targeted by Empresa projects, three of which overlap. The map in Figure 2 depicts the departments that are dropped from the sample. We disaggregate MCA versus Empresa treatments in the analysis to clean MCA projects of what appears to be significant selection bias in the Empresa projects and acknowledge that MCA and Empresa projects had different labor market goals.

Non-WB4 Projects

WB4 projects coincided with other road improvement and bridge improvement projects implemented by MTI and funded by other development agencies. MTI has provided data on the municipios targeted by all such projects between the last two rounds of EMNV survey data in 2005 and 2009, the period during which WB4 projects were implemented. Seven municipios were targeted for both WB4 and non-WB4 road improvement projects. The subsequent analysis will control for the presence of other non-WB4 road improvement projects; however, we are constrained by the absence of data on the specific dates of project commencement and completion for non-WB4 projects.

3 Data Description and Summary Statistics

Our data include administrative data on both WB4 and non-WB4 road improvement projects provided by MTI and three rounds of a publicly available household survey from 2001, 2005, and 2009, which we match by municipio identifiers. Taken together, these data provide a repeated cross-section of different individuals and households within similar municipios across three years. All of our projects took place during the 2005-2009 period, thus 2001-2005 may be considered our pre-intervention period.

The household survey we use, the *Encuesta Nacional de Hogares sobre Medicion de Nivel de Vida* (EMNV), was designed according to the World Bank's Living Standard Measurement Study (LSMS) methodology and conducted by the Nicaraguan National Statistics and Census Institute. The 2001 EMNV includes 4676 households and was designed to be representative on the regional level based on the 1995 census. The 2005 EMNV included households from the 2001 EMNV, allowing for a household panel in the pre-intervention period, and added additional households based on the 2005 census to be representative on the department level with 7871 households. The 2009 EMNV dropped the prior panel households and resampled among 2005 census segments based on a similar sampling frame as the 2005 EMNV with a total of 7520 households.

The EMNV presents some key constraints. The 2009 EMNV survey, for example, did not consider the WB4 projects and, therefore, does not identify household level participation in MCAs. Questions and categories of answers sometimes change across survey rounds, requiring subjective data harmonization. Geographic precision for households is limited to the municipio level, a limiting factor given that our WB4 administrative data on road projects is more geographically precise. The 2009 EMNV dropped the household panel from prior rounds, so we cannot control for unobservable time invariant household characteristics with household fixed effects. We also cannot merely collapse and aggregate our data to create a balanced municipio panel both because the data are not designed to be representative at the municipio level and because we would be left with very few treated observations. Treating each round as a cross-section and assigning the treatment at the municipio level, we construct a repeated cross-section with rich individual and household variables and a number of outcomes of interest that are observed across the pre- and post-intervention period despite real data limitations.

Table 3 reveals summary statistics of household and individual controls for our last pre-intervention round in 2005 between treated and control municipios across three alternative sample specifications, which are described in greater detail below. Our preferred "base" specification appears to be the most balanced between treated and control groups. In the base specification, household size averaged seven people and differences are not statistically different among treatment vs. control groups. Nevertheless, treatment groups do appear somewhat better off in other meaningful ways. Households in treated municipios have more rooms, greater cell phone use, are more likely to enjoy indoor plumbing, and are more likely to access to the electricity grid, which explains their reduced use of biomass for cooking. They are also slightly older with an average age of 26.5 years.

Table 4 provides summary statistics of household and individual outcomes of interest by round between treatment and control groups for the preferred base sample specification. Mere comparison of differences in averages between groups is insufficient to infer causal impacts of project treatments for which we apply more sophisticated econometric methods. Nevertheless, interesting information can be gleaned from these basic summary statistics, which are also useful references for benchmarking the magnitude of implied project impacts derived from the later regression analysis.

As in the subsequent analysis, outcome variables are summarized within the appropriate demographic groups. Household outcomes assume one unit of observation per household regardless of the number of household members and include: whether the community's access road is paved, whether the household is poor or extremely poor, and whether the home has concrete walls and/or cement floors. Individual labor market outcomes are reported only for individuals ten years or older and include: a) whether or not an individual worked in the past year, past week, or ever before, and b) whether the individual's most recent primary source of employment was as a laborer/employee or as self-employment/entrepreneurship conditional on having worked during the past week. Health outcomes are limited to incidence of respiratory illness in the past month and include all respondents. The "Medical Visit" indicator is a proxy for preventive care and is limited to women of approximate childbearing age between 15-45. The variable captures whether a woman has visited a clinic for medical consultation or medicine in the past month despite reporting otherwise good health. "Matriculated" is an indicator for whether individuals age 7-18 have matriculated in the formal school system in the past year and "School too far" is an indicator for whether individuals of school age did not matriculate because of distance to school.

4 Identification and Counterfactuals to Treated Groups

Our principal empirical strategy applies a difference-in-difference (DD) design on repeated cross-sectional data, matching municipio level administrative data on WB4 project rollout to the household and individual level EMNV data from 2001, 2005, and 2009. We assign the "treatment" on the municipio level as it is the most precise geographic unit available in the EMNV data. In reality, only specific communities within municipios were targeted to participate in the MCAs.

Conventional DD designs allow unbiased estimates provided project targeting/selection bias is constant over time, which may be a strong assumption in our case. WB4 projects were selected from a pre-existing list of roads prioritized for improvements, introducing two layers of potential selection bias into our treatment assignment. Additionally, mayors and project leaders within targeted municipios had considerable flexibility in recruiting households and individuals to participate in the MCAs. Furthermore, WB4 projects coincided with other road improvement initiatives funded by other donor agencies, potentially contaminating the non-treated group in a selected manner.

To control for potential sources of bias that may vary over time, we control for observable characteristics of treatment and counterfactual municipios, including the presence of other non-WB4 road improvement projects. We also compare estimates with and without controls for observable characteristics of households within municipios and of individuals within households to confirm the robustness of our estimation. Equation 1 captures our preferred DD regression specification:

Equation 1:

$$Y_{ihmt} = \alpha_t + \gamma_m + \beta_1 MCA_{mt} + \beta_2 EMPRESA_{mt} + MTI_{ihmt} + X_{ht} + X_{it} + \upsilon_{mt} + \varepsilon_{it}$$

Y_{ihmt} is an individual or household by municipio by time outcome
α_t are time fixed effects for 2001 and 2005
γ_m is a municipio fixed effect
MCA_{mt} is an indicator for treatment with an MCA project in year t
$EMPRESA_{mt}$ is an indicator for treatment with an Empresa project in year t
MTI_{mt} is an indicator for the presence of an non-WB4 road project in year t
X_{ht} is a set of household level by year controls
X_{it} is a set of individual level by year controls
υ_{mt} is an unobservable municipio level error term that may vary over time
ε_{it} is an stochastic i.i.d individual error term

We adjust our sample frame to create three alternative counterfactuals to observations in our treated municipios. Our "naïve" sample specification reduces the raw EMNV data to exclude households in municipios that are not surveyed across all three year of the EMNV and allows for comparison between WB4 targeted municipios and all other non-dropped municipios in the data. In addition to our naïve specification, we tailor additional samples to construct alternative counterfactual control groups. In our primary "base" sample specification, we drop all municipios from departments that do not contain a WB4 project. This should reduce potential bias in our estimation to the extent that departments not targeted by WB4 interventions differ on unobservable characteristics from targeted departments in ways that may change over time. Notably, the dropped departments include six of the seven least population dense departments and include the two autonomous regions in the Eastern part of the country. As reflected in Table 3, observable characteristics are better balanced between treated and control municipios in the base sample. Additionally, as a robustness check, we construct an "uncontaminated" sample by dropping from the naïve sample any municipios that were targeted by non-WB4 projects, including road and bridge improvement projects funded by other development agencies.

As an essential first step for validating the use of a DD method, we will demonstrate that unobservable characteristics are not differentially affecting the pre-intervention linear trends in key outcomes of interest - the parallel trends assumption. We will provide visual evidence for the naïve sample and statistical evidence for our preferred base sample.

5 Testing Parallel Pre-Trends

WB4 projects were implemented in the second period of EMNV data, between 2005-2009, providing two pre-intervention observations (2001 and 2005 rounds) with which we can test that our outcomes of interest evolved along parallel linear trends between treated and control groups – a necessary assumption for the internal validity with the DD strategy. If trends are not parallel we may be concerned that treatment and control groups differ along unobserved characteristics that may differentially drive trends in the outcomes of interest in ways that would be spuriously associated with the project treatment, leading to mistaken inference about project impacts.

Figure 4 visually depicts the trends in various household and individual level outcomes over time, using the naïve sample, and Table 5 statistically tests the parallel trends assumption using the base sample. The regressions in Table 5 include a full suite of household and individual controls and cluster standard errors at the municipio level according to Equation 2 where β_1 and β_2 are our coefficients of interest.

Equation 2:

$$Y_{ihmt} = \alpha_{2005} + \gamma_m + \beta_1(MCA_m * \alpha_{2005}) + \beta_2(EMPRESA_m * \alpha_{2005}) + X_{ht} + X_{it} + v_{mt} + \varepsilon_{it}$$

Y_{ihmt} is an individual or household by municipio by time outcome
α_t is a time fixed effect for 2005
γ_m is a municipio fixed effect
MCA_m is an indicator for treatment with an MCA project
$EMPRESA_m$ is an indicator for treatment with an Empresa project
X_{ht} is a set of household level controls
X_{it} is a set of individual level controls
v_{mt} is an unobservable municipio level error term that may vary over time
ε_{it} is an stochastic i.i.d individual error term

Table 5 reveals that, using the base sample, municipalities display parallel linear trends between MCA treated and control groups for seven of eight key outcomes. This is consistent with the suggestive visual evidence in Figure 4. Only the coefficient on whether or not an individual worked most recently as a laborer is significantly different from zero and suggests that the likelihood of working as a laborer in control groups increased more than in treated groups in the pre-period. Therefore, to the extent we find likelihood of working as a laborer positively impacted by the MCA treatment, as indeed we do, the intervention effect would have to overcome this pre-trend. Although this pre-trend works in favor of our identification claims, in some sense, it does suggest that there may be unobservable characteristics that affect the likelihood of working as a laborer, which could be cause for concern if such unobservables have a non-linear relationship with the outcomes of interest. Notably, coefficients on the Empresa treatment variable are significant across extreme poverty, matriculation, and whether worked last as laborer, supporting our prior belief that Empresa projects were subjected to substantial selection bias and that MCA and Empresa projects should be disaggregated from a pooled WB4 treatment indicator in the subsequent analysis.

To further validate that robustness of our finding of parallel pre-trends and potentially increase the precision of our point estimates, we can exploit the fact that the 2005 EMNV round sought to include all households surveyed in the 2001 EMNV round to create an individual panel. Table 6 provides regression outputs for the individual panel sub-sample within the base sample estimated according to Equation 3, which exploits individual fixed effects.

Equation 3:

$$Y_{ihmt} = \alpha_{2005} + \gamma_i + \beta_1(MCA_m * \alpha_{2005}) + \beta_2(EMPRESA_m * \alpha_{2005}) + X_{ht} + X_{it} + \varepsilon_{it}$$

The results are largely consistent with the parallel trends claim. In some cases standard errors increase and in other cases they decrease compared to the repeated cross-section. Notably, the significant coefficients on the Empresa treatment disappear for extreme poverty.

Comparing Panel to Repeated Cross-Section Estimates

Our principal analysis applies a DD to repeated cross-sections of the EMNV, controlling for observable characteristics. However, we are limited in the characteristics we observe and can therefore use as controls. To check whether we should be concerned that time invariant but unobserved individual or household level characteristics significantly affect our estimates, we exploit the pre-intervention panel sub-sample used to test parallel trends in Table 6. Specifically, we conduct the same regression specified in Equation 2, treating the data as a repeated cross-section with municipal fixed effects, on the sub-sample of data in the 2001-2005 panel. Comparing estimates with municipio fixed effects in Table 7 vs. individual fixed effects in Table 6 reveals similar estimates on household outcomes of paved road access and poverty. Coefficients on Empresa treatment status do vary slightly for individual level outcomes in some cases, suggesting that unobservable time invariant individual characteristics may co-vary with Empresa treatment status. MCA treatments are our primary treatment of interest and MCA coefficients remain quite stable and insignificant across placebo specifications with the exception of the outcome "whether or not the individual works as a laborer" for which there is a small negative coefficient that is significant in the cross-sectional estimate and when including individual fixed effects, but insignificant when including panel municipio fixed effects. In conclusion, it appears that inference from the repeated cross-section with municipio fixed effects will not be substantially different than what would have been estimated from a true three period household panel with household or individual fixed effects.

6 Results

Examining both household and individual outcomes of interest, we find strong evidence that the WB4 projects are associated with improvement to road infrastructure and suggestive evidence of select labor market, education, and health related impacts.

Table 8 captures the estimated effects of pooled treatments and MCA-only treatment on three household outcomes across our three different sample specifications. Estimates are derived from OLS regressions with full household controls, year and municipio fixed effects, and control for the presence of other road and bridge improvement projects with standard errors clustered at the municipio level. Point estimates significant at 95% confidence are highlighted in bold and italics and marginally significant point estimates are highlighted in italics only.

Road Quality and Access

As a first step, we want to validate that the road improvement projects actually improved road quality and access, a key goal of the WB4 projects and likely an essential channel through which other social and economic benefits of the program may be realized. For a variety of reasons program funding might not immediately translate into improved roads. Weak government institutional capacity or corruption could lead to fund leakage, inefficient project implementation, etc.. In fact, one of the explicit goals of the WB4 funding was to increase the capacity of MTI. Additionally, potential confounding factors could include the low skill level of MCA participants, small footprint of MCA projects, and short period between the commencement of many of the projects and our endline survey. Finally, we want to verify that there is indeed sufficient signal in our data, which are derived from a national household livelihood survey that was not designed with MCAs in mind.

We do not have objective measures of pre- and post-intervention road quality, but can exploit the survey question regarding whether or not a household's principal access route is paved or cobble stoned ("adoquinado") from the question: "What is the principal access route to arrive in the community in which you live?"[2] Responses to this question are the dependent variable "Paved" in Table 8. The table illustrates that the substantial impact of pooled WB4 road improvement projects and the MCA-only WB4 projects on the likelihood of having one's principal access point paved is a robust finding across all specifications. The estimates of pooled impacts vary from 13% to 23% increased likelihood of paved access and estimates of MCA impacts vary from 12% to 23% increases.

To confirm that effects attributed to MCA projects are not spuriously capturing variation in outcomes truly attributable to other infrastructure improvements, we control for indicators such as hookup to the electric grid, landline and mobile telephone ownership, and indoor plumbing. Table 9 and Table 10 demonstrate that introducing these other infrastructure indicators as controls in regressions with paved access roads as an outcome does not meaningfully alter the point estimate on a pooled WB4 treatment indicator (WB4 Started 2009) or MCA-specific treatment indicator (WB4 MCA 2009).

[2] The actual Spanish version of the question reads: "Cual es el principal via de acceso para llegar a la comunidad / barrio donde se encuentra ubicada su vivienda?"

Despite robust findings that MCAs improve access to paved roads, evidence for other non-WB road building is not as convincing. Tables 9 & 10 show estimates on indicators for the presence of non-WB4 road improvement projects implement through MTI. The indicator for other MTI road improvement projects actually has a negative coefficient and the indicator for other MTI bridge improvement projects is large but insignificant, though both have wide standard errors. Controlling for other MTI projects actually increases estimates on the MCA indicator, suggesting that non-MCA control municipios may be contaminated by positive effects of other MTI projects. The Empresa indicator has smaller and at best marginally significant coefficients compared to MCA projects, though we hesitate to interpret the coefficients on either Empresa or other MTI project indicators as reflecting causal impacts given the lack of parallel trends for those projects and their counterfactual controls in the pre-intervention period.

Poverty

Given the apparent impacts of MCAs on road improvements, we turn to the impact of WB4 projects on household poverty. Even in the short-term MCAs could hypothetically reduce poverty through a variety of channels. Given that MCAs were designed to target the unemployed, direct employment effects could boost incomes and reduce poverty. Nevertheless, emergence of poverty impacts may require longer time periods than the relatively short interval between most of the WB4 project commencement dates and the 2009 endline survey. In fact, estimates with full controls across all three different sample specifications in Table 8 reveal relatively precise estimates of zero effect on poverty and extreme poverty. Results not provided here controlling for heterogeneous effects on poverty by duration of project exposure and gender of head of household are similar. We hesitate to take too seriously the results of a null effect on poverty given that the majority of projects were still under construction by the time of the endline and that the poverty indicator may be noisy given that it is binary and based on spatially adjusted aggregates of self-reported consumption. Nevertheless, the small local nature of MCA road improvements and short-term nature of the intended employment generation may genuinely constrain potential impacts on poverty.

Employment

Whether or not MCAs in fact stimulated employment as intended is a first order question. Table 11 provides individual labor market outcomes across our three sample specifications. Surprisingly, on the left hand side of the table, we do not find any robust associations between road projects and employment within the past week, past year, or entire prior lifetime for anything other than bridge projects, which are clearly targeted and whose coefficients vary widely with the inclusion of controls in Tables 9 & 10. So why might we not see short-term or medium-term employment effects in the case of a workfare project explicitly designed to stimulate employment?

As previously discussed, the process of recruiting MCA laborers may be highly endogenous, subject to the whim of local officials and project managers or reflecting the local labor market conditions that lead some people to remain on the unemployment rolls in a non-random manner. To explore this hypothesis we test whether the type of work in which people engage is affected by the intervention. The right side of Table 11 estimates the association between the various interventions and an indicator for whether the position of the occupation in which the respondent worked the most in the prior week was that of an employee/laborer (*empleado/obrero*), employer/entrepreneur (*empleador/cuenta propia*), or farmhand (*jornalero/peon*). As expected, we do see a 7.4%

increase in the likelihood of working as a laborer associated with MCA projects in our base specification that is robust to using our naïve specification as well. Other analysis that is not reported in the tables suggests that the increase in working as a laborer is larger and more statistically significant for men (8.3%, p<.02) than women (5.3%, p>.23). This result may be consistent with the lower participation of women in MCAs relative to men as reported by MTI. To see if this increase in likelihood of working as a laborer reflects substitution away from other positions, as could be the case if employed but locally connected people were positively selected into MCA projects, we estimate outcomes on entrepreneurship and work as farmhand. Our estimates do not find any association with entrepreneurship, but do reveal small negative coefficients on work as a farmhand that are only statistically significant for the naïve sample. Nevertheless, any substitution away from farmhand cannot fully explain the association between the intervention and increases in likelihood of working as a laborer given the small magnitude of the estimates combined with the fact that there are approximately 20% fewer farmhands than laborers in our pre-intervention years.

Viewed together, these mixed results on individual labor outcomes present a puzzle, which may at least be partially explained by lack of perfect survey harmonization across EMNV rounds. In 2009 the relevant question asks if the respondent worked for at least an hour with salary or in a family business. In 2005 the question merely asks if the respondent worked. And, in 2001, the question asks if the respondent worked even without pay. Given the importance of labor market impacts to WB4-style interventions, our mixed results, and the limitations of the EMNV data, we recommend that future data gathering exercises more precisely focus on short-term employment generation questions. Specifically asking about participation in MCAs and counterfactual employment, i.e. what respondents would otherwise be doing, would overcome many of our current data limitations. Additionally, specifically exploring the selection process with questions about local political affiliations, family connections, gender dynamics, etc. could shed light on the recruitment process – an important facet of program implementation with relevance beyond this specific intervention.

Education and Health

Further exploring individual level social outcomes, we find suggestive, but again not conclusive evidence, for project impacts on education and health related outcomes. Table 12 provides regression results for whether or not school age individuals matriculated in schools within the past 12 months across a variety of specifications using our preferred base sample. Consistent with results from Mu and van de Walle (2011) and Khandker and Koolwal (2011), we only find robust associations between the intervention and schooling among primary age children 7 to 12 years old with stronger results for girls. Columns (1)-(3) of Table 12 respectively estimate impacts of the pooled project sample, of all MCA projects, and of only the four MCA projects that were completed before the endline on all individuals age 7-18 with no significant results. Column (4) estimates impacts of the four completed MCA projects on matriculation for children 7-12, revealing that completed projects are associated with a 6.3% increase in matriculation. This estimate is similar in size to that of column (3) with all MCA projects, but more precisely estimated. The intuition for this specification is straightforward. Schools are only more accessible when roads are actually completed. Prior researchers have explained that effects may be isolated to primary school given the lower attendance in and, often, great distances between secondary schools in many low-income countries. Another hypothesis suggests that better access to more distant secondary schools actually spurs demand for pre-requisite primary education. Column (5) tests for heterogeneity by gender, finding a highly significant 7.3% percent increase for primary school girls and a 3.4% statistically insignificant estimate for boys (F-statistic of 1.49).

In general, our findings on education impacts are consistent with the existing literature. However, in future data collection efforts it would be useful to try to decompose the precise channel for increased school attendance. Is it lower time of travel, lower monetary cost of travel, or more complex dynamics, such as increased expected returns to education or positive household income shocks from short-term employment that mediate increased matriculation? Additionally, MCA work was only intended for adults, but it is conceivable that secondary school age individuals could be induced to drop out of school in the presence of new MCA employment opportunities consistent with findings on the negative matriculation impacts of the growth of low-skilled manufacturing jobs in Mexico (Atkin, 2009).

Our final outcomes of interest concern the potential effects of MCAs on health. We examine whether the predictions that i) better quality roads reduce incidence of respiratory illness and ii) our intervention stimulates increased access to primary care for women of childbearing age are consistent with our data. Comparison of Table 13 columns (1) and (2) reveals a 4.7% decline in respiratory illness that is isolated to completed MCA projects only. This is consistent with the story that better paved roads reduce ambient dust from traffic, but roads under construction would have no such effect. Column (3) reveals that reductions in respiratory illness are isolated to men. Columns (4) suggests MCAs are associated with increased visits to health facilities during times of good health for individuals aged 14-45 (consistent with non-emergency primary care), though the coefficient is small in magnitude and identified off of a small sample of individuals. Column (5) confirms that increased healthcare access is limited to women of childbearing age who we expect are more likely to seek primary care than men.

7 Discussion

We have conducted a variety of analyses to gather evidence regarding the impacts of the Fourth Roads Rehabilitation and Maintenance Project funded by the World Bank on a variety of outcomes of interest. Matching administrative data on road improvement project implementation between 2005-2009 provided by Nicaragua's Ministry of Transport and Infrastructure, we estimate a difference-in-difference model with municipio fixed effects on three years of repeated cross-sectional data from household surveys designed on the model of the LSMS. We disaggregate MCA projects intended as both workfare and road improvement interventions and Empresa projects constructed purely by professional contractors, dropping municipios in administrative departments without WB4 projects in our preferred base sample specification. Our results reveal meaningful associations between MCA projects and increased access to paved roads, increased likelihood of working as a laborer, increased matriculation in primary school among girls, decreases in respiratory illness among men, and increased access to primary care among women of childbearing age. In general, our findings are consistent with the existing literature on short-term impacts of road improvement projects and heterogeneity by age and gender. Notably, we find null associations with poverty and puzzling evidence on the associations with employment that beg important questions about the process of local recruitment of laborers into MCAs.

While we demonstrate parallel trends in treated and control groups using graphical and econometric methods across a variety of outcome variables and specifications, we hesitate to make strong claims of causal impacts given multiple layers of project and household selection, and key limitations in our data.

The presence of compelling findings despite data limitations, begs for further analysis of future road improvement efforts in Nicaragua for which our analysis may serve as a guide. Key to future data collection efforts should be pinning down the process of selecting individuals into MCAs and the channels through which participation in MCAs may impact more general outcomes such as matriculation, access to health clinics, and decreases in respiratory illness. Better understanding the source of heterogeneity by gender, for example, may point to more specific policy levers for stimulating gender equity and women's empowerment. Future surveys should investigate the local selection process with questions about local political affiliations, family connections to local leadership, and gender dynamics within MCA recruitment. Efforts should be made to rigorously document the commencement and conclusion of project construction.

In addition to specifically asking about participation in MCAs, questionnaires could capture perceptions of what respondents believe would be the counterfactual to MCA employment to estimate the extent to which MCAs stimulate new employment or encourage substitution away from other occupations. Data collection efforts could further consider comparing workfare wages to prevailing local wages, identifying whether there is excess demand for workfare, and observing how that demand may vary seasonally. In terms of educational outcomes, it would be useful to try to decompose the precise channel for increased primary school attendance and understand why we do not find evidence of impacts on secondary school matriculation. Is it lower time of travel, lower monetary cost of travel, or more complex dynamics, such as increased expected returns to education or positive household income shocks from short-term employment, that mediate increased primary school matriculation? Finally, in the process of collecting household data, surveyors should collect data on the frequency, location, and history of local markets and businesses to assess the effect that road improvements may have had on market institutions.

References

Atkin, David, “Endogenous Skill Acquisition and Export Manufacturing in Mexico," Yale, 2015.

Banerjee, Abhijit and Duflo, Esther and Qian, Nancy (2012). "On the Road: Access to Transportation Infrastructure and Economic Growth in China," NBER Working Papers 17897.

Casaburi, Lorenzo and Glennerster, Rachel and Suri, Tavneet. “Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone,” International Growth Center, November 2013.

Dercon, Stefan, et. al (2009). “The Impact of Agricultural Extension and Roads on Poverty and Consumption Growth in Fifteen Ethiopian Villages.” *American Journal of Agricultural Economics*, 91:4.

Faber, Benjamin. “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System.” *The Review of Economic Studies*, March 26, 2014.

Gibson, John and Rozelle, Scott. (2003). “Poverty and Access to Roads in Papua New Guinea,” *Economic Development and Cultural Change*, 52:1.

Gonzalez-Navarro, Marco, and Climent Quintana-Domeque. “Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects.” *Review of Economics and Statistics* 98, no. 2 (November 16, 2015): 254–67.

Jensen, Robert (2012). "Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India," *Quarterly Journal of Economics*, 127(2), p. 753-792.

Khandker, Shahidur R. and Bakht, Zaid & Koolwal, Gayatri B. (2009). "The Poverty Impact of Rural Roads: Evidence from Bangladesh," *Economic Development and Cultural Change*, University of Chicago Press, vol. 57(4), pages 685-722.

Khandker, Shahidur R. & Koolwal, Gayatri B., 2011. "Estimating the long-term impacts of rural roads: a dynamic panel approach," Policy Research Working Paper Series 5867, The World Bank.

Mu, Ren & van de Walle, Dominique (2011). "Rural Roads and Local Market Development in Vietnam," *Journal of Development Studies*, Taylor & Francis Journals, vol. 47(5), pages 709-734.

Subbarao, Kalanidhi; del Ninno, Carlo; Andrews, Colin; Rodríguez-Alas, Claudia. 2013. *Public Works as a Safety Net: Design, Evidence, and Implementation.* Washington, DC: World Bank.

World Bank Group. *Project Appraisal Document: Report No: 61418-NI*. November 10, 2011.

Table 1: Number of Municipios Targeted by Different Project Types across Samples

		Unique Munis in Preferred Sub-sample of EMNV	Unique EMNV Panel Munis	EMNV Raw	All Nicaragua
WB4:	*MCA*	15	15	20	25
	Empresa	8	8	9	9
	Both	3	3	3	4
	Subtotal	20	20	26	30
Other MTI:	*Roads*	20	27	36	38
	Bridges	6	10	13	14
	Both	2	2	2	2
	Subtotal	24	35	49	52

Table 2: Project Overlap by Municipios in Preferred Sample Specification

		WB4: *MCA*	*Empresa*	**MTI:** *Roads*	*Bridges*	**Pure Controls**	**Total**
WB4:	*MCA*	15					
	Empresa	3	8				
MTI:	*Roads*	6	1	20			
	Bridges	0	0	2	6		
Pure Controls						31	
Total							68

Table 3: Treatment vs. Control Municipalities in 2005 Pre-Intervention Cross-Section across Alt. Sample Specifications

	NAÏVE - All Obs.					BASE - Treated Depts. Only					UNCONTAMINATED - No New Road Controls				
	Treated		Control			Treated		Control			Treated		Control		
	Mean	n	Mean	n	t-stat	Mean	n	Mean	n	t-stat	Mean	n	Mean	n	t-stat
Hhld size	6.95	6,560	7.08	24,763	-2.70	6.95	6,560	7.03	10,981	-1.42	6.94	6,730	7.05	12,745	-2.12
Multi hhld	0.11	6,560	0.06	24,763	15.41	0.11	6,560	0.09	10,981	3.50	0.11	6,730	0.07	12,745	10.24
No. rooms	2.74	6,560	2.56	24,763	9.61	2.74	6,560	2.54	10,981	9.17	2.72	6,730	2.54	12,745	8.63
Tele - Land	0.13	6,560	0.08	24,763	13.37	0.13	6,560	0.13	10,981	1.01	0.13	6,730	0.11	12,745	4.61
Tele - Cell	0.28	6,560	0.16	24,763	21.40	0.28	6,560	0.22	10,981	8.92	0.27	6,730	0.17	12,745	15.82
Farmhouse	0.01	6,560	0.03	24,763	-8.92	0.01	6,560	0.01	10,981	-1.24	0.01	6,730	0.03	12,745	-8.54
Shack	0.03	6,560	0.02	24,763	2.00	0.03	6,560	0.03	10,981	-2.98	0.03	6,730	0.02	12,745	0.75
Plumbing	0.36	6,560	0.20	24,763	26.10	0.36	6,560	0.30	10,981	7.94	0.35	6,730	0.23	12,745	17.85
Electric grid	0.82	6,560	0.56	24,763	40.37	0.82	6,560	0.75	10,981	12.10	0.81	6,730	0.58	12,745	33.57
Biomass fuel	0.63	6,543	0.73	24,727	-16.52	0.63	6,543	0.67	10,965	-5.35	0.64	6,713	0.70	12,730	-8.98
Age	26.50	6,560	24.23	24,763	8.49	26.50	6,560	25.85	10,981	2.13	26.41	6,730	24.69	12,745	5.87
Male	0.48	6,560	0.50	24,763	-2.02	0.48	6,560	0.49	10,981	-1.41	0.48	6,730	0.50	12,745	-1.58
Single	0.40	6,560	0.38	24,763	3.97	0.40	6,560	0.41	10,981	-0.93	0.40	6,730	0.38	12,745	2.82

Note: Means computed across individual observations in full 2005 EMNV (equivalent of weighting household variables by the number of individuals in the household). "Multi hhld" refers to multiple households living in the same dwelling. Farmhouse and Shack refer to the lowest quality forms of housing reported.

Table 4: Description of Outcome Variables over Time in Base Sample

	2001				2005				2009			
	Treatment		Control		Treatment		Control		Treatment		Control	
	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n	Mean	n
Paved	0.45	911	0.486	1,659	0.54	1,298	0.577	2,133	0.75	979	0.718	4,352
Poverty	0.32	911	0.353	1,659	0.36	1,298	0.392	2,133	0.27	979	0.2	4,358
Extreme Pov.	0.09	911	0.115	1,659	0.08	1,298	0.13	2,133	0.05	979	0.044	4,358
Concrete Wall	0.27	911	0.357	1,659	0.39	1,298	0.408	2,133	0.55	979	0.612	4,353
Cement Floor	0.50	911	0.521	1,659	0.54	1,298	0.541	2,133	0.62	979	0.716	4,353
Wrk past wk	0.49	3,564	0.49	6,631	0.52	5,240	0.50	8,560	0.52	3,568	0.49	15,781
Wrk past yr	0.55	3,564	0.55	6,631	0.55	5,240	0.55	8,560	0.57	3,568	0.54	15,781
Wrk ever in past	0.65	3,564	0.65	6,631	0.66	5,240	0.65	8,560	0.69	3,568	0.69	15,781
Laborer	0.43	1,835	0.41	3,395	0.40	2,826	0.40	4,535	0.51	1,888	0.52	7,983
Entrep	0.28	1,835	0.28	3,395	0.30	2,826	0.31	4,535	0.33	1,888	0.36	7,983
Respiratory sick.	0.14	2,278	0.14	4,282	0.24	3,265	0.22	5,378	0.20	2,376	0.18	10,331
Med visit	0.01	1,552	0.01	2,955	0.03	2,029	0.04	3,507	0.04	1,456	0.03	6,852
Matriculated	0.77	1,603	0.75	3,052	0.76	2,134	0.77	3,574	0.77	1,333	0.80	5,511
School too far	0.06	370	0.03	773	0.01	510	0.02	830	0.02	304	0.02	1,114

Note: Outcome variables are summarized within the appropriate demographic groups using Base specification (dropping all departments that did not have a treated municipio). Household outcomes include: whether the community's access road is paved, whether the household is poor or extremely poor, and whether the home has concrete walls and/or cement floors. Individual labor market outcomes are reported only for individuals ten years or older and include: a) whether or not an individual worked in the past week, past year, or ever before, and b) whether the individuals most recent primary source of employment was as a laborer/employee or as self-employment/entrepreneurship conditional on having worked in the past week. Health outcomes are limited to incidence of respiratory illness in the past month and include all respondents. The Medical Visit indicator is a proxy for preventive care and is limited to women of approximate childbearing age between 15-45. The variable captures whether a woman has visited a clinic for medical consultation or medicine in the past month despite reporting otherwise good health. Matriculated is an indicator for whether individuals age 7-18 have matriculated in the formal school system in the past year and "School too far" is an indicator for whether individuals of school age did not matriculate because of distance to school.

Table 5: Parallel Trends Test - 2001 & 2005 Cross-sections

VARIABLES	(1) Paved	(2) Poverty	(3) Extr Pov	(4) Med Acc.	(5) Matric.	(6) Wrk yr	(7) Wrk wk	(8) Laborer
MCA 2005	0.026	0.012	-0.003	-0.017	-0.016	0.005	0.012	-0.039*
	(0.085)	(0.028)	(0.016)	(0.017)	(0.028)	(0.014)	(0.013)	(0.023)
Empresa 2005	0.056	0.027	-0.024*	-0.023	-0.045*	-0.002	0.009	0.047*
	(0.148)	(0.022)	(0.013)	(0.027)	(0.025)	(0.018)	(0.017)	(0.025)
Observations	6,001	6,001	6,001	4,756	9,631	23,995	23,995	12,591

Table 6: Parallel Trends Test - 2001 & 2005 Individual Panel with Individual FE

VARIABLES	(1) Paved	(2) Poverty	(3) Extr Pov	(4) Med Acc.	(5) Matric.	(6) Wrk yr	(7) Wrk wk	(8) Laborer
MCA 2005	0.035	0.014	-0.014	-0.030	0.035	-0.009	0.006	-0.040**
	(0.025)	(0.024)	(0.017)	(0.025)	(0.025)	(0.015)	(0.015)	(0.020)
Empresa 2005	0.042	0.047	-0.001	-0.010	0.003	0.003	0.009	0.051*
	(0.035)	(0.033)	(0.023)	(0.038)	(0.034)	(0.021)	(0.021)	(0.027)
Observations	4,119	4,119	4,119	3,489	7,024	17,542	17,542	9,083

Table 7: Parallel Trends Test - 2001 & 2005 Individual Panel with Municipio FE

VARIABLES	(1) Paved	(2) Poverty	(3) Extr Pov	(4) Med Acc.	(5) Matric.	(6) Wrk yr	(7) Wrk wk	(8) Laborer
MCA 2005	0.036	0.023	-0.015	-0.025	0.027	-0.009	0.000	-0.046
	(0.080)	(0.031)	(0.015)	(0.018)	(0.029)	(0.015)	(0.016)	(0.028)
Empresa 2005	0.050	0.042	0.000	-0.021	-0.070**	0.003	0.022	0.044
	(0.143)	(0.025)	(0.017)	(0.029)	(0.032)	(0.020)	(0.017)	(0.031)
Observations	4,119	4,119	4,119	3,489	7,024	17,542	17,542	9,083

Notes for Tables 5,6, & 7: Table 5 regressions are estimated using the Base specification (dropping all departments that did not have a treated municipio). Table 6 regressions use the subset of observations from the Base specificaion for which a panel could be constructed and individual fixed effects. Table 7 regressions use the panel sub-sample with municipio fixed effects. Household outcomes include: whether the community's access road is paved, and whether the household is poor or extremely poor. Med Access is a proxy for preventive care and is limited to individual women of approximate child-bearing age between 15-45. The variable captures whether a woman has visited a clinic for medical consultation or medicine in the past month despite reporting otherwise good health. Matric is an indicator for whether individuals age 7-18 have matriculated in the formal school system in the past year. Labor market outcomes are reported only for individuals ten years or older and include: a) whether or not an individual worked in the past year, past week, or ever before, and b) whether the individuals most recent primary source of employment was as a laborer/employee conditional on having worked in the past week. All estimates report robust standard errors. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 8: Household Outcomes across Three Specifications

		NAÏVE - All obs		BASE - Treated Depts. Only		UNCONTAMINATED - No New Roads	
		Pooled	**MCA**	**Pooled**	**MCA**	**Pooled**	**MCA**
Paved	Effect Size	***0.125***	*0.122*	***0.169***	***0.164***	***0.228***	***0.235***
	se	***0.064***	*0.072*	***0.067***	***0.076***	***0.067***	***0.081***
	p-value	***0.054***	*0.094*	***0.014***	***0.035***	***0.001***	***0.005***
	n	15674	15674	11274	11274	11563	11563
Poverty	Effect Size	-0.001	0.020	-0.001	0.029	-0.020	0.016
	se	0.031	0.031	0.030	0.031	0.037	0.034
	p-value	0.980	0.512	0.982	0.358	0.586	0.633
	n	15675	15675	11275	11275	11564	11564
Extreme Pov	Effect Size	-0.012	-0.018	-0.013	-0.016	-0.003	-0.010
	se	0.015	0.016	0.014	0.016	0.019	0.020
	p-value	0.413	0.276	0.386	0.314	0.884	0.621
	n	15675	15675	11275	11275	11564	11564

Note: "MCA" refers to point estimates from a treatment indicator for WB4 supported MCA projects while controlling for WB4 supported Empresa projects. "Pooled" refers to point estimate from pooling indicators for WB4 supported MCA and Empresa projects into a single treatment indicator. Estimates are derived from OLS regressions with full household controls, year and municipio fixed effects, and control for the presence of other road and bridge improvement projects with standard errors clustered at the municipio level. Point estimates significant at 95% confidence are highlighted in bold and italics, and point estimates marginally significant at 90% confidence are highlighted in italics only.

Table 9: Validate Pooled Treatment Estimates with Project Type and Household Controls

VARIABLES	(1) Paved	(2) Paved	(3) Paved	(4) Pov	(5) Pov	(6) Pov
WB4 Started 2009	0.161**	0.205***	0.169**	-0.040	-0.054	-0.001
	(0.071)	(0.074)	(0.067)	(0.042)	(0.041)	(0.030)
MTI Other Road 2009		-0.082	-0.049		0.019	-0.001
		(0.074)	(0.070)		(0.041)	(0.031)
MTI other Bridge 2009		0.252	0.126		-0.158***	-0.052
		(0.155)	(0.127)		(0.054)	(0.035)
Hhldsize			-0.006*			0.098***
			(0.003)			(0.009)
Hhldsize Sq.			0.000*			-0.003***
			(0.000)			(0.000)
Multi Hhlds			0.045**			-0.304***
			(0.020)			(0.017)
No. Rooms			0.015***			-0.049***
			(0.004)			(0.007)
Tele - Land			0.094***			-0.046***
			(0.016)			(0.008)
Tele - Cell			0.026			-0.109***
			(0.016)			(0.010)
Farmhouse			0.014			0.152***
			(0.042)			(0.037)
Shack			0.010			0.152***
			(0.037)			(0.026)
Plumbing			0.104***			-0.041***
			(0.020)			(0.012)
Electric Grid			0.230***			-0.194***
			(0.026)			(0.019)
Biomass Fuel			-0.116***			0.186***
			(0.024)			(0.014)
Observations	11,332	11,332	11,274	11,338	11,338	11,275
R-squared	0.266	0.268	0.347	0.147	0.148	0.395

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The outcome variables Paved and Pov indicate whether the community's access road is paved and whether a household is poor, respectively. WB4 Started 2009 is a "pooled" treatment dummy indicator for municipios that had WB4 supported roadbuilding projects that started (but may not have been completed) between 2005 and 2009 built by MCA teams and/or contractors with heavy machinery. MTI Other Road 2009 and MTI Other Bridge 2009 are dummy indicators for MTI road or bridge building projects between 2005 and 2009 that were not supported by WB4. The remaining variables are controls that include the number of people in the household and the square of that term, whether there are multiple households living in the dwelling, the number of rooms in the dwelling, the presence of a fixed line or cellular phone, whether the dwelling is a farmhouse or a shack (two low quality housing types associated with extreme poverty), the presence of plumbing in the home, access to a power grid, and whether or not the household uses biomass for fuel.

Table 10: Validate MCA-Specific Estimates with Project Type and Household Controls

VARIABLES	(1) Paved	(2) Paved	(3) Paved	(4) Pov	(5) Pov	(6) Pov
WB4 MCA 2009	0.162**	0.209**	0.164**	-0.023	-0.030	0.029
	(0.075)	(0.081)	(0.076)	(0.042)	(0.041)	(0.031)
WB4 Empresa 2009	0.157*	0.143	0.124	-0.118**	-0.120**	-0.084***
	(0.093)	(0.086)	(0.087)	(0.052)	(0.053)	(0.031)
MIT Other Road 2009		-0.079	-0.043		0.003	-0.017
		(0.077)	(0.074)		(0.039)	(0.029)
MTI Other Bridge 2009		0.254	0.127		-0.157***	-0.050
		(0.155)	(0.127)		(0.053)	(0.034)
Hhldsize			-0.006*			0.098***
			(0.003)			(0.009)
Hhldsize Sq.			0.000*			-0.003***
			(0.000)			(0.000)
Multi Hhlds			0.045**			-0.304***
			(0.020)			(0.017)
No. Rooms			0.015***			-0.049***
			(0.004)			(0.007)
Tele - Land			0.094***			-0.047***
			(0.016)			(0.008)
Tele - Cell			0.026			-0.108***
			(0.016)			(0.010)
Farmhouse			0.014			0.154***
			(0.043)			(0.037)
Shack			0.010			0.152***
			(0.037)			(0.026)
Plumbing			0.103***			-0.040***
			(0.020)			(0.012)
Electric Grid			0.229***			-0.195***
			(0.026)			(0.019)
Biomass Fuel			-0.115***			0.186***
			(0.024)			(0.014)
Observations	11,332	11,332	11,274	11,338	11,338	11,275
R-squared	0.268	0.269	0.348	0.148	0.148	0.396

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The outcome variables Paved and Pov indicate whether the community's access road is paved and whether a household is poor, respectively. WB4 MCA 2009 is a dummy treatment indicator for municipios with WB4 supported MCA projects between 2005 and 2009. WB4 Empressa 2009 is dummy indicator for municipios that have WB4 supported roadbuilding projects constructed by contractors with heavy machinery and not MCA teams between 2005 and 2009. MTI Other Road 2009 and MTI Other Bridge 2009 are dummy indicators for MTI road or bridge building projects between 2005 and 2009 that were not supported by WB4. The remaining variables are controls that include the number of people in the household and the square of that term, whether there are multiple households living in the dwelling, the number of rooms in the dwelling, the presence of a fixed line or cellular phone, whether the dwelling is a farmhouse or a shack (two low quality housing types associated with extreme poverty), the presence of plumbing in the home, access to a power grid, and whether or not the household uses biomass for fuel.

Table 11: Individual Labor Market Outcomes across Three Specifications

Wrk past wk	Naive	Base	Uncontaminated
WB4 MCA 2009	0.017	0.017	-0.029**
	(0.022)	(0.025)	(0.013)
WB4 Empresa 2009	-0.011	-0.008	-0.002
	(0.024)	(0.024)	(0.017)
MTI Other Road 2009	0.007	0.013	0.074***
	(0.017)	(0.025)	(0.027)
MTI Other Bridge 2009	0.033**	0.004	
	(0.014)	(0.025)	
Observations	61,042	43,247	44,509

Job Last as Worker	Naive	Base	Uncontaminated
WB4 MCA 2009	0.057**	0.074**	0.013
	(0.026)	(0.028)	(0.030)
WB4 Empresa 2009	-0.003	0.002	0.005
	(0.047)	(0.049)	(0.042)
MTI Other Road 2009	0.001	-0.009	0.065
	(0.020)	(0.025)	(0.039)
MTI Other Bridge 2009	-0.045	-0.131***	
	(0.055)	(0.049)	
Observations	31,557	22,394	23,030

Wrk past yr	Naive	Base	Uncontaminated
WB4 MCA 2009	0.021	0.020	-0.024*
	(0.024)	(0.025)	(0.014)
WB4 Empresa 2009	0.025	0.027	0.034
	(0.022)	(0.022)	(0.024)
MTI Other Road 2009	0.0143	0.019	0.080***
	(0.019)	(0.025)	(0.029)
MTI Other Bridge 2009	0.039**	0.026	
	(0.015)	(0.041)	
Observations	61,042	43,247	44,509

Job Last as Entrep	Naive	Base	Uncontaminated
WB4 MCA 2009	0.000	-0.013	0.015
	(0.022)	(0.024)	(0.035)
WB4 Empresa 2009	-0.025	-0.033	-0.029
	(0.031)	(0.032)	(0.029)
MTI Other Road 2009	-0.002	0.002	-0.024
	(0.016)	(0.020)	(0.038)
MTI Other Bridge 2009	0.035	0.138***	
	(0.048)	(0.051)	
Observations	31,557	22,394	23,030

Wrk ever prior	Naive	Base	Uncontaminated
WB4 MCA 2009	0.016	0.019	-0.023
	(0.027)	(0.030)	(0.032)
WB4 Empresa 2009	0.010	0.007	0.018
	(0.025)	(0.025)	(0.034)
MTI Other Road 2009	0.015	0.009	0.075*
	(0.019)	(0.028)	(0.043)
MTI Other Bridge 2009	0.064**	0.025	
	(0.024)	(0.061)	
Observations	61,042	43,247	44,510

Job Last as Farmhand	Naive	Base	Uncontaminated
WB4 MCA 2009	-0.030**	-0.016	-0.007
	(0.015)	(0.017)	(0.015)
WB4 Empresa 2009	0.083***	0.078***	0.078**
	(0.028)	(0.029)	(0.030)
MTI Other Road 2009	0.015	-0.010	-0.019
	(0.018)	(0.021)	(0.017)
MTI Other Bridge 2009	0.006	0.013	
	(0.016)	(0.044)	
Observations	31,557	22,394	23,030

Note: Coefficients and standard errors are reported for OLS regressions on dummy indicators for the individual labor market outcomes: "whether worked in the past week", "whether worked in past year", " whether ever worked previously", "whether respondents last job was as a non-agricultural laborer", "whether respondents last job was as an entrepreneur", and "whether respondents last job was as a farmhand." For each outcome, regressions are run on three alternative Naive, Base, and Uncontaminated sample specifications. WB4 MCA 2009 is a dummy treatment indicator for municipios with WB4 supported MCA projects between 2005 and 2009. WB4 Empressa 2009 is a dummy indicator for municipios that have WB4 supported roadbuilding projects constructed by contractors with heavy machinery and not MCA teams between 2005 and 2009. MTI Other Road 2009 and MTI Other Bridge 2009 are dummy indicators for MTI road or bridge building projects between 2005 and 2009 that were not supported by WB4. All specifications use a full suite of household controls, as well as municipio and year fixed effects and report robust standard errors clustered at the municipio level. *** p<0.01, ** p<0.05, * p<0.1

Table 12: Individual Education Outomes: Matriculated within past 12 months

VARIABLES	(1) Pooled Treatment Matric. 7 to 18 yrs	(2) All MCA Matric. 7 to 18 yrs	(3) Completed MCA Matric. 7 to 18 yrs	(4) Completed MCA Matric. 7 to 12 yrs	(5) Completed MCA Matric. 7 to 12 yrs
WB4 Pooled 2009	-0.008				
	(0.031)				
WB4 MCA 2009		0.007			
		(0.037)			
WB4 Empresa 2009		-0.049*			
		(0.026)			
MCA Completed			0.054	0.063***	0.073***
			(0.037)	(0.018)	(.015)
Empresa Completed			0.010	0.125***	0.126***
			(0.021)	(0.011)	(0.011)
Male					-0.021***
					(0.006)
MCA Completed x Male					-0.017
					(0.035)
MTI Other Road 2009	0.033	0.024	0.027	0.016	0.016
	(0.025)	(0.027)	(0.024)	(0.016)	(0.015)
MTI Other Bridge 2009	0.037	0.038	0.041	0.015	0.016
	(0.066)	(0.066)	(0.066)	(0.018)	(0.018)
Observations	15,932	15,932	15,932	7,898	7,898
R-squared	0.062	0.062	0.062	0.070	0.072

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Regressions are estimated with "matriculation within the past 12 months" as the outcome variable for individuals aged 7-18 years in Columns (1)-(3) and 7-12 years in Columns (4)-(5), using the Base specification. Columns (1)-(3) respectively estimate impacts of the pooled project sample of all MCA projects and of only the MCA projects that were completed before the endline on all individuals age 7-18. Column (4) estimates impacts using the completed MCA and Empresa projects on the sub-sample of 7-12 year old individuals only, controlling for incomplete projects. Column (5) adds an interaction term for male individuals to the Column (4) specification. WB4 Pooled 2009 is a dummy treatment indicator for municipios with any WB4 supported projects between 2005 and 2009. WB4 MCA 2009 and WB4 Empressa 2009 are dummies for WB4 MCA built and contractor built projects between 2005 and 2009, respectively. MTI Other Road 2009 and MTI Other Bridge 2009 are dummies for MTI road or bridge building projects between 2005 and 2009 that were not supported by WB4. All specifications use a full suite of household controls, as well as municipio and year fixed effects and report robust standard errors clustered at the municipio level.

Table 13: Individual Health Outcomes: Incidence of Respiratory Illness and Medical Clinic Visits

VARIABLES	(1) All MCA Respiratory	(2) Completed MCA Respiratory	(3) Completed MCA Respiratory	(4) All MCA Med Access 14-45 yrs	(5) All MCA Med Access 14-45 yrs
WB4 MCA 2009	0.003			0.012**	0.020**
	(0.026)			(0.005)	(.009)
WB4 Empresa 2009	-0.058**			0.023**	0.023**
	(0.022)			(0.010)	(0.010)
WB4 MCA Complete		-0.047*	-0.009		
		(0.025)	(.017)		
WB4 Empresa Complete		-0.120***	-0.120***		
		(0.016)	(0.016)		
Male			-0.001		-0.009***
			(0.004)		(0.003)
WB4 MCA Complete x Male			-0.076**		
			(0.037)		
WB4 MCA 2009 x Male					-0.016*
					(0.009)
MTI Other Road 2009	0.048*	0.052**	0.052**	0.001	0.001
	(0.024)	(0.021)	(0.021)	(0.005)	(0.005)
MTI Other Bridge 2009	-0.019	-0.019	-0.019	0.005	0.005
	(0.031)	(0.030)	(0.031)	(0.013)	(0.014)
Observations	52,977	52,977	52,977	18,617	18,617
R-squared	0.020	0.020	0.021	0.010	0.013

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Regressions using the Base specification. For outcomes variables, Columns (1)-(3) use a dummy variable for self-reported incidence of respiratory illnesss and Columns (4)-(5) use dummy variables for whether or not individuals age 14-45 report using health facilities during times of good health (e.g. non-emergency primary care). Columns (1) estimates impacts of all MCA and Empresa projects started before the endline. Column (2) estimates impacts of only projects that were completed before the endline, controlling for incomplete projects. Column (3) adds an interaction term for male individuals to the Column (2) specification. Column (4) estimates impacts of all MCA projects and Column (5) adds an interaction term for male individuals to the Column (4) specification. WB4 MCA 2009 and WB4 Empressa 2009 are dummies for WB4 MCA built and contractor built projects between 2005 and 2009, respectively. MTI Other Road 2009 and MTI Other Bridge 2009 are dummies for MTI road or bridge building projects between 2005 and 2009 that were not supported by WB4. All specifications use a full suite of household controls, as well as municipio and year fixed effects and report robust standard errors clustered at the municipio level.

Figure 1: Nicaragua Map with Municipios, Departments, and Road Improvement Projects 2005-2009

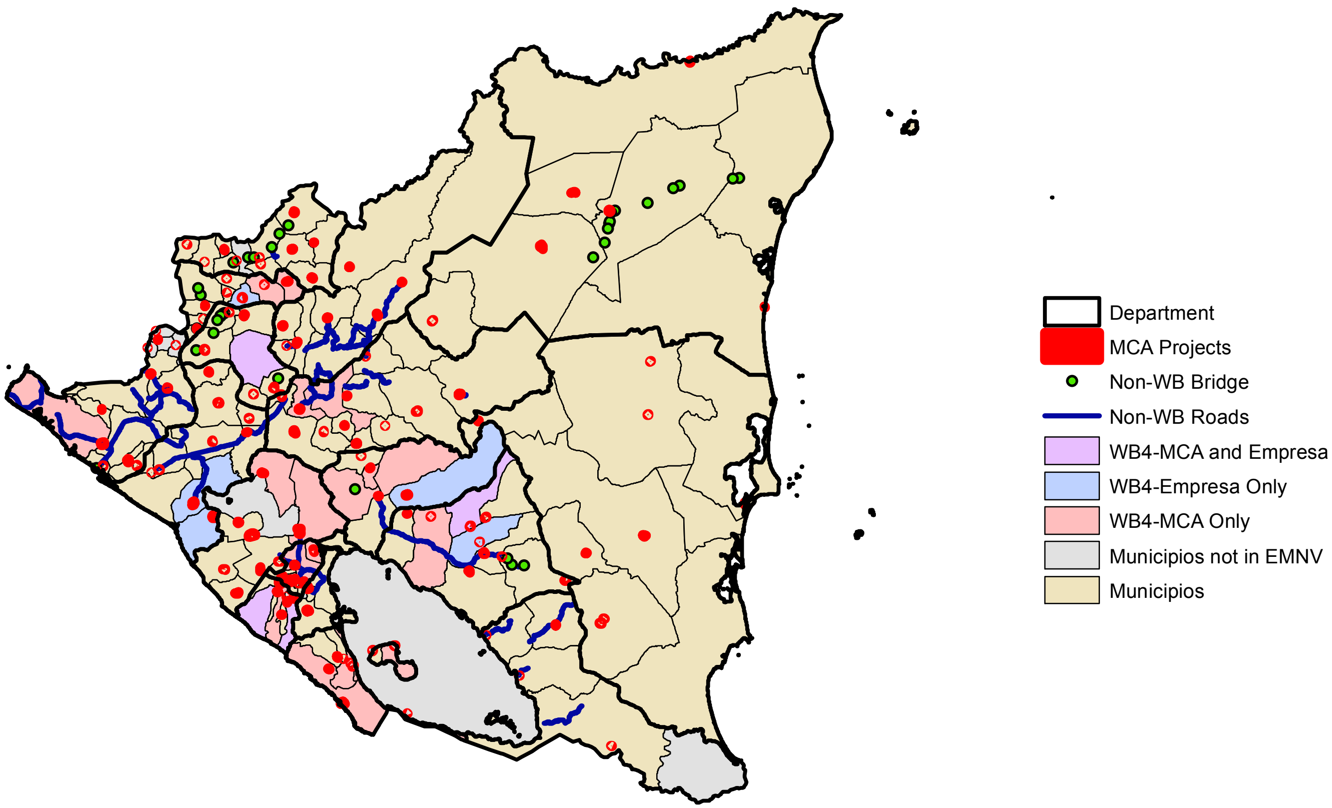


Figure 2: Nicaragua Map with Departments Dropped as in Base Specification

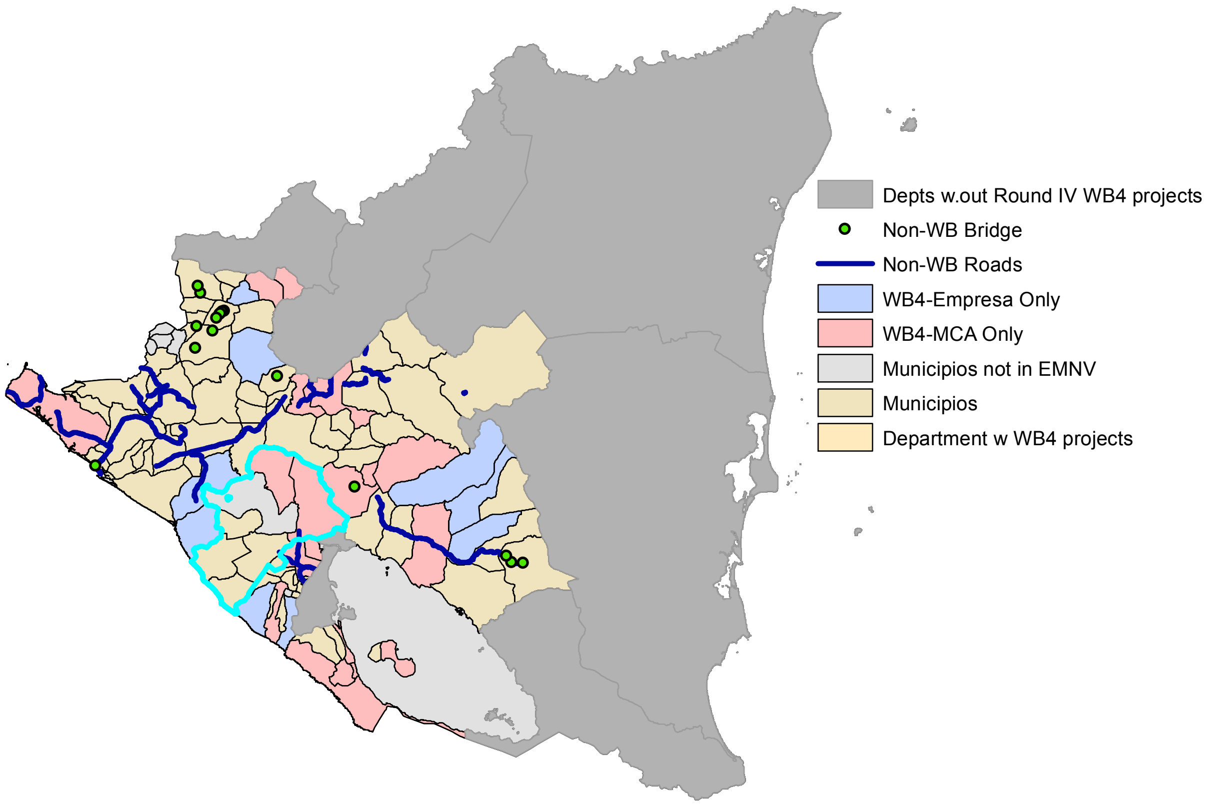


Figure 3: Treatment Exposure by Time (days), Road Length (km), and Cost (C$)

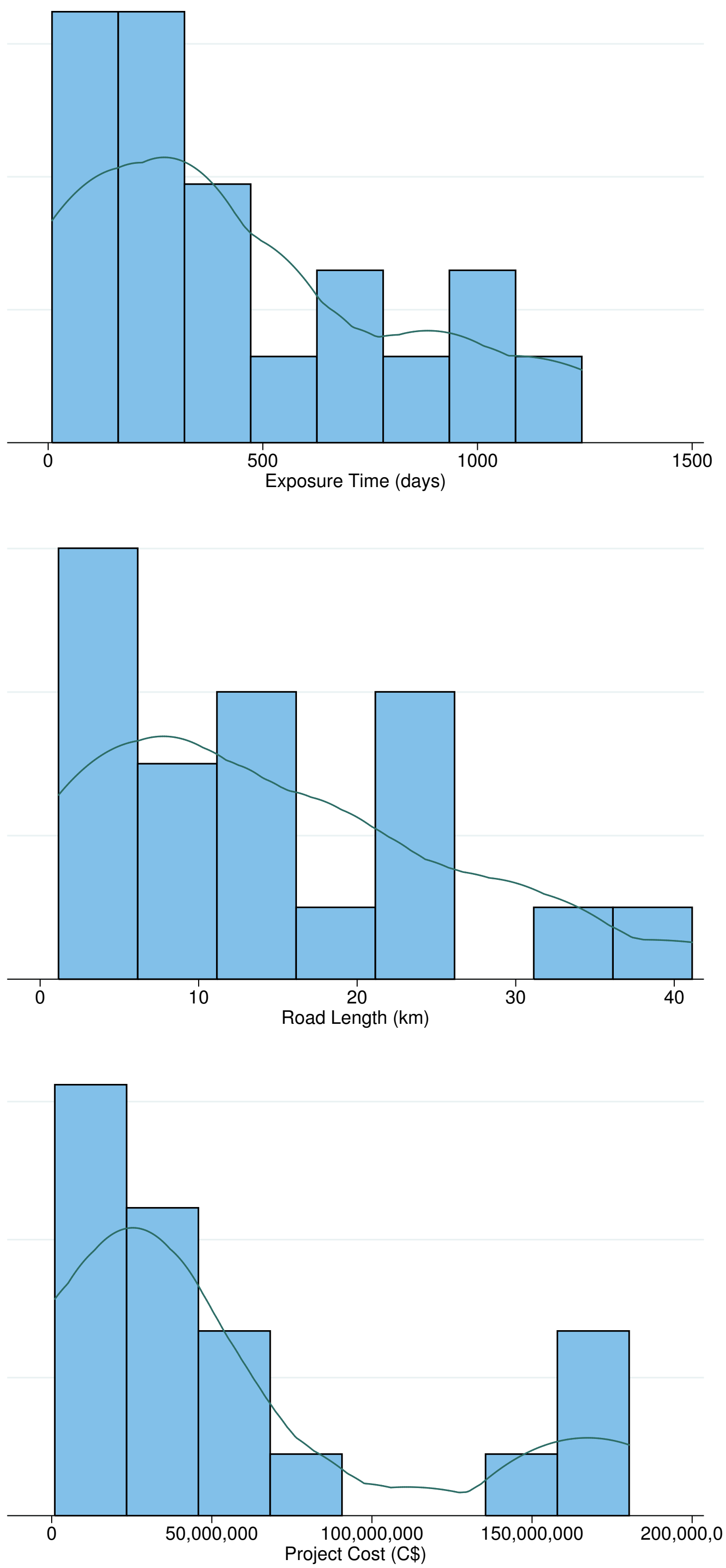
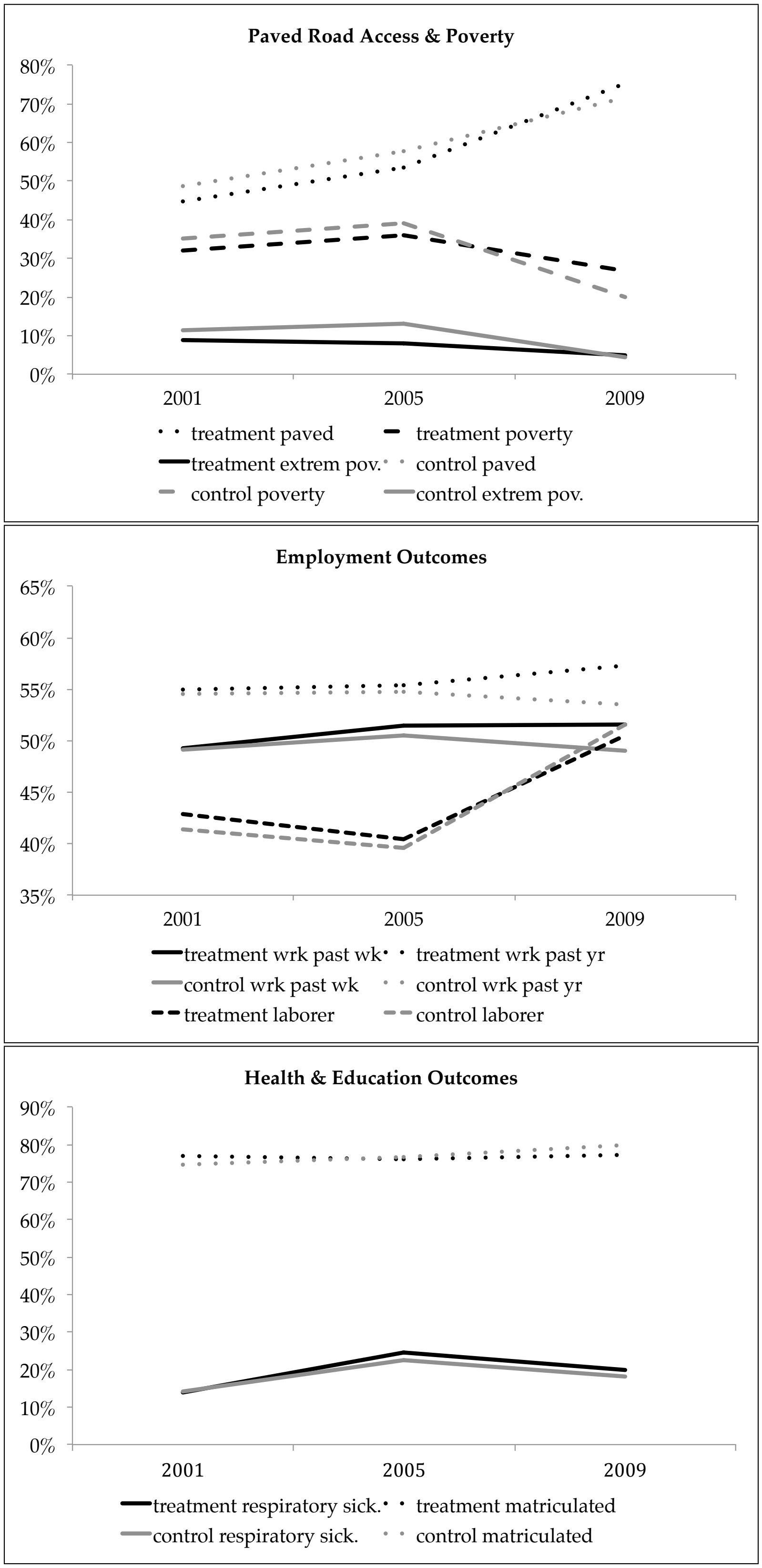


Figure 4: Trends in Key Outcomes over Time



Being Reminded to Save for College: The Impact of Inattention and Expected College Cost

Seth Garz

April 23, 2013

Abstract

This study investigates two potential mechanisms that mediate household decisions to save for distant future expenditures with particular relevance for education-oriented savings. Employing a randomized experiment in the natural setting of a school district-wide college savings program, the results of the experiment here support previous findings for an *attention effect*, i.e. that inattention to lumpy future expenditures may explain undersaving among intended savers. Among accountholders that had previously saved with the program, receiving a basic postcard reminder to save increased the likelihood of making a savings deposit within seven days by 10%. Extending findings from previous empirical work that has tested the attention effect on populations that select into savings programs, this study finds no effect for the +2400 accountholders that have never previously made a deposit, suggesting the attention effect is more appropriately an effect of *attention on the intentional*. Varying the messages of postcard savings reminders with information about either the availability of financial aid or the increasing cost of tuition allows this study to target a second mechanism - the effect of persuasive information about cost on household savings decisions. This investigation is of practical relevance for anticipating the effects of pervasive recent college tuition increases and related policy attempts to subsidize college costs, though the findings related to the *cost effect* are ambiguous and provide fertile ground for further investigation.

1 Introduction

Post-secondary degree attainment is associated with higher income and better employment outcomes, particularly among disadvantaged student populations (Bailey and Dynarski, 2011). The high returns to college may explain the common concern that steep growth in the cost of college threatens to reduce college access, in general, and further disadvantage students of low socio-economic status, in particular. Such concerns are largely consistent with the established literature on effects of price and financial aid on college attendance (E., 1997; Hemelt and Marcotte, 2011; Dynarski, 2003). To address concerns about college access, policymakers and education advocates have developed numerous policy interventions beyond traditional financial aid, one of which was exploited to conduct the experiment described herein.

Much of the extensive literature on college cost exploits variation in tuition and eligibility for financial aid to identify the effects of cost on college attendance and performance (Dynarski, 2003; Hemelt and Marcotte, 2011; E., 1997; Bettinger et al., 2009; Deming and Dynarski, 2009). The associated human capital models assume that these sources of variation affect the marginal cost of college, which induces variation in attendance, but are largely silent about the detailed behavioral and financial dynamics mediating the effects. College savings, student performance, and student debt are rarely observed, such that it is difficult to discern whether reducing college cost net of tuition expenses, financial aid subsidies, and college debt is effective because of its impact on marginal cost, student motivation, substitution away from time-costly part-time employment, substitution towards complementary college earmarked savings, etc.. This study contributes to the investigation of these mediating mechanisms, focusing on the effect of changes in college cost on college savings, in particular.

This study employed a randomized experiment to test the effects of postcard reminders to save for college on savings behavior. The experiment randomly assigned the full population of participants in a school-district wide college savings program to three treatment categories and one control. The treatment households received postcard reminders to save for college with messages emphasizing either the steeply increasing cost of tuition, the wide availability of financial aid, or a generic reminder to save. With data on the full deposit history before and during the experimental period for all accountholders in the program, this study compares the average likelihood of depositing and the average level of savings during the experimental periods across the the experimental groups. Comparing differences between groups that receive information implying higher college cost (tuition group) versus lower cost (financial aid group) enables the study to test how changes to expected college cost affect college savings.

Though the results prove to be ambiguous, this comparison is of relevance in anticipating how subsidizing college cost through financial aid or discounted tuition may affect household savings for college. Financial aid remains the primary means by which the federal government promotes college access in the United States with the government spending greater than \$33 billion on Pell Grants in FY2012 (Fain, Paul, 2012). State governments, in addition to providing state level financial aid, often promote college access through discounted tuition to public institutions. Given the scale of post-secondary education subsidies, even the weak evidence provided below of a negative savings effect from providing financial aid information begs key policy related questions:

Does financial aid discourage savings for college?
Classical consumer models would anticipate that financial aid, through its direct effect on net college cost, should reduce savings at least at the intensive margin among those that are currently saving. Alternatively, models that include uncertainty or psychological factors might anticipate increases in savings among both marginal and infra-marginal savers who believe that they are more likely to be able to afford and attend college, net of financial aid.

Can public programs mobilize college-oriented household savings?
In the education context specifically, better understanding household savings behavior may inform the design and calibration of programs to promote college access. Furthermore, as hypothesized by some researchers and advocates of asset-based economic development strategies, the presence of assets and savings, in particular, may be instrumental in improving student academic performance, educational aspirations, and college-going identity (Sherraden and Sherraden, 1991; McKernan and Sherraden, 2008).[1] This *instrumentalist* argument will not be tested here, but at least implicitly, if not explicitly, provides the theoretical foundation for current policy experiments to mobilize college savings through college savings accounts.[2] The administrators of one such experiment in San Francisco were close collaborators for the experiment described herein.

Beyond the educational and local experimental context, better understanding household savings behavior is of general economic interest and an active focus of ongoing research. Recent work on default savings options, savings technologies, and savings reminders lend support to the theory of "mental accounting" (Thaler, 1990; Dupas and Robinson, 2011; Ashraf et al., 2006), which refines the traditional assumption that money is fungible. A variety of models have been developed and tested to describe the stylized fact that individuals often report undersaving for expenditures as normal as retirement (Benartzi and Thaler, 2007). In a survey of the literature on persuasion, DellaVigna and Gentzkow (2010) categorize alternative models of persuasive communications. Two basic distinctions are made between belief based models in which a) rational Bayesian receivers update their beliefs in the context of persuasive information or b) persuasive information enters a receiver's utility function directly and effects her preferences independent of beliefs. Both classes of models have been used to explain undersaving among low income individuals.

Within the class of preference based models, Banerjee and Mullainathan (2010) propose a model of *temptation goods* in which time-inconsistent individuals have two different utility functions for temptation goods consumed in the present versus in the future. The authors use this model to explain why rational, hardworking poor individuals may consume non-productive goods, such as cigarrettes, and remain chronically poor if the percentage of one's

[1]See the Corporation for Enterprise Development (CFED) website `http://cfed.org` for research and advocacy documents related to asset-building policy.

[2]The Saving for Education, Entrepreneurship, and Downpayment (SEED) piloted asset-building accounts for children and youth in 12 communities, including 1,200 low-income children and their families, from 2003 to 2008. The experience with SEED informed the development of the Kindergarten-to-College program and a new randomized trial of college savings accounts for high school students sponsored by the U.S. Department of Education through the GearUp program.

personal budget spent on temptation goods naturally declines with income.

Within the belief based class, limited attention models have been developed to explain behavior as diverse as voting, retail consumption, and stock picking. Karlan et al. (2011), which will be the basis for the model developed herein, proposes a model that differentiates between regularly occuring consumption that is perfectly attended to in every time period and lumpy expenditures that may be periodically forgotten. The limited attention to lumpy expenditures explains undersaving in the simplified setting where a utility maximizing actor would perfectly smooth consumption. Among individuals that select into savings programs run by retail banks, Karlan et al. (2011) finds that reminders can significantly increase savings amounts by as much as 6% and are more effective when the intended use of the funds is mentioned. Karlan et al. (2011) argues that a variety of findings on personal savings associated with the mental accounting framework, such as the effects of labeling, non-binding commitments, default options, etc. may be explained by the mechanism of inattention. This study closely builds on the model and the empirical methods of Karlan et al. (2011). Consistent with Karlan et al. (2011), the model here depicts college costs from the perspective of guardians of current period kindergarteners as a distant lumpy expenditure.

In addition to modeling attention effects on a general non-selected population, this model and this study seek to contribute to the empirical literature on persuasion by addressing: can persuasion effect personal savings? Generally stated, the question is not terribly interesting as it is hard to imagine that under no circumstances would providing new information ever effect personal savings. The case considered here, however, is of more specific, practical and policy relevance. In this experiment, the persuasive messages were not deceptive advertising, but were factual with external relevance for simulating the real direction of the effect of changes in tuition and financial aid availability on household college savings. Thus, this experiment addresses the micro-behavioral dynamics of households facing changing financial aid and tuition environments, focusing on the savings mechanism in particular. To foreshadow a prediction of the model, the rank of outcomes among the various treatment groups will be used to test whether college cost information has a persuasive "cost effect" on savings.

Two other studies provided key methodological inspiration for this experiment. Bhargava and Manoli (2012) reinforces findings on the potential impact of reminders with an experiment involving IRS reminders to claim the Earned Income Tax Credit. The study is methodologically similar in employing mail-based reminders from a public agency targeting low income Californians. Key differences from this study include the fact that Bhargava and Manoli (2012) specifically targeted individuals who were likely eligible for, but had not claimed the public benefit and that the public benefit was often quite substantial (as much as $5,657) and did not require any personal financial contribution. Notably, the results do not clearly confirm the existence of the inattention mechanism as the main findings emphasize the importance of informational simplicity and display of program benefits, with practically no effect for a basic reminder. The social psychological study Destin and Oyserman (2009) inspired this study's use of persuasive college cost information. In Destin and Oyserman (2009) researchers read information about either the high cost of college tuition or the availability of financial aid to randomly selected groups of seventh grade students from an underperforming inner-city school. They find that, compared to the college tuition exposed group or a control group that received no new information, students that are read financial aid information report higher expected grades in their class and anticipate spending more time on a homework assignment. Similiar to Destin and Oyserman (2009), the messaging of two of the postcard treatments in the experiment described here emphasizes tuition versus financial aid. However, the fact that it is the guardians of students that are targeted in this study and that the outcome of interest is savings behavior, not self-reported perceptions, renders this experiment substantially different.

Findings from this study reveal clear results for the attention effect and ambiguous results for the cost effect potentially due to lack of statistical power. Consistent with an attention effect, the study's main finding confirms prior evidence that reminders increase savings amounts and increase the probability of savings. Within a week following receipt of the reminder among households that have previously saved (the intensive margin), those that receive a reminder are 9.7% more likely to save than those that do not. This effect is equivalent to increasing the probability of savings across the entire sample of existing program enrollees by approximately 1.5%. Extending prior empirical work with a general population of accountholders that is not limited to a sample of individuals that select into a savings programs, this study finds no effect of savings reminders on existing program enrollees that have never saved (the extensive margin). The attention effect is only relevant for accountholders that have previously demonstrated a willingness to save, which herein is referred to as the effect of *attention on the intentional*. Findings for the cost effect are ambiguous. Among new program enrollees, differences in the effects of introducing information about increasing tuition expense versus widespread financial aid subsidies are weakly consistent with the predictions of a simple persuasion model. Among existing enrollees that have previously contributed to their college savings accounts, there does not seem to be a cost effect. Collapsing attention and cost effects into a general reminder effect on the whole population of accountholders, not just those that had previously saved, reveals weak evidence of a main average effect for reminders, which is largely explained by the significant effect of reminders on the Chinese subset of accountholders. The ambiguity in some of the results seems to suggest that the experiment was statistically underpowered likely reflecting the low rate of participation among enrollees.

The paper is structured as follows. Section 2 provides an overview of the savings program, account structure, incentives, and program implementation. Section 3 describes the experimental design, reminder messages, and data. Section 4 develops a model, clarifies key assumptions, and presents testable model predictions. Section 5 provides summary descriptions of the accountholders and savings data. Key empirical considerations are described in Section 6 and results are presented in Section 7. Main results are presented for a general reminder effect on the full sample and on sub-samples by language group, and specific mechanisms are explored by testing the model's predictions.

2 Savings Program Description

Kindergarten to College (K2C) is a college savings program administered by the Office of Financial Empowerment (OFE) within San Francisco's Office of the Treasurer in collaboration with the San Francisco Unified School District (SFUSD).[3] By default, OFE opens a savings account for each new kindergartener incoming to SFUSD unless a student's guardians opt out of the program, which less than one percent of families do.[4]

In addition to opening accounts, OFE seeds new accounts with a base amount of $50, which is increased to $100 if a student is enrolled in the federal Free and Reduced Lunch program (FRL) and signs a consent form allowing release of FRL data from SFUSD to OFE. OFE matches family contributions one-to-one up to a lifetime cap of $100 and offers a $100 bonus if families contribute at least $10 for six successive months - called the "Save Steady" bonus. Anyone may contribute to a student's account online, by mail, or via a Citibank retail branch up to an annual account cap of $2,500 per year. The account is earmarked for post-secondary education related expenses, including college tuition, in accordance with federal qualified 529 account rules. Deposited funds earn the prevailing (very low) retail banking savings account interest rate. Similar to 529 accounts, income from K2C accounts in the form of interest is not taxed. Unlike 529 accounts, however, K2C accounts do not affect eligibility for public assistance or college financial aid, and are not considered assets of the family. Withdrawals must be made prior to the student turning twenty-five years old. Thereafter, remaining funds will be remitted to the student excluding any match or bonus funds. Further discussion of the accounts and incentives is provided in Appendix A.

Figure 1 illustrates how K2C was rolled out in three phases with 25% of schools with Kindergartens eligible in the school year beginning in 2010, 50% in 2011, and 100% in 2012. Prior to universal enrollment among all schools with Kindergartens in Phase 3 (2012), OFE prioritized schools with higher rates of students enrolled in FRL (a common proxy for socio-economic status) while ensuring that at least one school from each San Francisco Supervisor district was included in each wave. The sequence of this rollout implies that at the time of the experiment some students in the second grade had maintained accounts for over 18 months, some first graders had access to accounts for over one year, and some students (approximately 50% of the sample) had received access to the accounts just before or after round one of the experiment. This variation in experience with the account will be meaningful for testing the predictions of the model below.

3 Description of the Experiment

3.1 Overview

There were three rounds to this experiment. The first two rounds sought to use monthly mailed postcard reminders to influence guardians' expected cost of sending a student to college in the future net of tuition and financial aid. The third round took place less than a week prior to Christmas Day with postcard messages encouraging families to give their children the gift of college savings for the holidays with an emphasis of either individual giving or families and friends giving depending on the treatment. The subsequent analysis

[3]Kindergarten to College. http://k2csf.org/

[4]OFE staff reported this percentage estimate, which is broadly consistent with data on accountholders and publicly available censuses on SFUSD students.

will focus on the college cost messages in round one and two, though savings following round three may reflect the interaction effects of cost messages and holiday messages. All rounds of the experiment targeted guardians of K2C enrolled students that received monthly mailed reminder postcards issued by OFE. Postcard reminders were an existing feature of OFE outreach activities and families from the first two waves of K2C had received numerous prior postcard reminders. Further discussion on the relevance of postcard reminders, mailing procedure, and messaging in the third round of the experiment is available in Appendix B.

3.2 Postcard Messages

In the first two rounds, targeted guardians were randomly assigned to one of four experimental categories, including three treatments and one control. Images of the reminder postcards are included in Appendix D for reference. Two treatment categories were designed to influence the recipients expectations of future net college cost. The front of the tuition treatment postcard read:"Tuition is rising steeply and can substantially increase the cost of sending your child to college" with facts about the rate of tuition increase in California colleges as well as facts about the K2C program on the back. The front of the "financial aid treatment" postcard read: "Financial aid is widely available and can substantially reduce the cost of sending your child to college" with facts about the availability of financial aid in California as well as facts about the K2C program on back. The experiment hypothesized that even such limited information could effect behavior in light of very low financial aid application rates among eligible families in California (Jackson, Orville, 2013).[5] Both tuition and financial aid messages were composed to be reference independent, suggesting the direction of changes in college cost, but not the actual cost levels. This was thought to be important given the likely heterogeneity of prior college cost expectations among families. The third treatment category was designed to be a generic reminder and to avoid making college cost salient to the recipient. The front of the "generic treatment" repeated the simple K2C program slogan "Save Steady. Dream Huge," which had been featured on the front of all reminder postcards mailed prior to this experiment. The back of the generic postcard only included K2C program facts. Those assigned to the fourth experimental category did not receive postcards. In all rounds the postcards were distributed in English, Spanish, and Chinese according to the preferred language previously indicated by the guardians. Effort was made to construct each of the postcards to conform as much as possible in design, color, grammar, word use, etc. across treatment categories and languages.

3.3 Program Waves

The first round of the experiment included all individuals from Wave 1 and Wave 2 with valid addresses as well as Wave 3 students that receive English language materials from K2C, which constituted over half of Wave 3 students and are referred to as "Wave 3a." Newly incoming Chinese and Spanish speakers were not included in the first round of the experiment, but were included in the second and third, and are subsequently referred to as

[5] California Financial Aid Tracker. `http://financialaid.edtrustwest.org/`

"Wave 3b." The year of the experiment was the first year that the entire cohort of incoming kindergarteners was included in K2C, providing a massive increase in the administrative burden on the program. English speaking families had only just received welcome packets and bank cards necessary to begin depositing in their accounts a couple of weeks prior to round one of the experiment and those materials had not yet been translated into Spanish and Chinese. Thus, the first round of the experiment could only include part of Wave 3 - specifically Wave 3a. Wave 3a was assigned an identical treatment category in round two in which Wave 3b was first included.

Randomization was stratified on schools, existing enrollees (Wave 1 and Wave 2) vs. new enrollees that had not previously received reminders (Wave 3), and savers vs. non-savers in the existing enrollee cohort.

3.4 Rounds of Experiment

Figure 2 illustrates the experimental design over time across waves and experimental rounds. Postcards from the first round were mailed and postmarked on November 7, 2012. Thirty-three days later, postcards for the second round were mailed and postmarked on December 11, 2012. Ten days later, postcards for the third and final round were mailed and postmarked on December 21, 2012. The final deposit data available is for January 11, 2013, twenty-one days after the third round was mailed. Notably, the window of timing following the mailings in which deposits could have accumulated ranges from ten to thirty-three days.

3.5 Data Description

The author had access to the full anonymous deposit history for all K2C accounts from the start of the program in April 2011 through January 11, 2013 as well as a variety of individual non-identifiable student covariates, such as school, wave, gender, language preference, etc..

4 Model

The following model is largely based on the attention model in Karlan et al. (2011), which falls within the belief class of persuasion models described in DellaVigna and Gentzkow (2010). A simple extension of the model is developed to consider selection and program salience with three distinct predictions that can be empirically tested with available data.

4.1 Extension of the Basic Attention Model

As in Karlan et al. (2011), this model assumes consumption is divided between regular expenditures in every period, c_t, and lumpy expenditures that occur infrequently, x_t. Here the model will consider one distant lumpy expenditure, which will represent the expected future cost of college tuition in the fourteenth year for a household's kindergarten child, $x^{\star} = E[College]$. Assuming no initial or final period wealth, constant income, and no discounting [6], a household will choose current period consumption, c_t, to optimize lifetime indirect utility, $v(c_t, x_t)$, subject to a budget constraint:

$$\max\, v(c_t, x_t) = \sum_{t=1}^{T} u(c_t) + u(x_T^\star) = u(c_1) + u(c_2) \cdots u(c_{14}) + u(x_{14}^\star) + u(c_{15}) \cdots + u(c_T)$$

$$\text{s.t} \sum_{t=1}^{T} y_t = \sum_{t=1}^{T} c_t + x_T^\star$$

Assuming the utility function $u(c)$ is concave with respect to consumption and the lumpy college expenditure $x^\star$ is fixed in time (in the fourteenth period with respect to college), reveals the well establish result that households will set marginal utilities equal across periods by smoothing consumption:

$$u'_{c_1} = u'_{c_2} = \cdots = u'_{(c_T)} \text{ s.t } \sum_{t=1}^{T} y = \sum_{t=1}^{T} c_t + x^\star \Longrightarrow y - \frac{x^\star}{T} = \bar{c}$$

internalizing savings, s, in current period consumption decisions equal to the smoothed lumpy expenditure as:

$$s^\star = \frac{x^\star}{T} = y - \bar{c} \tag{1}$$

Following Karlan et al. (2011), the model can incorporate inattention to explain why individuals may save less than intended. Let $\theta \in [0, 1]$ represent the probability of attending to the future lumpy expenditure where θ will increase on average if you are mailed a savings reminder, $D_i \in \{0, 1, 2, 3\}$ where $D_0 =$ the control, such that:

$$\theta = f(D_i) \text{ and } f'_{D \neq 0} > 0$$

Thus,

$$s^\star = \theta(\frac{x^\star}{T}) \text{ and } s'_{D \neq 0} > 0 \tag{2}$$

Thus far, the model has treated the lumpy expenditure as unavoidable. In a more general case, one could include an uncertainty parameter that captured how confident a household was that their child would one day attend college and incur the cost of college. This uncertainty parameter could be modeled as a function of information about college cost. However, this approach with uncertainty weighted cost of the lumpy expenditure presents the empirical challenge of decomposing the effect of interventions on expected cost versus the effect on the probability of incurring the cost. Fortunately, in the experiment described herein, there are strong disincentives for saving among those that are uncertain about their children attending college. K2C accounts are legally earmarked for college and the child, not

[6] Assuming no discounting simplifies the explication of the model. However, incorporating basic discounting would not meaningfully alter the predictions of the model.

the parents who are the depositors, redeem the account balance should the child not attend college. Thus, it seems reasonable to assume away this uncertainty effect.[7]

As mentioned above, the salience of savings opportunities may mediate the effect of increased attention for future lumpy expenditures. Slightly departing from the DellaVigna (2009) formulation in which salience is endogenous to the attention function, the model here separates a salience function from the attention function. Importantly, salience in this case does not refer to the salience of the messages.[8] Rather, salience here refers to the salience of the K2C savings program, which may alternatively be understood as desire to select into saving with K2C. Salience in this selection sense will be empirically identified from whether or not existing accountholders have ever saved.[9] Therefore, salience here is simplified to an indicator function that equals one for households that expect children to attend college with high likelihood and equals zero for families who have less certainty about the college attendance of their children, such that:

$$s^{\star} = \mathbb{I} * \theta(\frac{X^{\star}}{T}) \tag{3}$$

This seemingly trivial extension of the equation (1) is important for considering the differential effects of reminders on increasing savings for the general population versus a sample that selects into a savings program. In the context of K2C which universally enrolls students (with less than one percent opting out) prior private contributions to the K2C account provide a proxy for families that may have selected into a similar savings progam if it were structured as an opt-in rather than an opt-out program. Differently stated, the absence of even one private deposit into the K2C accounts among existing enrollees from Wave 1 and Wave 2 serves as a good proxy for which households do not find the K2C opportunity salient and would not have selected into a similar progam. This insight yields the model's first prediction:

[7]Assuming this uncertainty effect may seem to directly contradict one the findings from Destin and Oyserman (2009) that originally inspired this investigation. However, the Destin and Oyserman (2009) findings do not seem externally relevant for the savings choices of parents for a variety of reasons. In that study, it is children (seventh graders) that are targeted with tuition and financial aid messages and the outcomes observed are self-reported estimates of the grades they would receive in a class and the time they intended to invest on a homework assignment. It seems reasonable to believe that college cost information could change a "naive" child's self-image from that of "non-college going" to "college going" while leaving a parent's estimated probability of a child's college-going future uneffected. This discussion reveals the nuanced theory of change implicitly underlying asset-building policy arguments to encourage college savings. The presence of savings may encourage and promote achievement in children. But, the possibility of achievement may not be sufficient to overcome parent uncertainty and motivate parents to incur the opportunity cost of additional saving on behalf of their children.

[8]The messages were designed carefully to look almost exactly the same and share the same syntax. The belief of the author was that, if anything, the availability of financial aid message should be more salient in the sense of attention enhancing (i.e. salience endogenous to the attention function $\theta(salience, D_i)$) than the increasing cost of tuition message. Thus any confounding of salience on the identification of a persuasion effect should be to attentuate the difference between financial aid and tuition groups.

[9]Neither "salience" nor "selection" as used in the existing persuasion literature correctly capture what is referred to as salience of the savings program here. Elsewhere, selection refers to selection into demand for persuasive information and salience refers to the attention-enhancing effect of persuasive information.

Prediction 1a - Attention Effect on Prior Enrollees: Among Wave 1 and Wave 2 families that have already been exposed to the opportunity to save through K2C, the"attention effect" of reminders should only impact those accountholders that have previously saved. Among those accountholders that have previously saved, the generic reminder message group should save more and be more likely to save than the control group on average (the *intensive margin*). In contrast, generic and tuition groups among accountholders that have not previously saved should not be effected by treatment assignment (*the extensive margin*).

Prediction 1b - Attention Effect on New Enrollees: Among Wave 3 accountholders that save during the experimental period, those accounts assigned to the generic group should have higher savings on average than those accounts assigned to the control group.

4.2 Changes in Expected Future Lumpy Expenditures

Of central interest to this experiment is the question of whether providing persuasive information that may vary expected future costs of college affects current period savings. Equation (3) can be modified to incorporate the receipt of different types of information, D_i, on the expected cost of future lumpy expenditure:

$$s^\star = \mathbb{I} * \theta \left(\frac{E\left(X^\star | D_i\right)}{T} \right) \tag{4}$$

To simplify the comparative statics, the expected cost can be represented by a well-behaved monotonically increasing function $x(\cdot)$ of cumulative information about future cost such that:

$$s^\star = \mathbb{I} * \theta \left(\frac{E\left(X^\star | D_i\right)}{T} \right) = \mathbb{I} * \theta \left(\frac{x(D_{-i_{t-1}} + D_{i_t})}{T} \right) \tag{5}$$

where $D_{-i_{t-1}}$ represents an accountholder's prior expectation of future college cost and D_{i_t} can be positive or negative. The effect of changes in the expected cost of college on savings then depends on the effect of $\theta = f(D_i)$ on savings and the updating of the cost estimate following receipt of new cost-relevant information:

$$s'_{D_i} = f(D_i)x'_{D_i} + f'_{D_i}x(D_i) \text{ with } f(\cdot) \geq 0, x(\cdot) > 0 \tag{6}$$

For interventions that suggest college is lower cost than previously expected, the accountholder will either not save or save less than with a generic reminder depending on the extent to which the negative impact of the updated cost estimate does or does not outweigh the positive impact of increased attention. In either case, the savings impact of interventions that suggest college is more expensive than anticipated should increase savings more on average than interventions that lower expected costs.

The claim required for later empirical identification is that the tuition and financial aid reminders should have opposite effects on the expected cost of college with messages and facts about "steeply rising college tuition" implying higher net cost, and messages and facts about "widely available financial aid" implying lower net cost. Notably, the model does not require savers to hold accurate beliefs about college cost to be influenced in a predictable way by new information. The experimental messages were carefully crafted to increase or decrease accountholders expected costs of college regardless of the reference point of their prior belief (though the actual interpretation of the messages cannot be qualitatively verified because of restrictions on communicating with progam participants). If reminders are randomly assigned, prior cost estimates and levels of inattention should be equal on average between reminder groups. For those accountholders that do not expect to go to college, the different messages should have no effect as modeled by the indicator function, which zeroes out any potential impact of updating on savings. An alternative interpretation of this indicator function is that some households may never expect to use K2C accounts, but may have another pool of savings elsewhere. Complete household balance sheets are rarely observed, so this alternative interpretation is not merely a semantic difference, but may be important for interpreting null effects of reminders. These insights lead to the model's second set of predictions:

Prediction 2: Among Wave 1 and Wave 2 prior savers and new Wave 3 enrollees, the tuition reminder should have a larger positive effect on savings than the generic reminder, which should have a larger effect on savings than the financial aid reminder. The financial aid reminder may have a positive or negative effect compared to the control group.

5 Descriptive Analysis

5.1 Verifying Randomization

The following analysis focuses on the randomized college cost reminder experiment, which contained two rounds with three treatment groups and one control in each round. New Wave 3b participants were absorbed into the experiment in the second round balanced across all experimental categories, so round two treatment assignment is used in the subsequent analysis. Treatment assignment of the first round participants remained stable through the second round. The control group remained the control group across all rounds, though it grew to absorb new enrollees in round two.

Summary statistics by round two experimental category in Table 1 using pre-intervention data confirm that the randomization was successful. The t-statistics in column (1) compare the control group to the pooled treatment groups. The t-statistics in columns (2) through (4) compare each treatment group to the control group respectively. None of forty-eight tests find differences in means at even the ninety percent confidence level. There are only limited descriptive variables available for each participant, including whether or not accountholders saved in the pre-experimental period, are male, consented to disclosure of their FRL status, provided mobile phone number, provided email contact, and prefer to receive materials in Spanish or Chinese. The final variable, Regular Saver, is a proxy for whether or not the accountholder is a frequent saver or potentially has enrolled in direct deposit. The dummy is turned on if an accountholder has saved three consecutive times within 45 days, including at least once within 45 days prior to the first round of the experiment. This variable will prove to be a significant covariate in later regressions, but should not bias our results. All of the variables described except for pre-experimental cumulative savings in row one are dummy variables. Tests for round one and round three are omitted, but similarly confirm the success of the randomization.

One detail worth noting is that the rank of means of pre-experimental cumulative savings, Tuition > Generic > Financial Aid > Control, is consistent with the predicted rank addressed by Prediction 2 of the model, though the prediction is only meant to apply to the sub-group of Wave 1 and Wave 2 prior savers that are included in Table 1. Thus, regressions with Wave 1 and Wave 2 accountholders will control for pre-experiment cumulative savings. Attrition is not an issue and their were only minor data issues, but further discussion on attrition and data is available in Appendix C.

5.2 Descriptive Statistics

Accountholders varied among waves along key characteristics. Table 2 Panel A provides summary statistics for the entire experimental sample by K2C cohort wave. Wave 3 is divided into Wave 3a, which was enrolled immediately before round one of the experiment, and Wave 3b, which was enrolled just after round one. The number of observations approximately doubles with each wave, reflecting that K2C was phased in across SFUSD over the course of three years approximately doubling the number of schools included in each year. Consistent with the program description above, Wave 3a is exclusively accountholders that prefer English to Spanish or Chinese (i.e. not necessarily native English speakers, but not Spanish or Chinese speakers) and the vast majority of Wave 3b prefer Spanish or Chinese. To the extent that providing an email address to the school district is a proxy for socio-economic status, the increase in Email Info Provided across waves reflects that schools with on average lower income families were prioritized in the program rollout. The drop in percentage preferring Spanish between Wave 2 and Wave 3 likely reflects the higher average socio-economic status of Wave 3 accountholders. The low percentages of accountholders enrolled online reflects both the low level of active participation and the fact that families can contribute through retail banking branches or by mail, i.e. are not required to register online to make deposits.

Given the low level of participation in K2C, it is useful to describe the sub-sample of those that save at least once ever and compare them to those that at least save during the experimental period. Table 2 Panel B and Panel C provide descriptive statistics for these sub-samples of the K2C population: those accountholders that ever made deposits with K2C and those that made deposits at least once after the introduction of the experiment. Chinese speakers are overrepresented in both samples, constituting only 19% of the K2C population, but 34% of those that ever save. Spanish speakers are under-represented, constituting 23% of the K2C population, but only 13% of those that ever save. These trends by language group are even more pronounced among savers during the experimental period and confirm anecdotal evidence that Chinese speakers are the most active participants. The 44% of accountholders that signed FRL disclosure consent forms were much more likely to save, constituting 89% of savers. This may reflect that the poorer families that qualify for FRL are more sensitive to savings incentives or that only the most motivated savers choose to sign the somewhat complicated consent form. That SFUSD district-wide FRL rate was 61%

in the 2010-11 school year suggests that approximately half of FRL enrolled families signed consent forms in Wave 3, for example (CA Dept. of Education, 2013). Those that registered online only represent 6% of the accounts, but constitute 62% of accounts that ever save and 69% percent of accounts that saved during the experiment. This could reflect the ease of depositing online or, most likely, that those who choose to register online are the most eager to participate. Accountholders that provide email information are more likely to be savers and even more likely to save during the experimental period.

6 Empirical Considerations

6.1 Identifying Outcomes of Interest

To assess the effects of reminders and reminder messaging on savings behavior, the following analysis considers the probability of saving, level of cumulative savings, and log of cumulative savings in the experimental period as the key outcomes of interest. Each of these outcomes has it own shortcomings. Figure 3a reveals that the distribution of cumulative savings in levels is highly skewed right even when dropping the zeroes of those that do not save. Including the zeroes, which make up the vast majority of the sample, only amplifies the skewness of the distribution. Taking log's of the cumulative savings level normalizes the distribution of the long tail, but presents other problems in that a dollar must be added to each account balance to correct for the zeroes. Even more problematic is the fact that the minimum contribution to a K2C account is $10, creating a gap in the interval 0.0-2.3, as revealed in Figure 3b. For this reason, the dummy indicator for saving in the experimental period may provide the most consistent outcome of interest. Only reporting the probability of saving, however, would preclude the possibility of estimating the effect of different reminder messages on the magnitude of savings. This limitation would constrain analysis of the *cost effect* hypothesis (Prediction 2) if financial aid reminders demonstrate a positive *attention effect* that is not fully offset by the negative *cost effect*. In lieu of delving into more complex econometric estimators that each have their own limitations, multiple outcomes of interest are reported in the tables comparing means among treatment groups. For regressions with probability of saving as the outcome of interest, OLS regression are reported as results from non-linear estimators are almost identical. In the results, level, log, and likelihood are reported by rounds one, two, and three separately, for the "Cost Experiment," which consolidates rounds one and two, and for "Both Experiments," which consolidates all three rounds. In some cases, results are estimated from only the sample of deposits made within seven days following the postcard mailings. In other cases, the full population of deposits over time are used.

6.2 Considerations for Testing Model Predictions

A key assumption necessary for empirically testing Prediction 1a is the claim that, among previously enrolled accountholders (Wave 1 and Wave 2) the absence of any prior deposits in the K2C account is a good proxy for disinterest in ever utilizing the K2C savings program - what the model referred to as "salience" of the K2C program. At the more extreme, though not necessary for internal validity here, this indicator could be interpretted as a disinterest in or severely discounted expectation of one's young child ever going on to post-secondary education. There are a number of reasons why this claim is reasonable. It is unlikely that families are not sufficiently informed about the K2C program to be able to learn more if interested. By the time of the experiment in November 2012, Wave 1 accountholders had been enrolled in K2C for over a year and a half since approximately April 2011 and Wave 2 accountholders for at least a year since November 2011. K2C actively reaches out to SFUSD families, teachers, and administators through promotional events at schools, mailings, email, etc. Wave 1 had been mailed at least twelve reminder postcards since July 2011 by the time of the experiment, not to mention numerous consent forms and account statements. Nevertheless, the likelihood that families had at least heard of the program and received numerous K2C mailings, does not undermine the potential attention effect. Prior to the experiment the last reminder postcard was distributed months earlier with the majority of mailings targeting new enrollees. The earlier waves are also less likely to have provided an electronic address for regular email updates as indicated in Table 2.

7 Results

7.1 Average Treatment Effects

Before testing the detailed predictions of the model, it is worth pooling waves, prior savers, prior non-savers, and new accountholders to estimate the average treatment effects. From the perspective of policymakers examining the cost and benefits of mailing reminders, the effect pooling all accountholders is the key outcome of interest. Table 3 reports the outcomes of interest by experimental category for deposits made within seven days of mailing each round of reminders. A seven day window was chosen arbitrarily as a reasonable period of time within which reminder effects should be observable and after which reminder effects should likely fade.[10] Column (1) reports the average for the control category and reports t-statistics for the control compared to the three treatments pooled. The average pooled treatment is significantly larger than the control in log and likelihood across many outcomes except for those of round one and round three. These treatment effects are largely explained by the generic and tuition reminder groups with the greatest effect for the tuition group. The tuition group is significantly more likely than the control to save over \$100 and over \$350 across the three rounds (outcomes labeled "Plus 100 \$US Rnd 1,2,3" and "Plus 350 \$US Rnd 1,2,3" respectively), and is larger in level, log, and likelihood than the average of the financial aid group across all outcomes, often signficantly so. Results including all deposits, not just

[10]For practical adminstrative reasons, the window between reminder mailings differed between rounds with the gap between round one and two at 33 days, between round two and three at ten days, and between round three and the end of the dataset at 21 days. Maintaining the maximum window around the three rounds of reminders and two rounds of college cost specific reminders will boost statistical power, but begs the question of how long is reasonable to expect reminder effects to remain. If there is an attention effect it should evaporate by definition, though a cost effect should linger. In general, restricting outcomes to a smaller window of time following the reminder mailings should boost differences in treatments provided there is sufficient statistical power.

those within seven days, are not reported but reveal no significant differences in means across any outcomes though the direction of the differences is similar to those reported in Table 3. These results provide evidence that reminders, in general, and the generic and tuition reminders, in particular, have positive effects on savings behavior.

7.2 Treatment Effects within Language Groups

The actual experiment distributed nine different reminder messages, three for each of the primary language groups of Chinese, Spanish, and English. Table 4 provides a comparison of means by experimental category for Chinese speakers only, revealing patterns consistent with the main effects across language groups. Average level, log, and likelihood of pooled treatments are consistently larger than those of the control (except in the likelihood of saving in round three) and often significantly so. Effects for the tuition group are the strongest and the averages for the tuition group are larger than those of the financial aid group across most outcomes (except for level of round one savings and likelihood of saving the maximum match amount of one hundred dollars exactly), often significantly so. Though not reported, these results for the Chinese group are consistent when all deposits, not just those within seven days of the reminder mailings, are included. Also not reported here, results for Spanish speakers reveal no consistent patterns and results for the English speakers, though consistent in direction with the pooled treatment, generic, and tuition reminder effect for deposits within the seven day window, are rarely significant.

In general, results for the Chinese group seem to drive the main effects of reminders across the language groups, which are generally consistent with an attention effect and a cost effect. Null effects for Spanish and English groups cannot be rejected, though that result may reflect weak statistical power at least in the case of the Spanish group. To the extent stratification by language reduces statistical power, this more pronounced effect within the Chinese group may reflect higher rates of participation. Among the 1,035 Chinese accountholders, 150 (14.5%) contribute to K2C at least once. By comparison only 63 (4.2%) of 1,499 Spanish-preferred accountholders and 225 (6.3%) of 3,626 English-preferred language accountholders ever save with K2C. Alternatively, these results could reflect the particular attention enhancement effect among native speakers of receiving mail in Chinese language, which is likely less common than the more pervasive use of English and Spanish in San Francisco, though one would expect this effect across treatment groups. Given that the messages were translations, treatments could also have different interpretations in different languages. It is also worth noting that the English group is a catch-all for anyone that did not indicate a preference for Chinese or Spanish when registering with SFUSD.[11]

[11] The actual experiment distributed nine different reminder messages, three for each of the primary language groups of Chinese, Spanish, and English. A professional translation service was hired to translate the messages, which were originally composed in English, and the author verified the interpretation of the translations with native Chinese and Spanish speakers. Nevertheless, to the extent messages are interpreted differently in different languages, each language-message reminder may be considered its own treatment. Comparing results within language category can control for this variation across languages. However, language may also be interpretted as a proxy for ethnicity, socio-economic status, etc. with many bilingual schools in SFUSD actually organized around language communities. Therefore, effects within language groups should be interpretted as local effects that may not apply to the larger population. One additional

7.3 Testing Predictions at the Margin & Among New Savers

Having provided some suggestive evidence of a general reminder effect at least among the Chinese group, the analysis will turn to testing the specific predictions of the model. Prediction 1a anticipates an attention effect for the generic and tuition reminders at the intensive margin among prior savers from Wave 1 and Wave 2, but no effect at the extensive margin among non-savers. First examining the extensive margin, there are only ten accountholders from Wave 1 or Wave 2 that first began to save after the experiment commenced. By experimental category, two were from the generic group, five from tuition, one from financial aid, and two from the control. Differences between these groups are not close to significant, revealing that, as anticipated, the treatments had no effect on the extensive margin. The analysis can reject that any additional "attention" reminders bring to saving with K2C for non-saving families is effective at encouraging them to save. This result implies that any potential attention effect among Wave 1 and Wave 2 savers should be considered an effect of *attention on the intentional* that would not generally apply to a non-selected population of savers. Of course, this result may be local to this experiment or the K2C program where there are potential reasons other than inattention or perceived future college cost that may prevent families from contributing to K2C accounts. Candidate explanations include unattractive K2C incentives, distrust of the program, lack of discretionary funds to save, low expectations for attending college, low expectations for incurring significant college cost due to scholarships, etc.. It is also conceivable that many of those accountholders that do not save with K2C do have alternative college savings accounts and K2C reminders remind them to save in those other accounts. Nevertheless, given that there are +2400 accountholders from Wave 1 and Wave 2 that do not save prior to the experiment, it is surprising that there could be an effect at the intensive margin and none whatsoever at the extensive margin.

Results at the intensive margin for deposits within seven days of reminder mailings further confirm Prediction 1a's expectation of an attention effect. Regression results are reported according to the following OLS specification:

$$Y = \beta_i D_i + \gamma_1 \text{Chin.} + \gamma_2 \text{Span.} + \gamma_3 \text{Pre. Exper. Sav. Level} + \gamma_j X + \epsilon \tag{7}$$

where D_i are the three treatment dummies and the β_i are the coefficients of interest. The vector X captures a variety of controls, including basic controls (gender, FRL consent, phone, email), preferred language (not just Spanish, Chinese, or English), school, cohort wave, and phase school was absorbed into K2C. Results for log deposits and likelihood of depositing within seven days of reminder mailings for prior savers from Wave 1 and 2 are reported in Table 5.[12] The generic treatment is at least marginally significant for all but the fully saturated Log Both Experiments and P(Both) specifications in columns (3) and (6), which reveal similar point estimates. Results for outcomes in levels are not significant

but important nuance is the fact that there may be significant variation within language communities with the Chinese language group including Mandarin, Cantonese, Taiwanese, etc. speakers and the English group as a catch-all for anyone that was not included in the Chinese or Spanish groups. Reflecting the diversity of SFUSD, in fact, 49.7% of accountholders in the English group are registered in SFUSD as preferring a non-English primary language.

[12]Note that the coefficient on Chinese is often negative in these regressions because specific language group controls are included in every specification to increase precisions.

and are omitted from the table. For the preferred specification with the full P(Cost) model in column (12), which is not contaminated with the different treatment in round three of the experiment, the point estimate on the generic treatment implies that, among Wave 1 and Wave 2 prior savers, those that received the generic reminder were 9.7% more likely to save within seven days of round one or round two. Given that 15% of Wave 1 and Wave 2 accountholders save prior to the experiment, the attention effect attributed to the generic reminder should encourage approximately 1.5% more instances of deposits, though it may be the case that this savings increase is offset by a decrease in the future. The point estimates for the tuition treatment is slightly (though insignificantly) larger than coefficients on the generic treatment across all specifications except Log(Cost), which is consistent with a cost effect accumulating in addition to the attention effect. In a regression specification including all deposits, not just those within seven days, in Table 6, the null hypothesis of no attention effect cannot be rejected though the coefficients on the treatment variables are always positive. The smaller coefficients on the generic treatment when using all the deposit data suggests that the point estimate of the treatment effect, though significant in reduced form within one week, should be interpreted with caution as it may overstate any potential effect.

Prediction 1b suggests that among Wave 3 savers, cumulative savings in the experiment periods should be greater for those assigned to the generic and tuition groups than the control, consistent with an attention effect. The results here are less straightforward. Table 7 reveals that savings within seven days is significantly larger for generic than the control at 90% confidence in round two and Cost Experiment, but smaller, though not significantly so, in round three. Savings for the tuition group is significantly larger than the control in Both Experiments in levels and logs. The generic and tuition groups have a higher probability of making large deposits above \$100 and \$350. Together, these results neither clearly reject nor support Prediction 1b. This ambiguous result for new savers when combined with the analysis confirming Prediction 1a for prior savers is suggestive of, or at least does not clearly reject, an attention effect at least on the intentional savers.

The hypothesized cost effect due to the persuasive content of the reminder messages leads to Prediction 2 that, among prior savers in Wave 1 and Wave 2, and new Wave 3 enrollees, the magnitude of savings should be ranked according to: Tuition > Generic > Financial Aid. Coefficients on the treatments in regressions in Table 5 for prior savers in Wave 1 and Wave 2 conform to the predicted pattern for deposits within seven days. Comparing tuition to financial aid effects, F-tests reject the likelihood that coefficients on tuition and financial aid treatments are equal at 90% confidence in every specification with likelihood of saving across round one and round two as the outcome of interest (columns 10-12). However, when applying similar regressions to the full deposit data in Table 6 there is no stable ranking of coefficients by treatment. To the extent accountholder households update their expectation of college cost with information about tuition or financial aid from the savings reminders and that information should not be forgotten after seven days, these results seem inconsistent with a cost effect.

Data for Wave 3 is more consistent between the limited seven day deposit sample and the full sample, though is still not entirely convincing. Table 8 Panel A compares means for Wave 3, using all deposits, and reveals ambiguous results. The mean magnitude of savings of the tuition group is always greater than that of the financial aid group, though this difference is rarely significant. Tuition also has significantly more large deposits above \$100. The mean for tuition group savings is greater than that of the generic group when consolidating savings levels across all rounds in Both Exper Sav., but not when omitting round three in Cost Exper Sav.. Table 8 Panel B reports the outcomes within a seven day window. The results here are more consistent with the prediction. The mean for the tuition group is again always larger than the financial aid group, but only larger than the generic group when consolidating across all rounds.

In general, the cost effect seems somewhat consistent with data for new savers, but not for prior savers. Though not anticipated by the model, this may reflect that prior savers are already well informed about college cost or that cost information gets crowded out among people for whom savings through K2C is already salient. To the extent there is a cost effect for new savers, the evidence from Wave 1 and Wave 2 seems to suggest this will diminish over time.

7.4 Summary of Results

Across a variety of outcomes, the experimental results provide suggestive evidence that reminders generally have a positive effect on savings. There is significant heterogeneity with clear effects for tuition reminders among Chinese language preferred accountholders, no effects of any kind among Spanish speakers, and ambiguous results for English preferred recipients. Using differences in deposit histories among accountholders by wave at the intensive and extensive margins, and between prior and new program enrollees provides an opportunity to test predictions consistent with the mechanisms described in the above model with attention effects and cost effects. Throughout the analysis, the experiment seems to suffer from insufficient power with trends in the data often conditional on the specific outcome of interest. This lack of statistical power may reflect how few accountholders actively make deposits in K2C accounts and how little of the variation in deposit amounts is likely explained by reminder effects.[13] Nevertheless, the data do appear to confirm the presence of an *attention effect among the intentional*. For prior savers in Wave 1 and Wave 2, the estimated average attention effect of generic reminders is non-trivial, increasing the probability of saving within one week among the 15% of accountholders that were prior savers, by approximately 10%. There is thinner evidence for the cost effect. Consistent with a cost effect story, financial aid reminders have lower mean outcomes compared to the generic group and the tuition group for deposits made within seven days of the reminder mailings. However, when including the full sample of deposits, this pattern disappears for Wave 1 and Wave 2 and becomes less clear for Wave 3.

[13] Even in prior studies where there are effects of reminders, the effects are often small. In Karlan et al. (2011) when pooling interventions the effect was a 6% increase in savings.

8 Conclusion

Extending prior work on the effect of savings reminders, the findings from this study reinforce prior evidence that inattention plays a role in undersaving. At least in the setting of this study with earmarked college savings accounts and select savings incentives, the data suggest the effect of reminders on overcoming inattention - the *attention effect* - is only relevant for accountholders that intend to save anyway for whom the program is salient. Among such accountholders, the attention effect is not trivial, increasing the probability of savings within a week of being reminded by 9.7%.

The distinction between those accountholders for whom the program is and is not salient is particularly relevant for policymakers interested in mobilizing savings, suggesting that keeping people attentive to savings opportunities may be ineffective for the majority of the population. The lack of findings on the extensive margin suggests it may be prudent for K2C staff to target reminders to existing savers only after a fixed period following enrollment. Additionally, emphasizing the availability of financial aid does not seem like a pragmatic strategy for mobilizing more savings among new enrollees. Not surprisingly, more costly material incentives to save may be necessary to mobilize savings. In the context of mobilizing college-specific savings, further research into the appropriate incentives is warranted given low participation in national college savings programs such as tax-preferred 529 college savings accounts (GAO, 2012) and local pilot efforts such as SEED Oklahoma (Beverly et al., 2012). Though not addressed in this study, interventions to change individuals' preferences for saving, as opposed to their attention to saving, may be a promising direction for research aimed at informing efforts to mobilize saving. Work on lottery-based savings offers a pertinent example (Kearney et al., 2010).

Interventions to mobilize savings for expenditures, such as college, inevitably interact with policies to encourage college-going, such as financial aid subsidies. The extent to which expectations of receiving financial aid affect household college savings choices should be an important input into financial aid related policy formulation. Understanding whether households increase savings in the context of rising tuition is also of direct relevance to policymaking. This study sought to directly vary household expectations of college cost net of tuition expense and financial aid subsidies using differing messaging on postcard savings reminders. The findings for a *cost effect* from changing college cost expectations are ambiguous with weak suggestive evidence that a cost effect may only be relevant for new enrollees, but not experienced savers. This provides weak evidence that providing information on financial aid to new savers may, in fact, negatively impact savings. Even this weak evidence suggests further investigation is warranted given the pervasive use of finanical aid to promote college access.

References

Ashraf, N., Karlan, D., and Yin, W. (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *The Quarterly Journal of Economics*, 121(2):635–672.

Bailey, M. J. and Dynarski, S. M. (2011). Gains and gaps: Changing inequality in U.S. college entry and completion. Working Paper 17633, National Bureau of Economic Research.

Banerjee, A. and Mullainathan, S. (2010). The shape of temptation: Implications for the economic lives of the poor. Technical report, National Bureau of Economic Research.

Benartzi, S. and Thaler, R. (2007). Heuristics and biases in retirement savings behavior. *Journal of Economic Perspectives, forthcoming*.

Bettinger, E. P., Long, B. T., Oreopoulos, P., and Sanbonmatsu, L. (2009). The role of simplification and information in college decisions: Results from the H&R block FAFSA experiment. Working Paper 15361, National Bureau of Economic Research.

Beverly, S., Kim, Y., Sherraden, M., Nam, Y., and Clancy, M. (2012). Socioeconomic status and early savings outcomes: Evidence from a statewide child development account experiment. Technical Report No. 12 - 30, Washington University Center for Social Development.

Bhargava, S. and Manoli, D. (2012). Why are benefits left on the table? Assessing the role of information, complexity, and stigma on take-up with an IRS field experiment. *Working Paper*.

CA Dept. of Education (2013). DataQuest. http://dq.cde.ca.gov/dataquest/.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2):315–72.

DellaVigna, S. and Gentzkow, M. (2010). Persuasion: Empirical evidence. *Annual Review of Economics*, 2(1):643–669.

Deming, D. and Dynarski, S. (2009). Into college, out of poverty? Policies to increase the postsecondary attainment of the poor. Working Paper 15387, National Bureau of Economic Research.

Destin, M. and Oyserman, D. (2009). From assets to school outcomes how finances shape children's perceived possibilities and intentions. *Psychological Science*, 20(4):414418.

Dupas, P. and Robinson, J. (2011). Why don't the poor save more? Evidence from health savings experiments. Technical report, National Bureau of Economic Research.

Dynarski, S. M. (2003). Does aid matter? Measuring the effect of student aid on college attendance and completion. *American Economic Review*, 93(1):279–288.

E., Heller, D. (1997). Student price response in higher education: An update to Leslie and Brinkman. *Journal of Higher Education*, 68(6).

Fain, Paul (2012). Pell spending declines despite growth in grant recipients. *Inside Higher Ed*.

Fairbank, Maslin, Maullin, Metz & Associates (2011). Parent perceptions of the san francisco kindergarten to college program: Report of september 2011 focus group research. Technical report.

FDIC (2013). FDIC: 2011 FDIC national survey of unbanked and underbanked households. http://www.fdic.gov/householdsurvey/.

GAO (2012). Higher education: A small percentage of families save in 529 plans. Technical Report GAO-13-64, US General Accountability Office.

Hemelt, S. W. and Marcotte, D. E. (2011). The impact of tuition increases on enrollment at public colleges and universities. *Educational Evaluation and Policy Analysis*, 33(4):435–457.

Hill, Laura and Hayes, Joseph (2013). Undocumented Immigrants. Technical report, PPIC.

Jackson, Orville (2013). The cost of opportunity: Access to college financial aid in california. Technical report, The Education Trust - West.

Karlan, McConnell, Margaret, Mullainathan, Sendhil, and Zinman, Jonathan (2011). Getting to the top of mind: How reminders increase saving. *Working Paper*.

Kearney, M. S., Tufano, P., Guryan, J., and Hurst, E. (2010). Making savers winners: An overview of prize-linked savings products. Working Paper 16433, National Bureau of Economic Research.

LAO (2009). Higher education affordability: Fees and financial aid. Technical report, CA Legislative Analyst's Office.

Mason, L. R., Nam, Y., Clancy, M., Kim, Y., and Loke, V. (2010). Child development accounts and saving for children's future: Do financial incentives matter? *Children and Youth Services Review*, 32(11):1570–1576.

McKernan, S.-M. and Sherraden, M. W. (2008). *Asset Building and Low-Income Families*. The Urban Insitute.

Sherraden, M. and Sherraden, M. M. W. (1991). *Assets and the Poor: A New American Welfare Policy*. M.E. Sharpe.

Thaler, R. H. (1990). Anomalies: Saving, fungibility, and mental accounts. *The Journal of Economic Perspectives*, 4(1):pp. 193–205.

Table 1: Round Two Randomization Test

	Control	Generic	Tuition	Financial Aid
Pre-Exper. Sav.	14.61	21.41	22.08	15.73
	(-1.34)	(1.47)	(1.64)	(0.34)
Log Pre-Exper.	0.32	0.33	0.33	0.33
	(-0.02)	(0.04)	(0.00)	(0.01)
Saved Pre-Exper.	0.06	0.06	0.06	0.07
	(0.18)	(-0.15)	(-0.43)	(0.14)
Male Student	0.50	0.52	0.53	0.50
	(-1.03)	(1.04)	(1.55)	(-0.07)
Consent FRL Disclosure	0.42	0.44	0.44	0.44
	(-1.19)	(1.09)	(0.99)	(0.83)
Phone Info Provided	0.84	0.82	0.83	0.82
	(1.00)	(-1.20)	(-0.13)	(-1.13)
Email Info Provided	0.59	0.59	0.59	0.59
	(0.07)	(-0.21)	(0.13)	(-0.08)
Spanish	0.23	0.22	0.23	0.24
	(0.20)	(-0.80)	(-0.09)	(0.40)
Chinese	0.18	0.19	0.19	0.19
	(-0.40)	(0.33)	(0.26)	(0.40)
Regular Saver	0.02	0.02	0.02	0.02
	(-0.96)	(0.83)	(0.61)	(0.95)
Observations	6856	3429	3427	3426

In columns (2)-(4) t-tests reported in parenthesis compare the respective treatment groups to the control group. The observation count for the control group in column (1) includes all observations in the round two sample, reflecting that the t-test for column (1) compares the pooled treatment groups vs. control. "Pre. Exper. Sav." and "Log Pre. Exper." capture pre-experimental savings in levels and logs respectively. The remaining variables are all dummies indicating whether accountholders saved before the experiment, are male, consented to disclosure of their FRL status, provided a mobile phone number when registering with SFUSD, provided email contact when registering with SFUSD, and prefer to receive materials in Spanish or Chinese. The final variable, Regular Saver, is a proxy for whether or not the accountholder is a frequent saver or potentially has enrolled in direct deposit. The dummy is turned on if an accountholder has saved three consecutive times within 45 days of each other, including at least once within 45 days prior to the first round of the experiment.

Table 2: Summary Statistics

Panel A: All Accounts by Wave

	Wave1	Wave2	Wave3	Wave3a	Wave3b	All Waves
English	0.541	0.533	0.616	0.987	0.081	0.583
Spanish	0.260	0.279	0.203	0.000	0.495	0.232
Chinese	0.199	0.188	0.181	0.013	0.424	0.185
Male Student	0.491	0.517	0.520	0.521	0.518	0.515
Registered Online	0.079	0.093	0.044	0.050	0.035	0.062
Consent FRL Disclosure	0.670	0.530	0.339	0.346	0.328	0.436
Phone Info Provided	0.805	0.797	0.850	0.856	0.841	0.828
Email Info Provided	0.198	0.618	0.665	0.795	0.477	0.591
Ever Saves	0.151	0.153	0.066	0.063	0.070	0.102
Regular Saver	0.032	0.056	0.002	0.004	0.000	0.022
Any Round 1,2,3 Savings	0.043	0.060	0.065	0.063	0.068	0.061
Round 1, 2, 3 Sav. $US	4.966	4.633	7.243	6.414	8.436	6.196
Observations	894	1947	4007	2365	1642	6856

Panel B: Those that Ever Save by Wave

	Wave1	Wave2	Wave3	Wave3a	Wave3b	All Waves
English	0.452	0.515	0.589	0.993	0.061	0.532
Spanish	0.178	0.138	0.094	0.000	0.217	0.129
Chinese	0.370	0.347	0.317	0.007	0.722	0.339
Male Student	0.504	0.502	0.517	0.547	0.478	0.507
Registered Online	0.526	0.613	0.660	0.787	0.496	0.615
Consent FRL Disclosure	0.926	0.889	0.868	0.913	0.809	0.888
Phone Info Provided	0.852	0.822	0.872	0.860	0.887	0.848
Email Info Provided	0.385	0.785	0.857	0.967	0.713	0.736
Ever Saves	1.000	1.000	1.000	1.000	1.000	1.000
Regular Saver	0.215	0.370	0.034	0.060	0.000	0.213
Any Round 1,2,3 Savings	0.281	0.394	0.985	0.993	0.974	0.598
Round 1, 2, 3 Sav. $US	32.889	30.372	109.516	101.132	120.452	61.038
Observations	135	297	265	150	115	696

Panel C: Savers during the Experiment by Wave

	Wave1	Wave2	Wave3	Wave3a	Wave3b	All Waves
English	0.421	0.590	0.590	0.993	0.054	0.575
Spanish	0.105	0.085	0.096	0.000	0.223	0.094
Chinese	0.474	0.325	0.314	0.007	0.723	0.332
Male Student	0.526	0.530	0.510	0.544	0.464	0.517
Registered Online	0.658	0.744	0.667	0.792	0.500	0.688
Consent FRL Disclosure	0.974	0.940	0.866	0.913	0.804	0.897
Phone Info Provided	0.921	0.795	0.874	0.859	0.893	0.856
Email Info Provided	0.237	0.838	0.858	0.966	0.714	0.796
Regular Saver	0.711	0.769	0.034	0.060	0.000	0.303
Any Round 1,2,3 Savings	1.000	1.000	1.000	1.000	1.000	1.000
Round 1, 2, 3 Sav. $US	116.842	77.098	111.195	101.810	123.679	102.121
Observations	38	117	261	149	112	416

Panel A includes all account holders. Panel B excludes accounts that never receive a private contribution. Panel C excludes accounts that do not receive a deposit after the start of the experiment. Waves indicate the years they enrolled in kindergarten, which correspond to waves in the rollout of K2C. Wave 3 is divided between Wave 3a, mostly accountholders that prefer to receive English language materials, and Wave 3b, mostly accountholders that prefer non-english language materials. Wave 3b was not yet enrolled at the time of experimental Round 1, but was included in Round 2 and Round 3.

Table 3: Main Results: Averages by Treatment within Seven Days

	(1) Control	(2) Financial Aid	(3) Generic	(4) Tuition	(5) Tuition v. Fin. Aid
Round 1 Sav. $US	0.36	0.56	0.77	0.76	0.76
	(-1.08)	(0.64)	(1.34)	(1.10)	(0.43)
Round 2 Sav. $US	0.50	0.43	0.96	0.83	0.83^{+}
	(-0.90)	(-0.29)	(1.16)	(1.15)	(1.70)
Round 3 Sav. $US	0.51	0.95	0.72	1.57^{+}	1.57
	(-1.34)	(1.21)	(0.57)	(1.83)	(0.96)
Cost Exper Sav. $US	0.86	1.00	1.73^{+}	1.58	1.58
	(-1.40)	(0.33)	(1.73)	(1.57)	(1.16)
Both Exper Sav. $US	1.38^{+}	1.95	2.45	3.15^{*}	3.15
	(-1.82)	(1.05)	(1.49)	(2.41)	(1.47)
Log Round 1 Sav.	0.03	0.02	0.05	0.03	0.03
	(-0.78)	(-0.38)	(1.55)	(0.66)	(1.02)
Log Round 2 Sav.	0.04^{+}	0.04	0.06^{*}	0.07^{*}	0.07^{*}
	(-1.95)	(0.41)	(2.04)	(2.42)	(2.05)
Log Round 3 Sav.	0.04	0.05	0.04	0.06	0.06
	(-0.87)	(0.79)	(0.11)	(1.24)	(0.47)
Log Cost Exper Sav.	0.06^{*}	0.06	0.10^{*}	0.10^{*}	0.10^{*}
	(-2.05)	(0.21)	(2.43)	(2.42)	(2.22)
Log Both Exper Sav.	0.09^{*}	0.11	0.13^{+}	0.15^{**}	0.15^{*}
	(-2.02)	(0.66)	(1.78)	(2.64)	(1.98)
Any Saving Round 1	0.01	0.01	0.01^{+}	0.01	0.01
	(-0.73)	(-0.63)	(1.67)	(0.56)	(1.18)
Any Saving Round 2	0.01^{*}	0.01	0.02^{*}	0.02^{**}	0.02^{*}
	(-2.15)	(0.64)	(2.11)	(2.69)	(2.08)
Any Saving Round 3	0.01	0.01	0.01	0.02	0.02
	(-0.60)	(0.59)	(-0.01)	(0.87)	(0.28)
Any Saving Cost Exper	0.02^{*}	0.02	0.03^{*}	0.03^{**}	0.03^{*}
	(-2.25)	(0.39)	(2.51)	(2.71)	(2.33)
Any Saving Both Exper	0.03^{*}	0.03	0.04^{+}	0.04^{**}	0.04^{*}
	(-2.02)	(0.63)	(1.77)	(2.64)	(2.03)
100 $US Round 1,2,3	0.01	0.00	0.01	0.01	0.01
	(-0.51)	(-0.47)	(0.42)	(1.18)	(1.64)
Plus 100 $US Rnd 1,2,3	0.00^{+}	0.00^{+}	0.00	0.01^{*}	0.01
	(-1.72)	(1.67)	(0.81)	(2.11)	(0.50)
Plus 350 $US Rnd 1,2,3	0.00	0.00	0.00	0.00^{+}	0.00
	(-1.53)	(1.41)	(1.41)	(1.73)	(0.45)
Observations	6856	3426	3429	3427	3427

T-statistics in parenthesis ** $p<0.01$, * $p<0.05$, $^{+}$ $p<0.1$. In columns (2)-(4) t-statistics reported in parenthesis compare the respective treatment groups to the control group. The observation count for the control group in column (1) includes all observations in the round two sample, reflecting that the t-test for column (1) compares the pooled treatment groups vs. control. Column (5) compares tuition to finanical aid group reporting the mean for the tuition group and the t-stat for the t-test of means. Average savings in level, log, and likelihood are reported by rounds one, two, and three separately, for the "Cost Experiment," which consolidates rounds one and two, and for "Both Experiments," which consolidates all three rounds. Outcomes labeled "100 $US Round 1,2,3," "Plus 100 $US Rnd 1,2,3," and "Plus 350 $US Rnd 1,2,3" refer to the probability of saving exactly $100, saving over $100, and saving over $350 across the three rounds respectively.

Table 4: Chinese Group - Seven Day Window

	(1) CH-Control	(2) CH-Fin. Aid	(3) CH-Generic	(4) CH-Tuition	(5) CH-Tuition v Fin Aid
Round 1 Sav. $US	0.40	0.86	1.49	0.43	0.43
	(-0.54)	(0.60)	(0.79)	(0.08)	(-0.60)
Round 2 Sav. $US	0.17*	1.04+	1.05+	2.17*	2.17
	(-2.06)	(1.71)	(1.89)	(2.66)	(1.27)
Round 3 Sav. $US	1.28	2.49	2.34	4.83	4.83
	(-0.86)	(0.74)	(0.57)	(1.22)	(0.75)
Cost Exper Sav. $US	0.57	1.90	2.53	2.61*	2.61
	(-1.57)	(1.44)	(1.36)	(2.38)	(0.62)
Both Exper Sav. $US	1.85	4.39	4.87	7.44+	7.44
	(-1.32)	(1.35)	(0.96)	(1.85)	(0.92)
Log Round 1 Sav.	0.03	0.04	0.05	0.06	0.06
	(-0.77)	(0.47)	(0.70)	(1.10)	(0.59)
Log Round 2 Sav.	0.03*	0.09+	0.10*	0.19*	0.19+
	(-2.45)	(1.77)	(1.99)	(3.27)	(1.69)
Log Round 3 Sav.	0.09	0.10	0.08	0.11	0.11
	(-0.19)	(0.27)	(-0.04)	(0.38)	(0.11)
Log Cost Exper Sav.	0.05*	0.12	0.15+	0.24*	0.24+
	(-2.34)	(1.50)	(1.88)	(3.27)	(1.80)
Log Both Exper Sav.	0.13+	0.20	0.20	0.32*	0.32
	(-1.76)	(1.18)	(1.11)	(2.68)	(1.50)
Any Saving Round 1	0.01	0.01	0.01	0.02	0.02
	(-0.92)	(0.43)	(0.78)	(1.41)	(1.03)
Any Saving Round 2	0.01*	0.03+	0.03+	0.06*	0.06+
	(-2.47)	(1.72)	(1.91)	(3.34)	(1.81)
Any Saving Round 3	0.03	0.03	0.02	0.03	0.03
	(0.22)	(-0.04)	(-0.32)	(-0.01)	(0.03)
Any Saving Cost Exper	0.02*	0.04	0.04+	0.08*	0.08*
	(-2.45)	(1.48)	(1.85)	(3.50)	(2.16)
Any Saving Both Exper	0.04+	0.05	0.06	0.09*	0.09+
	(-1.73)	(0.93)	(1.08)	(2.81)	(1.93)
100 $US Round 1,2,3	0.01	0.02	0.01	0.01	0.01
	(-0.41)	(1.11)	(-0.03)	(-0.01)	(-1.12)
Plus 100 $US Rnd 1,2,3	0.00	0.01	0.00	0.02	0.02
	(-0.87)	(0.56)	(-0.02)	(1.63)	(1.16)
Plus 350 $US Rnd 1,2,3	0.00	0.00	0.00	0.01	0.01
	(-1.11)	(0.99)	(0.99)	(1.41)	(0.59)
Observations	1271	600	603	596	602

T-statistics in parenthesis ** $p<0.01$, * $p<0.05$, $^{+}$ $p<0.1$. In columns (2)-(4) t-statistics reported in parenthesis compare the respective treatment groups to the control group. The observation count for the control group in column (1) includes all observations in the round two sample, reflecting that the t-test for column (1) compares the pooled treatment groups vs. control. Column (5) compares tuition to finanical aid group reporting the mean for the tuition group and the t-stat for the t-test of means. Average savings in level, log, and likelihood are reported by rounds one, two, and three separately, for the "Cost Experiment," which consolidates rounds one and two, and for "Both Experiments," which consolidates all three rounds. Outcomes labeled "100 $US Round 1,2,3," "Plus 100 $US Rnd 1,2,3," and "Plus 350 $US Rnd 1,2,3" refer to the probability of saving exactly $100, saving over $100, and saving over $350 across the three rounds respectively.

Table 5: Wave 1 & 2 Pre-Savers Only - The Intensive Margin - Seven Days

VARIABLES	(1) Log Both	(2) Log Both	(3) Log Both	(4) P(Both)	(5) P(Both)	(6) P(Both)	(7) Log Cost	(8) Log Cost	(9) Log Cost	(10) P(Cost)	(11) P(Cost)	(12) P(Cost)
Generic	0.40*	0.36*	0.23	0.12*	0.11*	0.08	0.41**	0.39**	0.32*	0.12**	0.12**	0.10*
	(0.21)	(0.21)	(0.21)	(0.06)	(0.06)	(0.06)	(0.18)	(0.18)	(0.18)	(0.05)	(0.05)	(0.06)
Tuition	0.41*	0.36*	0.26	0.12**	0.12*	0.09	0.40**	0.37**	0.31*	0.14**	0.13**	0.11**
	(0.21)	(0.21)	(0.21)	(0.06)	(0.06)	(0.06)	(0.18)	(0.18)	(0.18)	(0.05)	(0.05)	(0.05)
Fin Aid	0.16	0.17	0.07	0.05	0.05	0.03	0.11	0.11	0.06	0.03	0.04	0.02
	(0.21)	(0.21)	(0.21)	(0.06)	(0.06)	(0.06)	(0.18)	(0.17)	(0.18)	(0.05)	(0.05)	(0.05)
Pre-Exper Sav		0.00***	0.00***		0.00	0.00		0.00*	0.00*		0.00	0.00
		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)
Chinese	-0.71	-0.79	-1.89	-0.40	-0.41	-0.71	-2.58*	-2.63*	-3.29**	-0.76*	-0.76*	-0.97**
	(1.76)	(1.74)	(1.75)	(0.50)	(0.50)	(0.50)	(1.47)	(1.47)	(1.49)	(0.45)	(0.45)	(0.46)
Spanish	-0.02	-0.12	-1.16	-0.22	-0.24	-0.52	-2.12	-2.18	-2.78**	-0.64	-0.65	-0.84**
	(1.64)	(1.62)	(1.64)	(0.46)	(0.46)	(0.47)	(1.37)	(1.37)	(1.39)	(0.41)	(0.41)	(0.43)
Male Student			0.16			0.06			0.14			0.06
			(0.15)			(0.04)			(0.13)			(0.04)
FRL Disclose			0.78***			0.23***			0.49**			0.15**
			(0.28)			(0.08)			(0.24)			(0.07)
Observations	421	421	421	421	421	421	421	421	421	421	421	421
Adj R-sq	0.05	0.07	0.09	0.05	0.05	0.07	0.04	0.04	0.05	0.02	0.02	0.03
Lang. Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Basic Controls	N	N	Y	N	N	Y	N	N	Y	N	N	Y
Sch. Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Waves	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
School Phase	N	N	Y	N	N	Y	N	N	Y	N	N	Y
Jointly = zero	0.15	0.26	0.56	0.12	0.17	0.42	0.04	0.07	0.16	0.03	0.04	0.10
Tuition = Fin	0.23	0.36	0.37	0.21	0.27	0.30	0.10	0.14	0.15	0.05	0.07	0.09
Tuition = Gen	0.94	0.99	0.89	0.92	0.94	0.85	0.93	0.90	0.95	0.80	0.81	0.76
Fin = Gen	0.26	0.37	0.45	0.25	0.30	0.40	0.08	0.11	0.14	0.10	0.12	0.16

Columns (1)-(6) report OLS regressions for log savings and likelihood of saving within seven days of postcard mailings across all three rounds of the experiment. Columns (7)-(12) refer to only the first two rounds. Across all specifications, controls for language, school, and program wave are included. Basic Controls include dummies for whether accountholders are male, consented to disclosure of their FRL status, provided a mobile phone number when registering with SFUSD, and provided email contact when registering with SFUSD. School controls include a dummy for each school. Waves controls include dummies for each wave of program participation. School Phase controls refer to a dummies for whether the school of the accountholder was included in K2C in the first, second, or third year of the program roll-out. The last four rows report p-values for F-tests of joint equality. Non-linear regression estimation yields similar results to OLS and are, therefore, omitted. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Wave 1 & 2 Pre-Savers Only - The Intensive Margin

VARIABLES	(1) Log Both	(2) Log Both	(3) Log Both	(4) P(Both)	(5) P(Both)	(6) P(Both)	(7) Log Cost	(8) Log Cost	(9) Log Cost	(10) P(Cost)	(11) P(Cost)	(12) P(Cost)
Generic	0.36	0.27	0.13	0.09	0.07	0.04	0.26	0.17	0.06	0.07	0.06	0.02
	(0.27)	(0.26)	(0.27)	(0.07)	(0.07)	(0.07)	(0.25)	(0.24)	(0.25)	(0.07)	(0.07)	(0.067)
Tuition	0.38	0.26	0.15	0.11*	0.10	0.07	0.28	0.17	0.07	0.09	0.07	0.05
	(0.27)	(0.26)	(0.26)	(0.07)	(0.07)	(0.07)	(0.25)	(0.24)	(0.24)	(0.07)	(0.07)	(0.07)
Fin Aid	0.36	0.38	0.27	0.11	0.11*	0.08	0.27	0.29	0.21	0.10	0.10	0.08
	(0.27)	(0.26)	(0.26)	(0.07)	(0.07)	(0.07)	(0.25)	(0.24)	(0.24)	(0.07)	(0.07)	(0.07)
Pre-Exper Sav		0.00***	0.00***		0.00***	0.00***		0.00***	0.00***		0.00***	0.00***
		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)
Chinese	0.22	0.03	-1.22	-0.20	-0.23	-0.53	-0.75	-0.92	-1.97	-0.39	-0.42	-0.70
	(2.28)	(2.18)	(2.21)	(0.56)	(0.55)	(0.55)	(2.10)	(2.02)	(2.05)	(0.56)	(0.55)	(0.56)
Spanish	-0.55	-0.80	-1.93	-0.33	-0.37	-0.65	-0.84	-1.06	-2.03	-0.34	-0.38	-0.65
	(2.12)	(2.04)	(2.06)	(0.52)	(0.51)	(0.52)	(1.96)	(1.89)	(1.91)	(0.52)	(0.51)	(0.52)
Male Student			0.20			0.05			0.25			0.06
			(0.19)			(0.05)			(0.18)			(0.05)
FRL Disclose			0.91***			0.24***			0.77**			0.23***
			(0.35)			(0.09)			(0.32)			(0.09)
Observations	421	421	421	421	421	421	421	421	421	421	421	421
Adj R-sq	0.04	0.11	0.13	0.04	0.07	0.09	0.02	0.10	0.11	0.03	0.06	0.07
Lang. Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Basic Controls	N	N	Y	N	N	Y	N	N	Y	N	N	Y
Sch. Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Waves	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
School Phase	N	N	Y	N	N	Y	N	N	Y	N	N	Y
Jointly = zero	0.43	0.52	0.78	0.30	0.35	0.61	0.63	0.70	0.86	0.43	0.47	0.68
Tuition = Fin	0.93	0.64	0.63	0.91	0.81	0.80	0.97	0.62	0.58	0.94	0.680	0.65
Tuition = Gen	0.94	0.97	0.95	0.72	0.77	0.69	0.93	0.99	0.95	0.74	0.79	0.71
Fin = Gen	0.99	0.66	0.59	0.81	0.60	0.52	0.96	0.63	0.54	0.69	0.50	0.41

Columns (1)-(6) report OLS regressions for log savings and likelihood of saving across all three rounds of the experiment. Columns (7)-(12) refer to only the first two rounds. Across all specifications, controls for language, school, and program wave are included. Basic Controls are reported and include dummies for whether accountholders are male, consented to disclosure of their FRL status, provided a mobile phone number when registering with SFUSD, and provided email contact when registering with SFUSD. School controls include a dummy for each school. Waves controls include dummies for each wave of program participation. School Phase controls refer to a dummies for whether the school of the accountholder was included in K2C in the first, second, or third year of the program roll-out. The last four rows report p-values for F-tests of joint equality. Non-linear regression estimation yields similar results to OLS and are, therefore, omitted. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 7: Comparison of Means: Wave 3 New Savers Only - Seven Days

	(1) Control	(2) Financial Aid	(3) Generic	(4) Tuition	(5) Tuition v. Fin. Aid
Round 1 Sav. $US	6.81	1.61	9.65	16.74	16.74^{+}
	(-0.49)	(-1.54)	(0.61)	(1.14)	(1.71)
Round 2 Sav. $US	5.14	10.08	20.46^{+}	11.52	11.52
	(-1.60)	(1.27)	(1.84)	(1.54)	(0.29)
Round 3 Sav. $US	8.65	17.90	6.58	22.83	22.83
	(-0.95)	(1.11)	(-0.43)	(1.36)	(0.39)
Cost Exper Sav. $US	11.96	11.69	30.12^{+}	28.26^{+}	28.26^{+}
	(-1.52)	(-0.05)	(1.96)	(1.75)	(1.70)
Both Exper Sav. $US	20.61^{+}	29.60	36.69	51.09*	51.09
	(-1.84)	(0.97)	(1.61)	(2.31)	(1.43)
Log Round 1 Sav.	0.43	0.07*	0.50	0.49	0.49*
	(0.40)	(-2.06)	(0.29)	(0.26)	(2.07)
Log Round 2 Sav.	0.50^{+}	0.81	1.06*	0.85	0.85
	(-1.92)	(1.29)	(2.16)	(1.49)	(0.17)
Log Round 3 Sav.	0.51	0.79	0.41	0.80	0.80
	(-0.72)	(1.06)	(-0.49)	(1.09)	(0.03)
Log Cost Exper Sav.	0.87	0.88	1.48^{+}	1.34	1.34
	(-1.47)	(0.03)	(1.94)	(1.56)	(1.49)
Log Both Exper Sav.	1.35^{+}	1.56	1.86	2.12*	2.12
	(-1.80)	(0.63)	(1.51)	(2.26)	(1.55)
Any Saving Round 1	0.12	0.02*	0.12	0.10	0.10*
	(0.85)	(-2.28)	(0.13)	(-0.27)	(2.05)
Any Saving Round 2	0.16^{+}	0.24	0.29^{+}	0.26	0.26
	(-1.78)	(1.18)	(1.85)	(1.46)	(0.25)
Any Saving Round 3	0.14	0.21	0.11	0.20	0.20
	(-0.55)	(0.97)	(-0.64)	(0.89)	(-0.10)
Any Saving Cost Exper	0.25	0.26	0.38^{+}	0.36	0.36
	(-1.39)	(0.15)	(1.73)	(1.48)	(1.28)
Any Saving Both Exper	0.38	0.42	0.48	0.55*	0.55
	(-1.55)	(0.49)	(1.17)	(2.07)	(1.50)
100 $US Round 1,2,3	0.13	0.10	0.18	0.17	0.17
	(-0.45)	(-0.60)	(0.86)	(0.71)	(1.28)
Plus 100 $US Rnd 1,2,3	0.01^{+}	0.06	0.05	0.10*	0.10
	(-1.76)	(1.49)	(1.07)	(2.21)	(0.76)
Plus 350 $US Rnd 1,2,3	0.00	0.02	0.02	0.03	0.03
	(-1.19)	(1.06)	(1.03)	(1.42)	(0.49)
Observations	265	131	134	138	131

T-statistics in parenthesis ** $p<0.01$, * $p<0.05$, $^{+}$ $p<0.1$. In columns (2)-(4) t-stats compare the respective treatment groups to the control group. The observation count for the control group in column (1) includes all observations in the round two sample, reflecting that the t-test for column (1) compares the pooled treatment groups vs. control. Column (5) compares tuition to finanical aid group reporting the mean for the tuition group and the t-stat for the t-test of means. Average savings in level, log, and likelihood are reported by rounds one, two, and three separately, for the "Cost Experiment," which consolidates rounds one and two, and for "Both Experiments," which consolidates all three rounds. Outcomes labeled "100 $USD Round 1,2,3," "Plus 100 $US Rnd 1,2,3," and "Plus 350 $US Rnd 1,2,3" refer to the probability of saving exactly $100, saving over $100, and saving over $350 across the three rounds respectively.

Table 8: Comparison of Means - Wave 3

Panel A: All Deposits

	(1) Control	(2) Financial Aid	(3) Generic	(4) Tuition	(5) Tuition v. Fin. Aid
Cost Exper. Sav.	2.63	2.25	4.18	3.38	3.38
	(-0.73)	(-0.57)	(1.43)	(0.70)	(1.09)
Both Exper. Sav.	4.97	4.79	5.96	7.75	7.75^{+}
	(-0.95)	(-0.17)	(0.76)	(1.63)	(1.71)
Log Cost Exper. Sav.	0.16	0.13	0.17	0.13	0.13
	(0.35)	(-0.69)	(0.41)	(-0.64)	(0.04)
Log Both Exper. Sav.	0.25	0.22	0.25	0.28	0.28
	(-0.07)	(-0.54)	(0.05)	(0.64)	(1.17)
100 $US Round 1,2,3	0.03	0.03	0.03	0.03	0.03
	(0.11)	(0.13)	(-0.01)	(-0.41)	(-0.54)
Plus 100 $US Rnd 1,2,3	0.01	0.01	0.01	0.02^{+}	0.02^{+}
	(-1.15)	(0.00)	(0.85)	(1.90)	(1.90)
Plus 350 $US Rnd 1,2,3	0.00	0.00	0.00	0.00^{*}	0.00
	(-1.53)	(1.00)	(1.41)	(2.00)	(1.34)
Observations	4007	2001	2004	2004	2003

Panel B: Deposits within Seven Days of Mailings

	Control	Financial Aid	Generic	Tuition	Tuition v. Fin. Aid
Cost Exper Sav.	0.82	0.72	1.95^{+}	1.94	1.94^{+}
	(-1.31)	(-0.28)	(1.70)	(1.64)	(1.80)
Both Exper Sav.	1.42	1.83	2.38	3.51^{*}	3.51
	(-1.53)	(0.62)	(1.29)	(2.09)	(1.55)
Log Cost Exper Sav.	0.06	0.05	0.10	0.09	0.09
	(-1.05)	(-0.27)	(1.44)	(1.32)	(1.59)
Log Both Exper Sav.	0.09	0.10	0.12	0.15^{+}	0.15
	(-1.14)	(0.14)	(0.96)	(1.71)	(1.57)
100 $US Round 1,2,3	0.01	0.01	0.01	0.01	0.01
	(-0.28)	(-0.78)	(0.65)	(0.65)	(1.41)
Plus 100 $US Rnd 1,2,3	0.00	0.00	0.00	0.01^{*}	0.01
	(-1.64)	(1.34)	(1.00)	(2.12)	(0.90)
Plus 350 $US Rnd 1,2,3	0.00	0.00	0.00	0.00	0.00
	(-1.15)	(1.00)	(1.00)	(1.41)	(0.57)
Observations	4007	2001	2004	2004	2003

T-statistics in parenthesis ** p<0.01, * p<0.05, $^{+}$ p<0.1. In columns (2)-(4) t-stats compare the respective treatment groups to the control group. The observation count for the control group in column (1) includes all observations in the round two sample, reflecting that the t-test for column (1) compares the pooled treatment groups vs. control. Column (5) compares tuition to finanical aid group reporting the mean for the tuition group and the t-stat for the t-test of means.

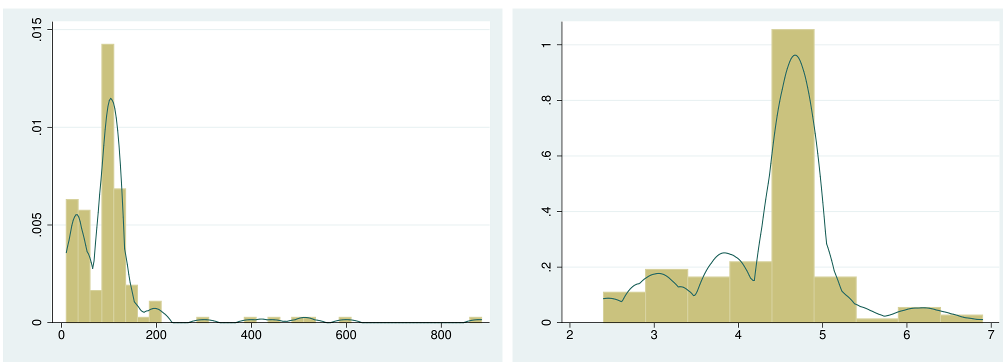
Figure 1: Number of Accountholders by Wave and Rollout Phase

	Wave 1 2010-2011	Wave 2 2011-2012	Wave 3a 2012-2013A	Wave 3b 2012-2013B	*Total*
Phase 1 schools	894	970	476	461	*2801*
Phase 2 schools		977	533	417	*1927*
Phase 3 Schools			1356	764	*2120*
Total	*894*	*1947*	*2365*	*1642*	*6848*

Figure 2: Treatment Groups across Waves and Experiment Rounds

	Round 1 Nov. 7 – Dec. 10	Round 2 Dec. 11 – Dec. 20	Round 3 Dec. 21 – Jan. 11
Wave 1 & 2	Control	Control	Control
	Fin Aid	Fin Aid	New
	Generic	Generic	Treatment
	Tuition	Tuition	Assignment
Wave 3a	Control	Control	Control
	Fin Aid	Fin Aid	New
	Generic	Generic	Treatment
	Tuition	Tuition	Assignment
Wave 3b		Control	Control
		Fin Aid	New
		Generic	Treatment
		Tuition	Assignment

Figure 3: Visualizing Deposit Distributions



(a) Cum. Savings Distrib. in Levels

(b) Cum. Savings Distrib. in Logs

A Appendix: Account Structure and Incentives

This default feature of account opening is experimentally meaningful in differentiating the experimental sample from that of other reminder experiments, such as in Karlan et al. (2011), in which subjects select into the savings program. Families with students in SFUSD are poorer than families in San Francisco generally with greater than 60% of students enrolled in FRL.[14] As a result, the experimental sample is not representative of the general population. Nevertheless, in contrast to Bhargava and Manoli (2012) in which families that had not claimed but were likely eligible for EITC benefits were actively targeted for reminders by experimenters, this study's sample is not biased beyond the natural bias of selection into or out of public school enrollment.[15]

Merely providing a mainstream financial savings instrument for students' families may be meaningful in mobilizing savings for a few reasons. First, K2C accounts do not require personal identification required by commercial retail banking institutions, such as drivers licenses and social security cards. This was designed with intention to include children of undocumented immigrants who lack such identification and whose families may benefit the most from K2C accounts. Undocumented families often lack access to mainstream financial services and make up a significant portion of San Francisco families and SFUSD students; thus, K2C may provide unique access to a mainstream financial savings product for these families.[16] Second, merely labeling an account for a specific savings goal, e.g. college, may encourage savings. Within the empirical literature on developing countries, Dupas and Robinson (2011) finds that merely providing a lock box to rural Kenyan women for savings intended for health expenditures increased savings 66% in comparison to controls, which the authors suggest as evidence for the impact of "labeling" within the mental accounting framework. Finally, with some reasonable simplifying assumptions, the K2C internal rate of return for the optimizing K2C saver is well above market returns for comparable accounts. The optimal saver would save \$10 for five months in a row and then save \$50 in the sixth month to optimally capture the full \$100 match and the Save Steady bonus, earning approximately 9.0%, about five times the rate of a 10 year Treasury bond and +40 times the

[14] Families with students enrolled in FRL must be within 185% of the Federal income poverty guidelines

[15] The high FRL rates in SFUSD did enable K2C planners to argue that they were targeting poorer families when advocating politically for the program initially.

[16] The FDIC (FDIC, 2013) estimates 18% of Californians are "underbanked." The Public Policy Institute of California estimates almost 8% of residents of the San Francisco Bay Area are undocumented (Hill, Laura and Hayes, Joseph, 2013).

interest rate of retail savings accounts.[17]

The likely effect of the earmarked feature of the K2C accounts is theoretically ambiguous. With a randomized experiment in the Philippines, Ashraf et al. (2006) demonstrates how many people have positive demand for illiquidity. Some households or individuals within households in San Francisco may have demand for the commitment device of an earmarked account like K2C to discipline their or their family members' near term consumption. Alternatively, families that may be willing to deposit money intended as college savings in a liquid account may choose not to participate in K2C because of income uncertainty or risk aversion. For example, Dupas and Robinson (2011) finds that savings accounts earmarked for preventitive health technology purchases decrease savings compared to liquid accounts, though earmarked accounts for health emergencies increased savings. Furthermore, the fact that families and their children may move between cities, school districts, states, and even countries may raise concerns that there will be significant administrative challenges for recovering saved funds in the future. Finally, in focus groups and surveys, guardians raised concerns that children will receive the funds even if they perform poorly in school and choose not to attend college (Fairbank, Maslin, Maullin, Metz & Associates, 2011).

There are a variety of reasons why guardians may choose not to contribute to their childrens K2C accounts despite the potential for an above-market rate of return. For families with access to mainstream financial instruments, including retail savings accounts, 529 college savings plans, etc., it may make little sense to invest more than \$100 to optimize the K2C incentives and such small incentives may not warrant the perceived administrative hassle. Evidence from a similar pilot program in Oklahoma with more lucrative incentives has failed to generate meaningful private contributions even with a seed investment in each account

[17]Estimating the net present value of K2C match and Save Steady bonus funds and the associated return on investment for savings deposits requires some simplifying assumptions. Assuming families only participate to optimize the match and Save Steady incentive, the optimal saver would save \$10 for five months in a row and then save \$50 in the sixth month to optimally capture the full \$100 match and the Save Steady bonus. Here it will be assumed that savers save as early as possible, which we can interpret as reflecting uncertainty about how long the incentives will be offered. In theory, it would be optimal for savers to save at the last possible moment to optimize returns to the deposits and minimize the opportunity cost of saving in a low interest earmarked financial instrument. It will also be assumed that savers realize their investment, i.e. cash out the account and pay for college 13 years after the first deposit is made. These assumptions yield a risk-free internal rate of return of approximately 9.0%, about five times the rate of a 10 year Treasury bond and +40 times the interest rate of retail savings accounts (without consideration for tax on interest income in traditional accounts).

of \$1000 and uncapped matched contributions (Mason et al., 2010). The small scale of the K2C seed contribution to the accounts and cap on the matched amount may be particularly uninspiring to savers. Nevertheless, it is worth considering that, combining a \$50-\$100 seed, \$100 deposit, \$100 match, and \$100 Save Steady bonus, the net \$350-\$400 would constitute over half a semester of tuition at California community colleges, which enjoy the lowest tuition in the country despite recent tuition increases (LAO, 2009).

Additionally, despite massive outreach efforts, student guardians may not be aware of the K2C program or understand its nuanced rules. This informational barrier may only increase among non-native English speaking families, despite OFE's multilingual outreach efforts. Even among elementary school teachers surveyed during a professional development event, there was significant misunderstanding of program rules. The complex legalese of the FRL consent form was anecdotally cited as a source of confusion (Fairbank, Maslin, Maullin, Metz & Associates, 2011). Given the wide variety of school related programming and overwhelming flow of information transmitted to students' families on a regular basis, information about K2C may be crowded out.

Lastly, there may be mistrust of the City of San Francisco and the major commercial bank managing K2C savings on behalf of families, particularly given the state budget woes and banking scandals that were regularly publicized during the initial years of K2C. Despite these potential explanations for non-participation, however, focus groups and surveys of guardians reveals positive interest and support for K2C.[18]

B Appendix: Postcards and Mailing Procedure

B.1 Validity of Postcards and Mailing Procedure

Postcard reminders were an existing feature of OFE outreach activities and families from the first two waves of K2C had received numerous prior postcard reminders. Only guardians with valid addresses were included in the experiment with addresses invalidated if prior postcards had been returned to sender by the postal service. The return to sender address feedback provided a method for verifying that the appropriate postcards were actually mailed to the

[18]In a phone survery of 300 K2C parents conducted by a professional firm, the average response to level of interest in K2C following a description of the program was very positive - 5.6 out of 7 on a Likert scale from "Not at all interested" to "Very interested." In focus groups, parents expressed high levels of trust for OFE, the Treasurer, and the bank managing the program (Fairbank, Maslin, Maullin, Metz & Associates, 2011)

intended recipients and addresses by the mail fulfillment service contracted by OFE. The unidirectional nature of traditional mail, however, renders it beyond the capability of the author to validate that postcards were received, read, or understood. There is independent survey evidence from before the experiment suggesting that postcards are received and read. In a phone survey of 300 parents conducted by a professional firm, "a letter or postcard in the mail" was identified as the most important source of information of seven options, including other parents, teachers, principals, etc. The median response to whether "you receive a great deal of information from that source [mail] or just a little" was "Yes, a great deal."

B.2 Third Round "Holiday" Messaging:

In the third round, messaging was changed significantly. OFE staff had noticed the seasonal increase in contributions around the holidays in previous years, which they partly attributed to holiday-oriented reminders, and wanted to conduct an experiment that incorporated the holiday theme. Thus, guardians were randomly assigned to one of two treatments or a control. Both treatments encouraged postcard recipients to "give the gift of a college degree" for the holidays. The first treatment was aimed at the guardians themselves and the front read "Do you need a holiday gift idea for your child? How about the gift of a college degree?" The second treatment aimed at encouraging guardians to attract savings contributions from friends and family with the message "Do your friends and family need a holiday gift idea for your child? How about the gift of a college degree?" Identical basic details about incentives and contact information were provided on the back of the postcards for both treatments. The control category did not receive a postcard.

C Appendix: Attrition and Minor Data Issues

In general, the default feature of enrollment in K2C accounts means attrition is not as much of a concern as it often is in randomized trials. Of course, non-use of the accounts in this case may be behaviorally similar to attrition in other studies, though non-use should be orthogonal to treatment assignment and should not bias our estimation. Nevertheless, there were some small data issues worth documenting.

A few accounts included in the available data were dropped. Nine of the accountholders included in the randomization and sent postcards were not included in the updated dataset that included accountholder covariates and were dropped. The deposit data included 153 deposits by 39 accountholders that were not included in the experiment and are discarded.

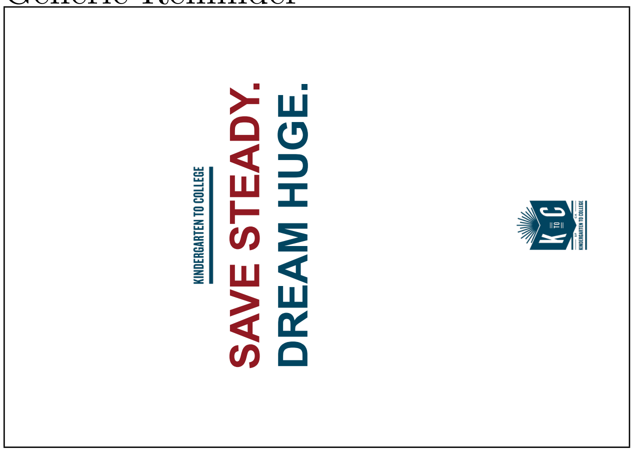
Some problematic accounts were not dropped to maintain the validity of the randomization. Ninety-seven targeted accountholders had database entries updated to indicate "incorrect addresses" likely because postcards bounced back to the sender.[19] These accounts were not dropped because they were not balanced across treatments. Twelve accounts either have new student identification numbers, changed language preferences implying the wrong language was used for the experiment, changed schools, or have contradictory data about the date of their enrollment. These accounts are *not* dropped.

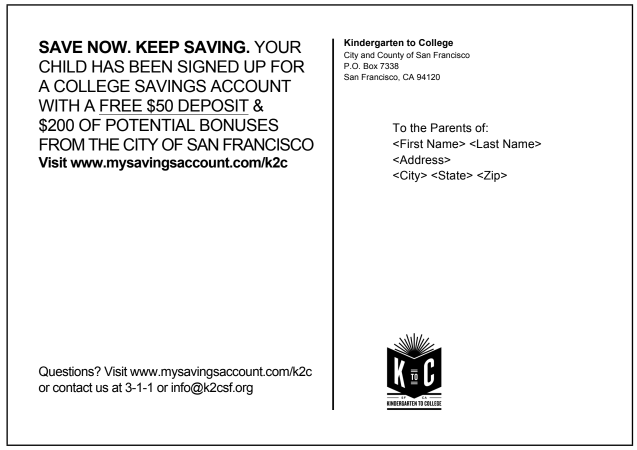
In the first two rounds there was a minor administrative error that caused one Croatian speaker to receive a Chinese reminder and 66 Taiwanese speakers to receive English reminders. The Taiwanese speakers were distributed approximately evenly across treatment groups with 15 in the control group, 14 in the generic group, 19 in the tuition group, and 18 in the financial aid group. The error was corrected in the third round and these accounts were retained without adjustment in the dataset used in this analysis.

[19] OFE staff updated the database when postcards bounced back and not every entry was verified by the author, though dozens were, mostly to validate treatment assignment and language assignment.

D Appendix: Postcard Samples

Generic Reminder



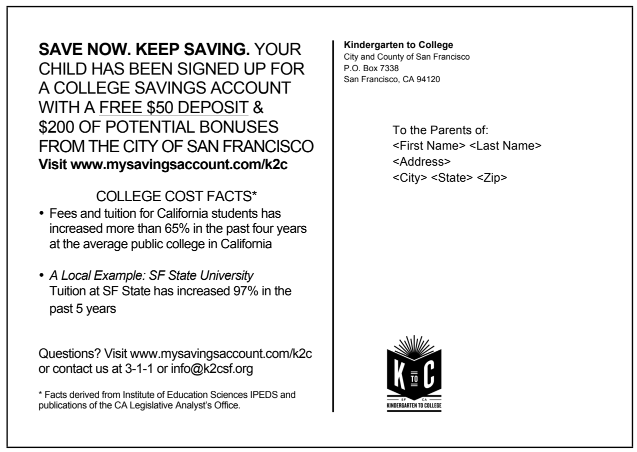

SAVE NOW. KEEP SAVING. YOUR CHILD HAS BEEN SIGNED UP FOR A COLLEGE SAVINGS ACCOUNT WITH A FREE $50 DEPOSIT & $200 OF POTENTIAL BONUSES FROM THE CITY OF SAN FRANCISCO
Visit www.mysavingsaccount.com/k2c

Questions? Visit www.mysavingsaccount.com/k2c or contact us at 3-1-1 or info@k2csf.org

Kindergarten to College
City and County of San Francisco
P.O. Box 7338
San Francisco, CA 94120

To the Parents of:
<First Name> <Last Name>
<Address>
<City> <State> <Zip>

Tuition Reminder


SAVE NOW. KEEP SAVING. YOUR CHILD HAS BEEN SIGNED UP FOR A COLLEGE SAVINGS ACCOUNT WITH A FREE $50 DEPOSIT & $200 OF POTENTIAL BONUSES FROM THE CITY OF SAN FRANCISCO
Visit www.mysavingsaccount.com/k2c

COLLEGE COST FACTS*

- Fees and tuition for California students has increased more than 65% in the past four years at the average public college in California
- *A Local Example: SF State University*
  Tuition at SF State has increased 97% in the past 5 years

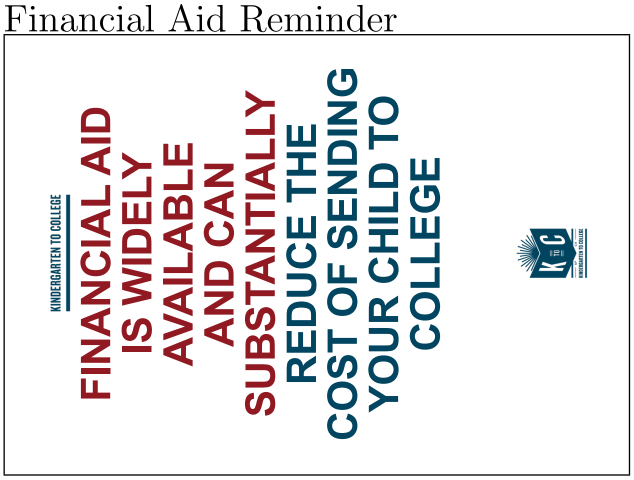
Questions? Visit www.mysavingsaccount.com/k2c or contact us at 3-1-1 or info@k2csf.org

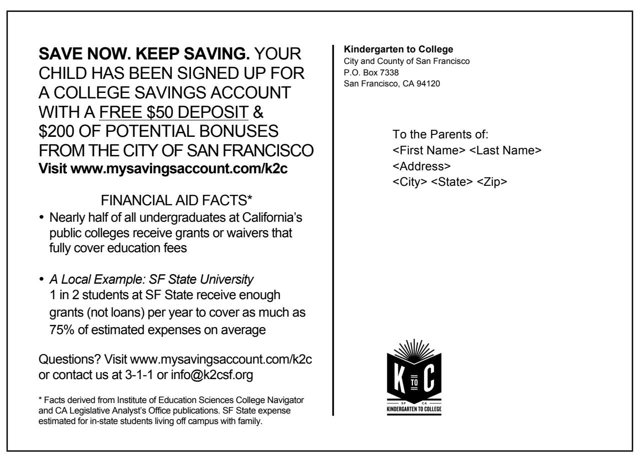
* Facts derived from Institute of Education Sciences IPEDS and publications of the CA Legislative Analyst's Office.

Kindergarten to College
City and County of San Francisco
P.O. Box 7338
San Francisco, CA 94120

To the Parents of:
<First Name> <Last Name>
<Address>
<City> <State> <Zip>

Financial Aid Reminder




SAVE NOW. KEEP SAVING. YOUR CHILD HAS BEEN SIGNED UP FOR A COLLEGE SAVINGS ACCOUNT WITH A FREE $50 DEPOSIT & $200 OF POTENTIAL BONUSES FROM THE CITY OF SAN FRANCISCO
Visit www.mysavingsaccount.com/k2c

FINANCIAL AID FACTS*

- Nearly half of all undergraduates at California's public colleges receive grants or waivers that fully cover education fees
- *A Local Example: SF State University*
  1 in 2 students at SF State receive enough grants (not loans) per year to cover as much as 75% of estimated expenses on average

Questions? Visit www.mysavingsaccount.com/k2c or contact us at 3-1-1 or info@k2csf.org

* Facts derived from Institute of Education Sciences College Navigator and CA Legislative Analyst's Office publications. SF State expense estimated for in-state students living off campus with family.

Kindergarten to College
City and County of San Francisco
P.O. Box 7338
San Francisco, CA 94120

To the Parents of:
<First Name> <Last Name>
<Address>
<City> <State> <Zip>