INSTITUTO DE ECONOMÍA

TESIS de DOCTORADO

2015

Essays in Economic Development

Claudio Alberto Mora.

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA
DOCTORADO EN ECONOMIA

TESIS DE GRADO
DOCTORADO EN ECONOMIA

Mora, García, Claudio Alberto

Junio, 2015

Essays in Economic Development

Claudio Alberto Mora

June 9, 2015

Acknowledgments

I am very grateful to my advisors Jeanne Lafortune and Tomás Rau for their unconditional guidance and support. I also thank Francisco Gallego, Marcela Perticará, Alejandra Traferri, Jeffrey E Harris, Robert A Margo, as well as EH Clio Lab members of the PUC, participants at the Conference in Applied Microeconomics PUC and at the SECHI 2014 meeting, and to all the professors who provided their comments during the construction of this research.

The malaria chapter would not had been possible without the help of Dr. Jose Luis Garces, Luis Rosales-González, and personnel from the Ministry of Health of Costa Rica, who I thank for their help in accessing the malaria campaign archives.

I also thank Cristián Crespo and Nicolas Andrade from the Miniterio de Desarrollo Social, and Andrea López from SENCE, for facilitating access to data from the SEJ chapter.

I owe an outstanding gratitude to all the friends I made at Santiago, and who became an essential part of my life there. Without noticing, they made a huge contribution to this research and to the fulfillment of my PhD. I owe special thanks to Carla Guadalupi for her friendship since the beginning of our doctoral studies. Also Juan Diego Chavarria and Gabriela Tejada for their support. I also owe Pinjas Albagli for several ideas and comments made to this thesis, I will never forget our long and boring (but fascinating for both of us) talks about identification and measurement. Many of other friends who also supported my work: Horacio Valencia, Cristián Figueroa, Vicente Loyola, Daniela Almeida, Faruk Miguel. And to those who showed interest in my work, Andrés García, Lelys Dinarte, Wilber Baires, Diego Escobar, Matías Covarrubias, Sebastián Otero.

More specially, I also thank my friends back in Costa Rica, who despite the distance always remembered this humble friend of theirs with a lot of affection and appreciation.

I owe this thesis to my family, because without their constant support this thesis would have never existed.

Introduction

Two important areas of government intervention are health and labor market programs. This thesis evaluates two public policies whose aim was to change the welfare of the most vulnerable. It particularly studies the existence of heterogeneous effects in public policies. On the one hand I study the consequences of eradicating malaria, and its heterogeneous effects between regions; and on the other I study the decision to take up a social program, and its heterogeneous peer effects among different groups.

Studying this heterogeneity is relevant because it allows to better understand the mechanisms by which the programs reached their consequences. In the case of malaria the estimated benefits of growing in a cohort with much decreased exposure to malaria has been very different in terms of human capital accumulation. While Bleakley (2010b), Cutler et al. (2010), and Lucas (2010) have found benefits to being born in an environment free of malaria, their estimates in terms of educational gains have been very different. Heterogeneity allows to study the interaction between the programs effectiveness and other measures prevalent before the intervention. In the case of the SEJ, heterogeneity allows to explain the mechanisms that helped the peer effects to become relevant.

In the first chapter, I analyze a malaria eradication campaign in Costa Rica that took place around 1940s and successfully lowered malaria rates around the country. I first study the impact of malaria eradication over children born during the eradication campaign and in high malaria regions over years of education, literacy rates, hours worked and wages. But the estimates of malaria eradication over human capital has been very different in similar literature. This is why the most important contribution of this chapter is to analyze heterogeneous effects in both schooling and children labor market conditions. I empirically test if the average marginal increase in the outcomes due to malaria eradication depend on the fraction of children employed and the number of students per school in each region. The second contribution of this chapter is to study the resurgence of malaria during 1967 due to a funding slowdown in the budget to combat mosquitoes, and its heterogeneous effects.

In the second chapter, I analyze the take up decision of a newly introduced Chilean social program called the Youth Employment Subsidy (SEJ) aimed to vulnerable youths. Using a rich and unique dataset, I study a causal endogenous peer effect in the decision to adopt SEJ at the level of the classmates and coworkers. Instrumental variables are used in order to identify causality. The IV exploit one of the eligibility rules, that admissible youths must have a vulnerability score below 11,734 points; the main assumption is that admissibility in a small window around this cut-off was as good as randomly assigned; I then average the fraction of peers at the left of this small window.

The results show important peer effects at the classmates network, that decay six months afterwards and become non-significant. Peer effects are stronger among younger youths inside small groups that graduated from *Técnico Profesional* high schools.

These two studies thus help us understand how learning about heterogeneous effects of government programs is relevant for public policy, as it provides new ways to increase the effectiveness of malaria eradication, and the effectiveness of informational externalities introduced by classmates.

Chapter 1

Can Benefits from Malaria Eradication be Increased? Evidence from the Costa Rican Malaria Campaigns

Abstract

The estimated benefits of growing in a cohort with much decreased exposure to malaria has been very different in terms of human capital accumulation. This paper presents the first estimation of the impact of a malaria eradication campaign in Costa Rica and explores if pre-campaign local characteristics can improve or damage the benefits of the health campaign. Results show that cohorts born during the first eradication campaign had positive gains in terms of human capital accumulation and wages earned by male workers. A one standard deviation decrease in malaria increases the years of education by 1.8%, and increases the wage earned by males by 5.2%. Results are robust to the inclusion of several controls. Point estimates for the malaria resurgence show that these human capital gains were almost completely eliminated when shortage of funding for eradication led to a resurgence of malaria, emphasizing the fragility of the benefits estimated. More importantly, when local characteristics are incorporated in the marginal effect, results show heterogeneous effects where cantons more concerned with educational attendance benefited more and cantons with a more developed child labor market benefited less from the eradication campaign. Motivation for education do not matter when the students are dropping out of school. But the outside option of the labor market seems to play an important role. Hence, health benefits may not be able to translate into educational gains when motivation for schooling is low, or when the child labor market provides a better investment opportunity than schooling.

1.1 Introduction

Malaria eradication has recently became a policy priority (Millennium Development Goals, UN (2013); Roberts and Enserink (2007)). Malaria control campaigns in Africa have decreased malaria mortality rates by more than 25% globally since 2000 (WHO, 2010), costing US$6.8 billion (WHO,

2012) between 2013 and 2015. As a result, the economic literature has given much attention to estimating the impact of growing in a cohort with decreased exposure to malaria. However, while Bleakley (2010b), Cutler et al. (2010), and Lucas (2010) have found benefits to being born in an environment free of malaria, their estimates in terms of educational gains have been very different. The same is true for other health interventions carried out in different countries and years (Maluccio et al., 2009; Bleakley, 2007; Baird et al. 2013; Miguel and Kremer, 2004). Bleakley (2010a) and Pitt, Rosenzweig and Nazmul (2012) have justified these results based on the fact that the added health and productivity could be used in increasing participation in the child labor market and not invested in education. This is the first paper to empirically tests this hypothesis, it does so by exploring heterogeneous effects in the malaria eradication campaign of Costa Rica, which began around 1945 and successfully reduced malaria transmission.

But according to the World Malaria Report 2012 (WHO, 2013) after a rapid expansion between 2004 and 2009, global funding for malaria prevention and control leveled off between 2010 and 2012, and progress in the delivery of some life-saving commodities has slowed. These developments are signs of a funding slowdown that could threaten to reverse the remarkable recent gains in the fight against malaria. Unfortunately no academic work has quantified the possible effect of reversion and has neither compared the benefits from eradication to the costs of its resurgence. The Costa Rican campaign suffered a large set back and is also studied for the first time to understand the impact of the re-introduction of malaria in an economy, and its heterogeneous effects across regions.

The history of malaria in Costa Rica can be divided in two episodes. The first one (called "during" eradication) took place between 1946 and 1963, and successfully lowered malaria rates around the country. However, beginning in 1963, the malaria campaign suffered a funding slowdown and by 1967 there was a new peak in malaria that more than tripled the previous rate. Authorities were not able to control this peak until 1968-70 (the "peak" episode).

Using manually recovered data from the archives and records of the Ministry of Health, this is the first work that explores heterogeneous effects of malaria eradication and resurgence across regions. To do so, this paper first quantifies separately the causal effects that early-life exposure to malaria at the "during" and "peak" episodes has on subsequent economic outcomes as adults—years of education, literacy rates, hours worked per week and monthly wage. Costa Rican censuses are used to obtain information on these outcomes for men and women separately and together, to study the causal effect of the "during" eradication and "peak" episode. In order to identify the effect, I exploit the timing of the campaigns and the pre-campaign variation of malaria rates between cantons. I show that cantons that benefited the most from the campaign where those with the highest malaria infection rates several years before the beginning of the campaign, in 1929; I also show that cantons that suffered the most from the malaria resurgence where those with the highest malaria infection rates in 1956.

Difference-in-differences results show that both men and women born during the first eradication campaign in regions where malaria was higher had positive gains in years of education and literacy rates due to the campaign. On average, a one standard deviation reduction in the malaria rate due to the first eradication campaign increased the years of education of women and men born during the eradication by 2%. These cohorts do not seem to have changed the amount of hours worked per week, but working men had a significant increase of 5.2% in the weekly wage earned.

The results are robust when other programs are taken into account, and whose intensity across regions could had been somehow correlated with the malaria intensity before the campaign. For example, the foundation of the health insurance institution "Caja Costarricense de Seguro Social" (C.C.S.S.) near 1943, and to the *Guerra del 48* and its proceeding school construction program. A visual test shows that the parallel trends assumption holds, and results are also robust to regional convergence. Other diseases not affected by indoor residual spraying are used as placebo, as to show that it was the spraying campaign and not other things that improved that caused the reduction of malaria through an effect over mosquitoes.

As for the peak episode, the point estimates show that, on average, a one standard deviation increase in the malaria rate due to a funding slowdown reduced the human capital stock of men and women; years of education decreased by up to 1.1% for women and 1.4% for men, literacy rates also fell by 1.3% for women and 0.2% for men. Moreover, results show significant evidence that men had to reduce the amount of hours worked by 0.36% in response to a one standard deviation increase in malaria. When comparing the results with the first campaign, men lost more than women during the resurgence of malaria, but gain almost the same with its eradication; moreover, despite the temporary nature of its resurgence, human capital gains were almost completely eliminated when shortage of funding for eradication led to a resurgence of malaria, emphasizing the fragility of the eradication.

The results from the eradication campaign can be compared to other similar works. Bleakley (2010) studies United States, Brazil, Colombia and Mexico. For the Latin American countries, he finds mixed evidence for the literacy and years of education of cohorts more exposed to the eradication efforts as children. But finds conclusive evidence of an increase in earnings for those countries and for the United States. On the other hand, Lucas (2010) studies Sri Lanka and Paraguay and finds that regions with the highest pre-eradication malaria rates experienced the largest gains in education as measured by years of completed primary schooling and literacy, despite the potential differences between the countries. Finally, Cutler et al. (2010) studies India, and finds no gains in literacy or primary school completion of men, and some evidence that women reduced literacy and primary schooling.

The hypothesis so far (Bleakley (2010) and Cutler et al. (2010)) to explain the mixed results in human capital accumulation is that malaria could affect children's productivity in both education and work. As in Grossman (1972), better health means more time available to spend between education, work and leisure. How much of this time will be invested in schooling remains an empirical question. In an economy where children can easily find a job the outside option of school gains value and children will prefer to use their improved health to enter into the child labor market, hence achieving fewer years of education and literacy rates. On the other hand, in an economy more concerned with educational attendance, the outside option lose value and health benefits may be more likely to generate higher human capital investments.

This is the first paper that empirically explores these alternatives by exploiting characteristics between cantons before the campaign in terms of educational motivation (number of students per school) and the degree of development of the child labor market (fraction of children employed), and interacting the marginal effect of the malaria with these two proxies. Even though the results for Costa Rica show an important average impact of malaria eradication and resurgence over human capital and labor market outcomes, these effects could be heterogeneous across regions in Costa Rica, depending on the initial conditions of each region. The heterogeneous results show that the marginal benefits of reducing malaria over years of education of both men and women are bigger in cantons with more students per school and with less children employed. Results for wages are mixed. As for the resurgence in malaria, results show no statistically significant evidence that the schooling conditions affect any of the outcomes of women, nor any of the human capital accumulation of men. This means that when children are dropping out of school because of malaria, they do it uniformly, no matter how good or bad the schooling conditions were.

This is not the first work to study the impact of improving the health environment during the first years of life over long-term human capital accumulation. The most similar works are Bleakley (2007), Bleakley (2010), Lucas (2010), and Cutler et al. (2010) who focus mainly on quantifying the long-term benefits of hookworms and malaria eradication campaigns in different countries. Many other authors have focused on estimating the effect of malaria. Among them, Barreca (2010) who uses instrumental variables and finds for the United States that in utero and postnatal exposure to malaria led to considerably lower levels of educational attainment and higher rates of poverty later in life. Venkataramani (2012) finds positive gains on cognition results in Mexico, Gallup and Sachs (2000) finds that countries with intensive malaria had income levels in 1995 of only 33% of countries without malaria. Chang et al. (2011) study colonial Taiwan and finds that malaria exposure leads to lower life-time educational attainment and to worse mental and physical health outcomes in old age. Rawlings (2012) analyzes the selection versus scarring effects of an unforeseen malaria epidemic in North East Brazil in 1938-1940 on subsequent human capital attainment. Barofsky et al. (2010) study the human capital and income consequences of a malaria eradication campaign in the Ugandan district of Kigezi. However none of them study heterogeneous effects between regions, i.e. the interaction between the marginal benefits of eradicating malaria and the pre-campaign cantonal characteristics in the schooling and labor market sectors to account for the different estimates.

On the other hand, this work is also related to the health literature that have attempted to evaluate the health consequences of malaria. Hong (2007) shows that Union Army recruits who spent their early years in malaria-endemic counties were shorter at enlistment due to malnutrition and were more susceptible to infections during the U.S. Civil War. However, the current work does not argue that the benefits found happen solely through a health channel but, instead, this work focuses on estimating a reduced form.

Other authors have focused on evaluating the impact of random health interventions (Miguel and Kremer, 2004; Baird et al., 2013; Maluccio et al. 2009; among others). Our present setup is not a randomized control trial, but instead focuses on a campaign that lowered the number of malaria cases, similar to the hookworm campaign evaluated in Bleakley (2007).

The next section will describe the history of malaria in Costa Rica, and explains the eradication campaign and the resurgence of malaria due to the funding slowdown. Section 1.3 describes the medical background on the malaria disease. Section 1.4 describes the sources of the data used in this work. Section 1.5 describes the characteristics that make up the research design. Section 1.6 explains the empirical strategy used to identify the effects of the first eradication campaign by canton, over outcomes of interest; present the results and robustness tests. Section 1.7 explains the empirical strategy used to identify the effect of the malaria resurgence, presents its results, and robustness tests. Finally, Section 1.8 presents the main results when interacting the marginal benefits with the schooling and child labor market proxies. Finally, Section 1.9 presents the conclusions.

1.2 A history of malaria in Costa Rica

This section describes the history of malaria in Costa Rica. Information on the malaria campaign was retrieved from the archives at the Ministry of Health (1930, 1939, 1940-1956, 1963, 1967), interviews undertaken from personnel that were on charge of the campaign, and several books and reports that described the campaign (Ministerio de Salud, 1956, 1970, 1973, 1974, 1981).

The first efforts to eradicate malaria began in 1925 when the first nationally malariometric survey was undertaken with the support of the Rockefeller Foundation (RF). The objective was to "determine prevalence of infection, nature of mosquito breeding, indicated methods of control, etc." (Rockefeller Foundation, 1925, pg. 24). "...children were examined for the presence of the plasmodia in their blood and for splenic enlargement..." (pg. 182). A malaria rate of 629 per 10,000 inhabitants was estimated (Ministerio de Salud, 2001). In 1929, the first available measure of malaria morbidity disaggregated at the canton unit (Ministerio de Salud, 1930), revealed a lower morbidity rate of 60.3 per 10,000 habitants. On July 9th, 1958, former minister of health, Dr. Antonio Jimenez Guard, describes the scenario of malaria during 1920 as follows:

> *En esos negros días poblaciones enteras eran azotadas por las "calenturas", y los enfermos amarillentos y tiritando se amontonaban en los salones de los hospitales, implorando unas cápsulas de quinina, como alivio transitorio y temporal* (La Nación, 1956).

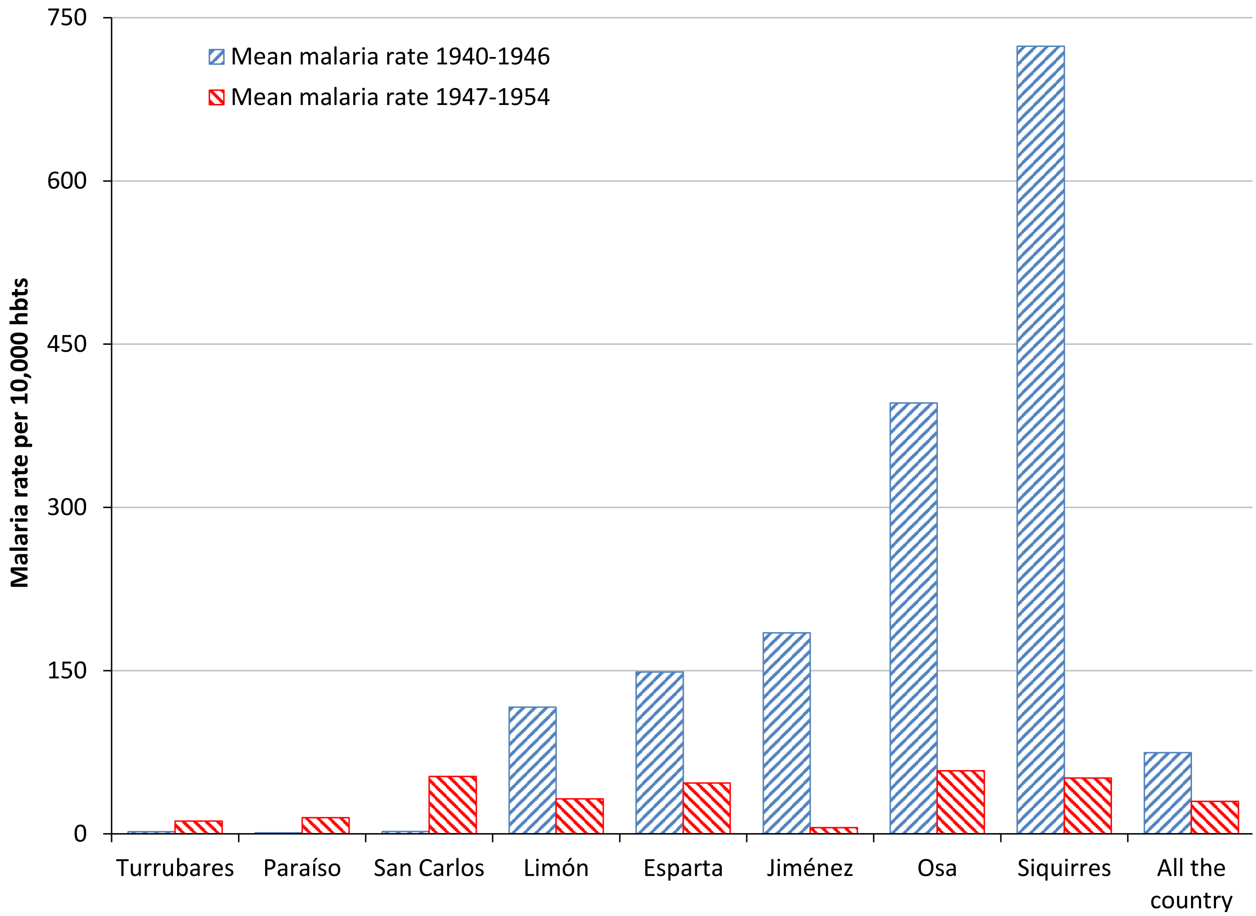
By 1938, Kumm and Ruiz (1939), made the first epidemiological malaria study, and found that the highest spleen rates were concentrated in Guanacaste and in the part of the Province of Puntarenas which occupies the southern end of the peninsula of Nicoya. According to the study, the localities with a high degree of endemic malaria were confined to areas with an elevation of less than 1,000 feet (pgs. 433-4). Of the 9,126 children examined, 1,240 per 10,000 had enlarged spleens. There are no records in The Rockefeller Foundation Annual Reports (1925-1937) or by the Ministry of Health, of any concrete effort to eradicate malaria between 1929 and 1938.

During 1938-1939 the country made its first efforts in addressing the problem through concrete actions that included modifying the mosquito environment via elimination of mosquito breeding sites (OPS, 2001), but with weak and ineffective weapons that resulted in neither positive nor durable results (La Nación, 1956).

In the year of 1946 the United Fruit Company (UFCo) began malaria control with the application of DDT as a measure to increase the productivity of their employees, but in a much focused way in houses inside their banana farms. In 1950, the government of Costa Rica began the first sprays with DDT. The sprayings began nationally after 1953. To the extent of my knowledge, the efforts to eradicate malaria between 1938 and 1956 did not follow any rule. Spraying was disorganized.

Figure 1 shows malaria morbidity rates per 10,000 habitants before and after the commencement of the eradication program, for the three least and five most malarious cantons, and the national mean rates. The figure shows that the program effectively reduced the malaria rates from 1940-46 to 1947-54. Moreover, malaria rates after the eradication campaign began (1946-54) fell more in cantons with the highest pre-eradication malaria rates (1940-1946). For example, the most malarious canton during 1940-46 was Siquirres with 723.8 malaria cases per 10,000 habitants and whose malaria rate was reduced in 92.9% to 51.4 malaria cases per 10,000 habitants. On the other hand, malaria rates remained quite constant among cantons with the lowest pre-campaign malaria rates. Malaria rates around the country were also reduced. This work refers to the period 1947-1963 as the "during" eradication campaign episode.

Figure 1.1: Mean malaria morbility rate per 10,000 habitants for the least and most malarious cantons before and during the first eradication campaign (1940-46).

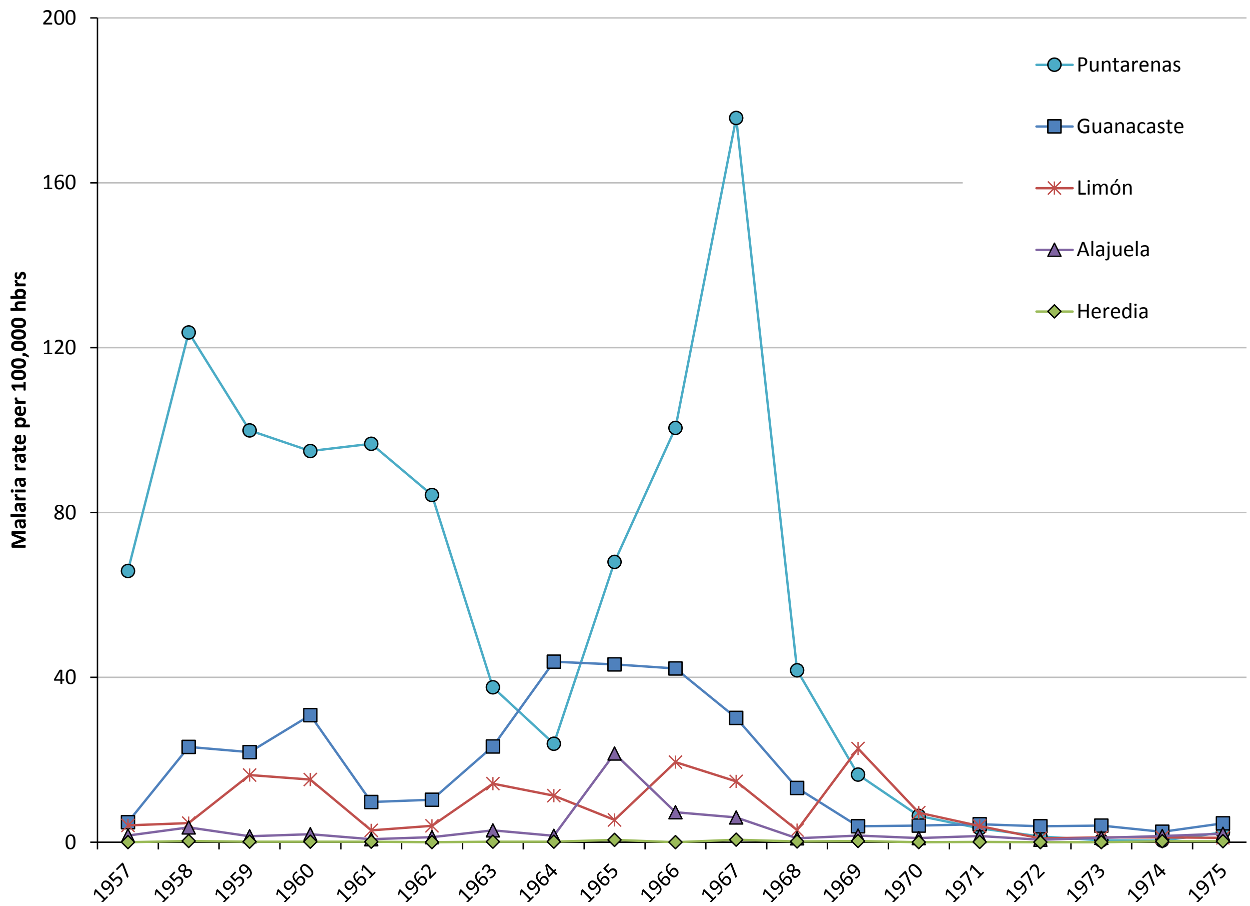


Source: author elaboration based on Memorias de Salud (different years) and WHO.

With the official start of the "national campaign" in 1956, according to Ministry of Health directors, the authorities better organized the already undergoing efforts to eradicate malaria. According to OPS (2001), sprayings were coordinated with this campaign, and the new strategy was the application of DDT in semiannual cycles and comprehensive coverage of malarious areas. The project was launched in 1957, when the National Service for Malaria Eradication (SNEM) was created. Each year the SNEM surveyed each canton for malaria incidence, and the data was used to control the DDT sprayings and other actions. Cantons were malaria rates were higher were more sprayed with DDT, and each canton was sprayed until its malaria rate was reduced to zero. By July 1962, the transmission had been interrupted in 74% of the originally considered malarious areas.

Figure 2 shows malaria morbidity rates per 10,000 habitants aggregated at the province unit, it shows the main result of the national eradication program. The most malarious province, the province of Puntarenas[1], had the biggest reduction of the malaria rate, while the malaria rate in the other provinces did not increase. But things did not work out as planned and by 1963, according to OPS (2001, pg. 13), the SNEM "program deteriorated due to administrative and financial reasons". The SNEM suffered from a funding slowdown because former president José Figueres took money out of the SNEM.

Figure 1.2: Malaria morbidity rate per 10,000 habitants by year and province, Costa Rica, 1957-1975.



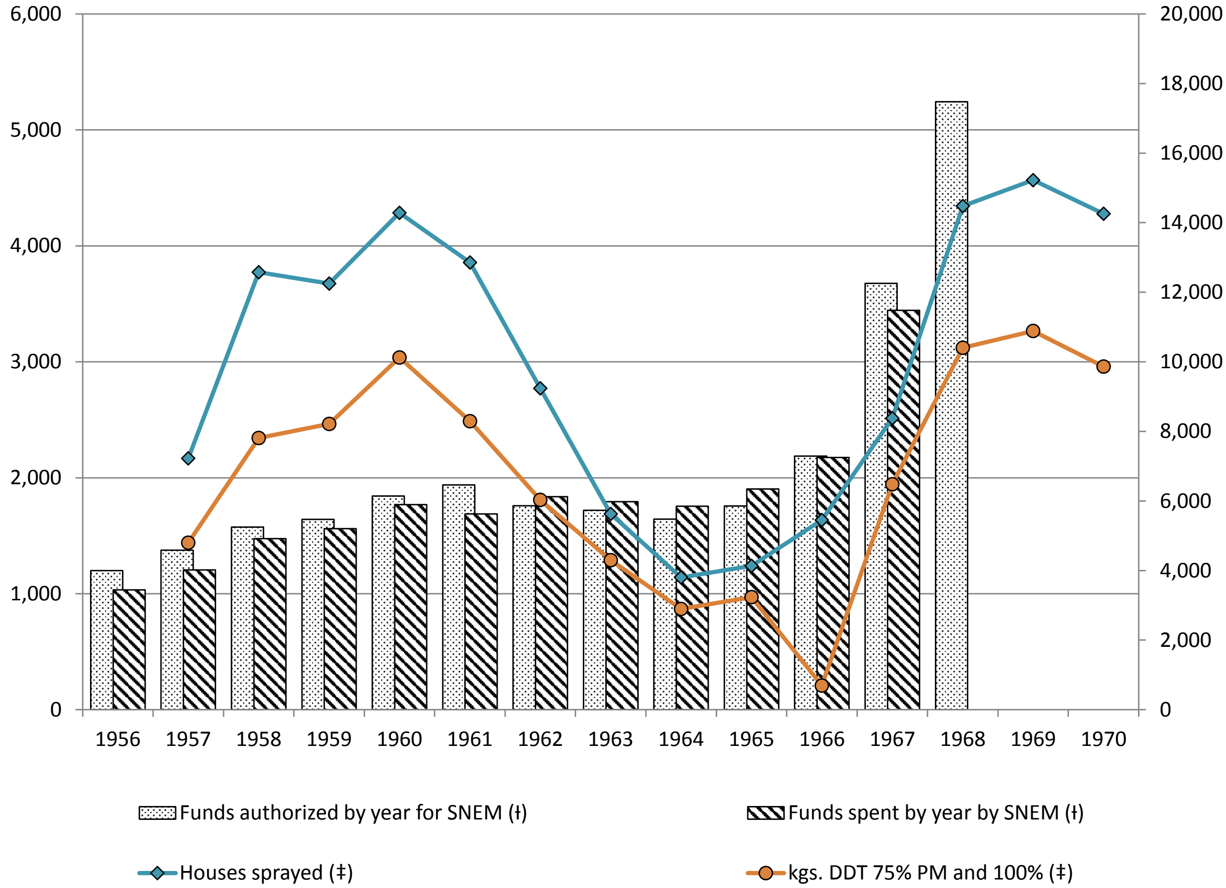
Source: author elaboration based on Memorias de Salud (different years) and WHO.

Figure 3 provides more information regarding the funding slowdown. Funds authorized for the SNEM were reduced beginning in 1962, and the SNEM sustained a deficit for several years. However, it was not enough to avoid the number of houses sprayed and the kilograms of DDT (both at 75% PM and 100%) from falling abruptly. Houses sprayed and kilograms of DDT used did

[1]Ministerio de Salud have enough information in their archives to calculate a time series of the cantonal malaria morbidity rate between 1957 and 1970, however this data was not retrieved by the author.

not increase until 1966, the same year when funds authorized for SNEM increased. Figure 3 also shows great variability in the funds authorized and spent by the Ministry of Health (Salud Pública).

Figure 1.3: Funding slowdown during the malaria eradication campaign of Costa Rica, 1956-1970.



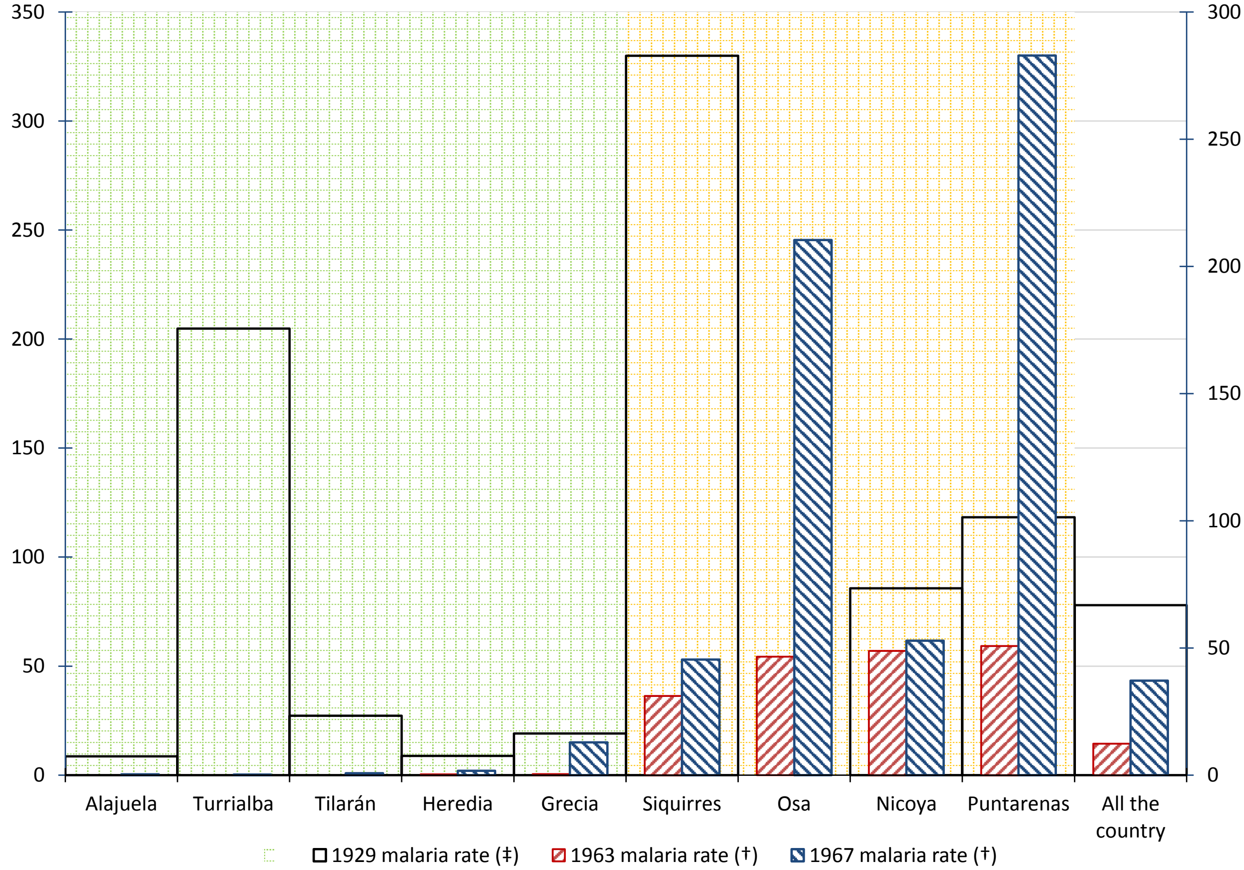
Notes: Salud Publica is the Ministry of Health during the respective years; (†) measured in thousands of colones, left axis; (‡) measured in tenths of colones, right axis.

There is some correlation (with a lag though) between the timing of this episode and, as shown in Figures 2 and A.1, the timing when the national malaria rate raised from 28.2 cases per 10,000 habitants in 1964 to 92.6 cases per 10,000 habitants in 1967. According to Figure 2, the increase was mostly concentrated in the cantons of the province of Puntarenas . This work refers to this event as the "peak" episode (1964-1970), because malaria rates returned to levels seen only before the eradication campaign.

Figure 4 shows mean malaria morbility rates per 10,000 habitants for the least and most malarious cantons before and during the peak episode (1963, 1967). It shows that cantons with the highest malaria rates in 1963 had the biggest increase in the malaria during 1967, while things remained quite stable among cantons with the lowest malaria rates. It also shows that malaria in 1929 was spread in a very different way along the cantons.

By 1968 malaria incidence reverted to levels before the "peak" episode and by 1970 malaria rates

Figure 1.4: Mean malaria morbility rates per 10,000 habitants for the least and most malarious cantons before and during the peak episode (1963, 1967).



Notes: (†) right axis, (‡) left axis.

Source: author elaboration based on Memorias de Salud (different years) and WHO.

achieved the lowest levels never seen before (6.3 per 10,000). For the next 20 years, malaria rates were kept in low levels.

1.3 Medical background on the malaria disease.

This section provides information on the malaria disease. Malaria is caused by contracting one of the several exiting *Plasmodium* parasites, which are transmitted only through direct ingestion with infected blood. One way is being bitten by an infected female *Anopheles* (a family of mosquito) or from a bitten mother to the fetus (while still at uterus). In Central America malaria is mainly caused by three different types of *Plasmodiums*, formally *vivax*, *falciparum* or *malariae*. As for selective attrition (see Almond, 2006), fortunately, the prevalent *Plasmodium* in Costa Rica is the less mortal malariae and vivax. Hence, the threat of mortality for child located at the lower tail of the distribution is not as big as to generate selection bias and influence the results. However, I cannot rule out an important presence of the more deadly parasite *falciparum* (the prevalent parasite in Sub-Saharan Africa). For example, by 1938, 25% of the total malaria cases were *falciparum*[2].

On the other side, recent medical investigations (see for example Singh et al. (1999), Nosten et al. (1999) and more lately Poespoprodjo et al. (2008)) have shown a link between the mother being infected of *vivax* and *falciparum* during any period of pregnancy or delivery and anemia and low weight of the fetus at birth[3]. Independently of the kind of malaria, the consequences of acquiring malaria by the fetus are congenital malaria (transmission of the *Plasmodium* through the placental barrier), abortions, low weight at birth (the most common complication, with a median reduction of 170grs of neonatal weight), prematurity, stillbirths and intrauterine growth retardation [4] (see WHO (2004), Purizaca (2008) and Mendez (1995)). A long-term consequence of contracting malaria at any age is anemia. In untreated cases of chronic malaria, fever, signs and general symptoms usually subside. Months or even years later after infection, dormant forms (hypnozoites) can emerge from the liver and produce new crises (Purizaca, 2008). Anemia leads to a lack of oxygen in the organs (including the brain). It can have negative long term consequences on the development of children, reducing their cognitive and productive skills. Moreover, according to Almond and Currie (2011) and the Fetal Origins Hypothesis (Barker, 1995), better conditions during fetus and first years of life lead to higher weight at birth, less preterm delivery, and increased cognitive skills due to, for example, less risk of anemia (Venkataramani (2012) finds evidence for Mexico that birth year exposure to malaria eradication is associated with increases in Raven Progressive Matrices test scores).

Hence, people who have better health while young will increase their labor productivity as adult. Workers can also be less days absent from work. But the final causal effect over hours worked and wages is a general equilibrium effect. It cannot measure how much people were willing to increase their labor supply, if they were given the opportunity to.

[2]There is no way of knowing what was the distribution of *Plasmodiums* in the population of Costa Rica during 1929, or what parasite was more dominant in each canton. But in the school surveys among children conducted by Kumm and Ruiz (1939) they found *P. falciparum* in 77 of the 168 localities visited (Table 4, pg.435), *vivax* in 62 and *malariae* in 63. Due to their measure there could be over one different type of *Plasmodium*. This does not establish the intensity inside each locality, but it does exhibit that the *falciparum* was very distributed along Costa Rica (especially in Guanacaste). As to the intensity, they found 217 *falciparum* cases (Table 5, pg.436), and 437 *vivax* and *malariae* cases. This could raise some suspects that the *falciparum* was an important problem in Costa Rica.

[3]Low weight at birth has recently been studied by Black et al. (2007) finding significant effects over adult height, IQ at age 18, earnings, and education; with selection bias most likely leading to an understatement of the effects of birth weight on adult outcomes.

[4]Also, according to WHO (2004), in high transmission areas, such as Sub-Saharan Africa, the high malaria infection among pregnant women causes considerable immunity levels that makes them be qualified as semi-immune; hence causing a low maternal mortality. In these areas, the biggest effect is low weight at birth and maternal anemia. In low or stable transmission areas (as Costa Rica), women during fertile age have low or null immunity; hence, pregnant women develop the disease as severe, with central nervous system complications, anemia, abortion, stillbirths and low weight at birth, with higher mortality rates for both the mother and the fetus.

1.4 Data

This section describes the sources of the data used in this work. Data for the outcomes of interest comes from the Costa Rican censuses; there are five censuses available electronically: 1963, 1973, 1984, 2000 and 2011. I use all five of them. This data was retrieved from the web page of the Centro Centroamericano de Población (CCP), which contains public electronic copies of these censuses and other surveys undertaken in Central America and Costa Rica. The CCP uses a PDQ-Explorer service[5]. With the difference from the Integrated Public Use Microdata Series (IPUMS), data is not available at the individual level, but averages of the outcomes are available and can be taken for each census at the sex, cohort, and canton of birth unit; hence this is the unit of measure used in this work. I prefer to use CCP data because it contains information on the canton of birth, while IPUMS only have it at the province level. The available universe of the censuses of 1973-2011 is the full sample; only the census of 1963 contains a 5% representative sample. This is a big improvement in the data used in this work. The usage of the census is an advantage over other studies (Cutler et al., 2001; Lucas, 2011), since it provides information with more statistical power than surveys.

Regarding the definition of the outcomes of interest (years of education, literacy rate, hours of work, and wage), each census has a question asking the years of education, sometimes divided by level attained (primary, secondary, technical secondary, college, university, etc.). In these cases, data was recoded by the total amount of years of education, independently of the kind. People over 5 years of age also were asked if they knew how to read and write. Hence, I can measure fraction of individuals that know how to read and write at each unit of measure and each census. In the censuses of 1973 and 1984 people with a working occupation were asked about the usual amount of hours worked per week. Hence hours of work is measured for those years and is restricted to working people with known hours of work. In the censuses of 1963 and 1973, workers with remuneration were asked about the income accrued during a particular month and the daily or monthly wage. This income or wage was later recoded as the monthly wage.

Moreover, this work also has a richer data set of active and passive detection rates for Costa Rica at the cantonal unit. Data on the number of malaria cases comes from archives of the Ministry of Health (1930, 1939, 1940-1967) and from Kumm and Ruiz (1939)[6]. This data is available at the canton or city unit. When available at the city unit, data was aggregated to the canton unit using the future closest canton division from Hernández (1980).

The pre-campaign cantonal-number of malaria cases in 1929 comes from Ministry of Health (1930). This is a passive detection rate because Ministry of Health (1930) is a registry of the number of patients that presented themselves at each health center and were ruled positive with malaria. The rate was calculated as the cantonal number of malaria cases reported in Ministerio de Salud (1930) divided by cantonal population in 1930 reported in Hernández (1985).

The pre-peak cantonal-number of malaria cases in 1956 comes from PEM (1963). After 1940, information on the number of malaria cases comes from active surveys undertaken by Programa de Erradicación de la Malaria (PEM), which took samples and recorded their results in separate

[5]For more information visit: `http://ccp.ucr.ac.cr/censos/index.php/censos_c/mostrarAyuda`

[6]Before 1940, so far as I know, there are only two "Memorias" available, the 1930s and 1939s, and data from 1929 is far more reliable and had more surveyed cantons than data on 1938.

documents that later on were printed in the Memoria del Ministerio de Salud or Salubridad, depending on the year. The malaria rate was computed as the number of positive slides examined divided by cantonal population in 1963 reported in Hernández (1985).

Outcomes of interest at each unit of measure (canton of birth x cohort x sex x census year) are then associated to the prevalent pre-campaign malaria measure at the canton of birth unit. Hence, early-life health shocks could be studied.

Data on banana potential production capacity (tons per hectare) comes from version 3.0 of the Global Agro-Ecological Zones (GAEZ) project run by the International Institute for Applied Systems Analysis (IIASA) and the FAO (IIASA/FAO 2012). Specifically, the variable is measured using the agro-ecological suitability and productivity for current cultivated land between 1961-1990 with an intermediate input level and gravity irrigation for banana/plantain crop. The GAEZ output is available for each five–arc-minute grid cell on Earth. The land area of such a cell varies by latitude but is 9.2 by 8.5 km at the Tropics (Costinot and Donaldson, 2012, pg. 456). Then the 1930 canton boundaries from Hernandez (1985) were manually matched with the grid cells, taking the mean GAEZ information at each canton. However, there were some cantons without information.

The census of 1927 and 1963 provide information on the number of children attending school. On the other hand, the Megabase of georeferenced data for primary [7] schools of Costa Rica (2000-2009)[8], has information on several characteristics of current existing schools in Costa Rica, among them the year of construction, location and a description if the educational institution is private or public. This information was combined to produce a proxy of the educational motivation at the cantonal unit for the years 1927 and 1963, calculated as the number of children in canton j attending school in 1929 (1963) divided by the mean number schools available in canton j in 1929 (1963). I also use the Megabase to construct the "other schools construction programs". It differs from the educational motivation measure because now it does vary per cohort and measures the mean number of schools that are open in canton j six years after the cohort c was born.

On the other hand, information for the number of children employed and the total number of children, both between 8 and 18 years old, was retrieved at the cantonal level from the censuses of 1927 and 1963. Then the fraction of children employed was calculated by dividing both numbers.

Data on the yearly rate of treated patients and the number of months that the health center remains open comes from Ministerio de Salud (1930). The number of schools during 1927 comes from The Megabase. On the other hand, Instituto de Estadística y Censos (1926) is a statistical compendium that contains information on the number of school facilities opened during 1925. Information on immigration and migration, and from the mean *manzanas*[9] per *finca* and mean *manzanas*

[7] A similar database is available for secondary schools, but it was not used in this work.

[8] This database was created jointly by the Research Program for Sustainable Urban Development (ProDUS) of the University of Costa Rica and the Program Nation State (PEN), during the years 2010 and 2011. The database groups a set of variables for the period 2000-2009 that were previously scattered in multiple bases. It currently contains information provided by the directors of the schools to the Departments of Statistical Analysis and Academic Assessment and Certification from the Ministry of Public Education (MEP). The dataset contains different characteristics at the schooling facility level for those facilities open between 2001 and 2011. It can be downloaded online at `http://www.estadonacion.or.cr/estadisticas/costa-rica/bases-de-datos/bases-en-linea/megabase-de-datos`.

[9] A *manzana* (or apple) is a measurement unit, and in most Central American countries is equivalent to approximately 1.72 acres or 6,961 m^2

per habitants comes from Jiménez (1956) who documented internal migrations during 1950.

Over time the geopolitical division of Costa Rica have changed, hence all cantonal boundaries are manually uniformed using Hernandez (1985) to the boundaries prevalent during 1927 or 1963, depending on the episode under study.

Table 1 presents descriptive statistics of the main variables used in this work. The second column shows means of the respective variable calculated for all the cantons and censuses. The third and fourth columns display means for subsamples separated by cantonal malaria intensity. As can be seen from this table, cantons with higher pre-campaign malaria prevalence had lower years of education and literacy rates, they also had lower wage per month even though they worked almost the same amount of hours per week. Children employment was quite similar to low malaria cantons, but with a higher variance. The number of children per school was also smaller; however there is more heterogeneity within low malaria cantons than within high malaria ones.

This data will be used in estimating the regressions whose research design is described in the next section.

1.5 Research design

The history of the malaria campaigns in Costa Rica has several features that facilitates the identification strategy. First, the timing of the introduction in Costa Rica of policies to combat mosquitoes was exogenous. Worldwide, the construction of the Panama Canal (1905-1910) and the US army occupation of Cuba (1906-1909) were the main causes that led to the discovery of modern tools that combated the transmission of malaria. The discoveries were, first, the DDT in 1939 as an effective insecticide against the *Anopheles* reproduction; and, second, the chloroquine pill against the *Plasmodium* parasites in 1934. The introduction of these tools into Costa Rica by the Rockefeller Foundation, and into the southern territories by the UFCo, due to the reasons explained in Section 2, can be taken as exogenous.

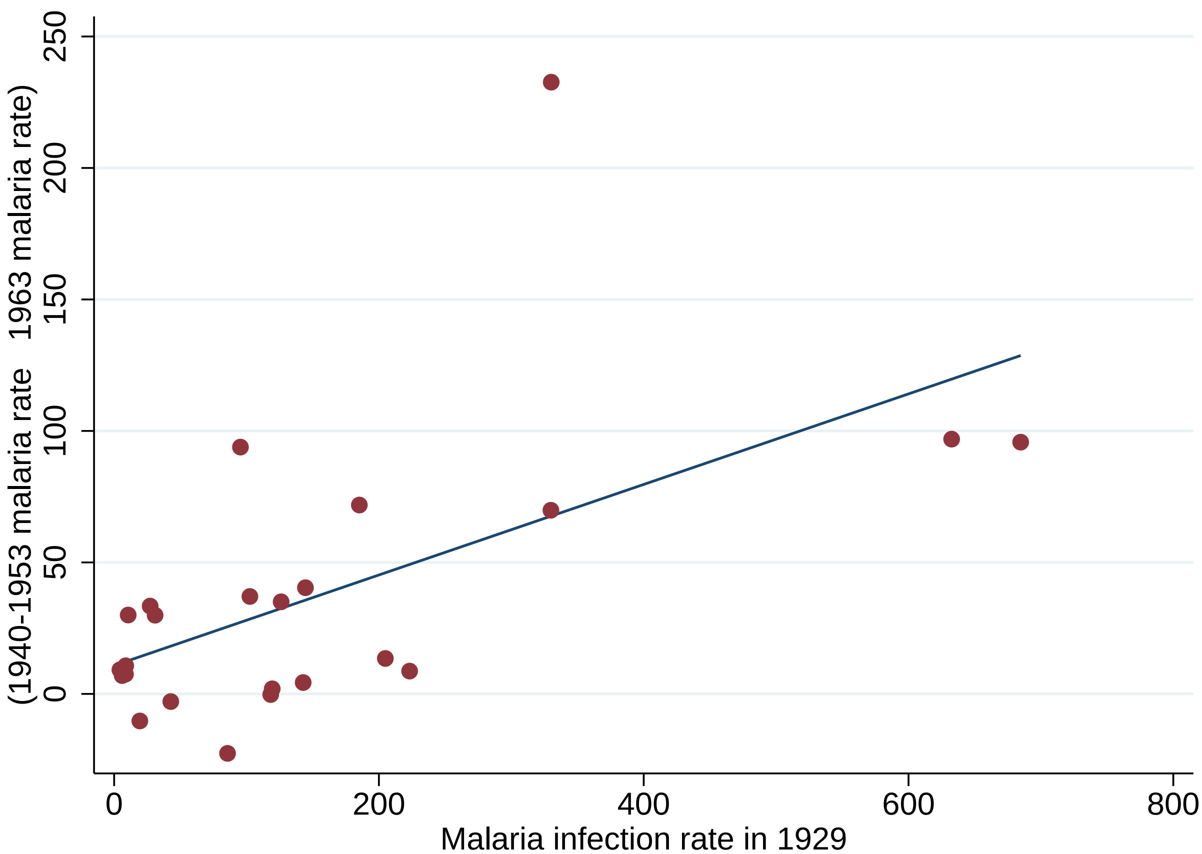
Second, cantonal differences in the ecology of both *Anopheles* and *Plasmodium* induced pre-campaign variation of malaria rates between cantons. Due to the ecology of malaria, high and cold lands as Moravia (1250 m.a.s.l.) are not endemic, whereas low and hot lands as Puntarenas (100 m.a.s.l.) are endemic of the disease. This means that people born in Moravia were not as vulnerable to suffer from malaria as people born near the coastline. But, moreover, malaria ecological differences along coastline and near coastlines cantons introduced differences in the stability of the Anopheles vector.

Finally, the different amount of years of exposure to the campaign by different cohorts. Cohorts born closer to or during the eradication have more to gain than cohorts born several years before the eradication.

A threat that could induce biased estimates is that highly educated cantons made more intense efforts to eradicate malaria than less educated cantons. However, according to published records, the spraying that took place during 1946 by the UFCo was only within their banana farms and, afterwards, by the government of Costa Rica, was made regardless of the possible future educational attainment of the infants or inhabitants compared to those in other regions.

Since this eradication campaign was basically national, this work assumes that "during" the eradication campaign (1947-1966), the cantons that benefited the most from the campaign where those with the highest malaria infection rates in 1929. It does not necessarily assume that the campaign efforts were related to the cantonal malaria incidence. This is adequate for this campaign, because each of the efforts to eradicate malaria before 1956, according to official records, were made independently and spraying with DDT did not follow any rule. Figure 5 shows that this assumption holds, by associating the change between the mean 1940-1953 malaria incidence rate and the 1963 incidence rate with the malaria incidence rate in 1929 for each canton. The linear fit has a slope of 0.17 ($t = 3.37$) with a goodness of fit (R^2) of 0.34. Notice that malaria rates were significantly reduced in all cantons, but cantons with the biggest reductions were those with the highest pre-campaign malaria rate.

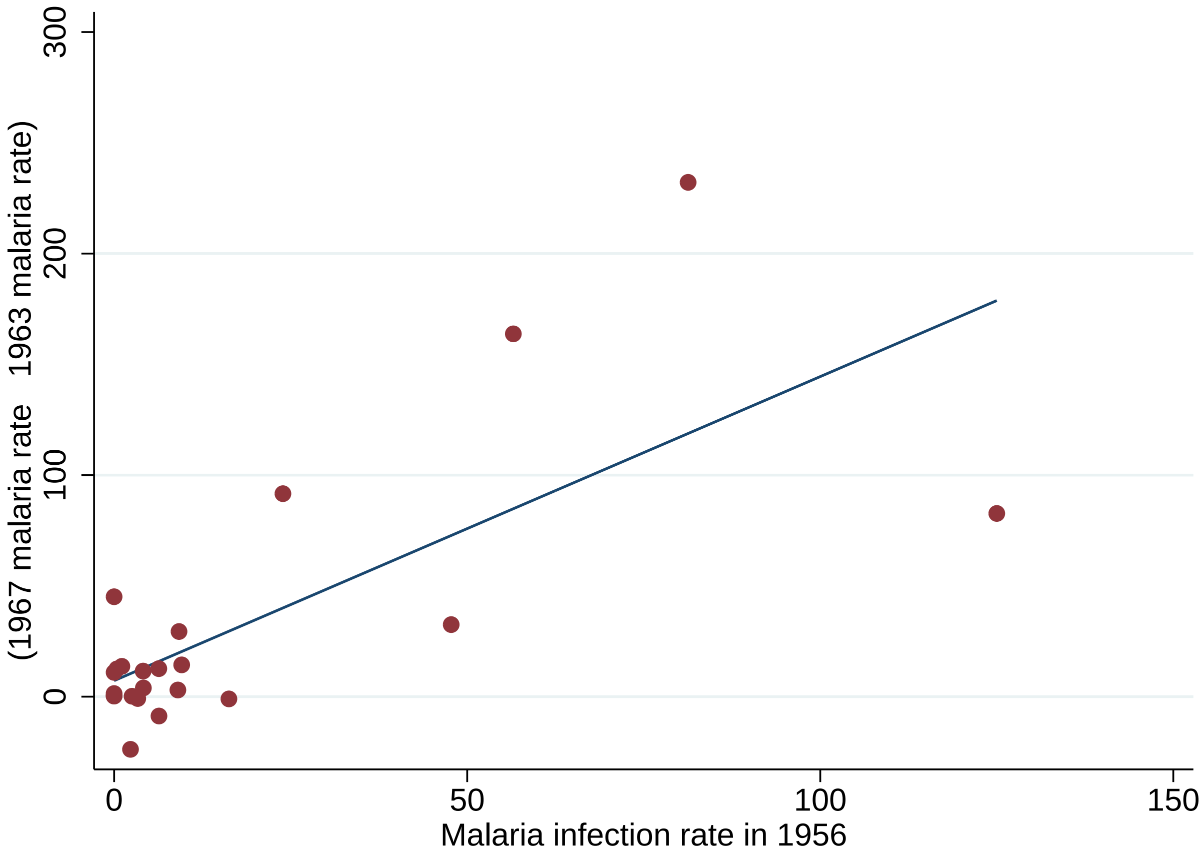
Figure 1.5: Malaria incidence rates of the respective canton "during" eradication.



Notes: Each dot in this figure associates the change between the mean 1940-1953 malaria incidence rate and the 1963 incidence rate with the malaria incidence rate in 1929 for each canton with no missing values. Since the decline is computed as the mean 1940-1953 incidence rate minus the 1963 incidence rate, hence a positive value means a decline and a negative value means an increase. β =0.17, t =3.37, R^2 =0.34.

On the other hand, this work also assumes that cantons that suffered the most from the resurgence of malaria in 1967 where those with the highest malaria infection rates in 1956. Figure 6 shows that this assumption holds. It shows that cantons with higher malaria infection rates in 1956 saw a bigger increase in their malaria infection rates between 1963 and 1967. The linear fit has a slope of 1.37 (t =4.77) and a goodness of fit (R^2) of 0.52. Figure A.3 in the appendix shows results when the outliers are excluded.

Figure 1.6: Malaria incidence rates of the respective canton during malaria "peak".



Note: Each dot in this figure associates the change between the 1963 malaria incidence rate and the 1967 incidence rate with the malaria incidence rate in 1956 for each canton with no missing values. Since the decline is computed as the 1967 incidence rate minus the 1963 incidence rate, hence a negative value means a decline and a positive value means an increase. β =1.37, t =4.77, R^2 =0.52.

This paper uses the malaria rate in 1956 instead of the rate in 1963 because the former is a better predictor of the malaria rate in 1967 than the latter[10] and, as Figure 2 shows, the provincial spread of malaria in 1967 was more similar to 1956 than to 1963.

[10] A linear regression between the change in malaria rate between 1967 and 1963 and the malaria rate in 1956 has a coefficient of 1.37 (t =4.77), while a regression with the rate in 1963 has a coefficient of 2.07 (t =3.76).

1.6 Evidence from the first eradication campaign (1947-1966).

This section describes the identification strategy used to identify the effect of the first malaria eradication campaign that took place between 1947 and 1966, over the outcomes of interest. This work employs a difference-in-differences framework, which allows comparing the incidence malaria rates in a point of time before the program with the long-term evolution of the outcomes of interest of cohorts born around the eradication program. The sample is restricted to both men and women born inside Costa Rica between 15 and 65 years old at the time of the census. This is a reasonable sample, because by the age of 15 most men and women in Costa Rica had finished elementary school or began to desert high school. Cohorts included for the first set of estimates are those born between 1936 and 1966[11]. The next equation is estimated using OLS for males and females separately and together:

$$Y_{jct} = \alpha + \beta(M_j^{Pre} \times During_c) + \delta_c + \delta_j + \delta_t + v_j ct \tag{1.1}$$

where Y_{jct} are the outcomes of interest, formally: years of education, literacy rate, log of hours worked, and log of wage in region j, cohort c, and census year t; M_j^{Pre} is the regional pre-eradication malaria intensity. This captures the idea that areas with high infection rates gained more from the campaign than areas with lower malaria rates. This work uses this passive detection rate and not other active detection rates from 1938 or 1940-1947 due to availability of data: Ministry of Health (1930) is the registry that contains the most number of observations, 58 cantons[12]. The δ_c, δ_t and δ_j are respectively cohort, census and canton fixed effects. The variable $During_c$ is a dummy variable that indicates membership to the eradication cohort. It is defined to take the value of one when year of birth ≥ 1947, one year after the date of commencement of sprayings by the UFCo, 0 otherwise. This is a reasonable date because, as mentioned in the last section, malaria eradication efforts made before 1946 were highly ineffective. Moreover, if there were any reduction in malaria before 1946 due to eradication efforts, these will be playing against finding any effect of the campaign. So, I can take the baseline estimates as a lower bound of the effect of the campaign. In other specifications I also check by adding γX_{jc} into equation (1.1), where X_{jc} is a vector of regional characteristics that vary by cohort and region that will be used as robustness checks. Since there are only 58 cantons and there could be correlation between units of measurement, following Bertrand, Duflo and Mullainathan (2004) the standard errors in equation (1.1) are clustered at the canton unit.

The parameter β quantifies how the change in the outcome after 1946 is related with the malaria rate intensity before the eradication. A positive parameter means that the program induced an increase in each outcome for cohorts born after 1946 and in cantons where malaria was highest.

Columns (1)-(3) of Table 2 show significant and important evidence that both men and women born during the eradication campaign increased their human capital, via an increase in their years of education; columns (4)-(6) suggests that the literacy rate of men also increased. Moreover, even though men and women had similar gains in years of education, benefits in terms of literacy were greater for men than for women. Row (A) presents the results on the coefficient of $M_j^{Pre} \times Post_c$

[11]Hence, I will compare cohorts born before the eradication campaign (1936-1946) with cohorts born during the eradication campaign (1947-1966).

[12]Results using the other measures can be requested from the author electronically.

from estimating equation (1.1). The point estimates of the coefficients in columns (1)-(3) shows that a reduction of one standard deviation in the malaria rate increases the years of education for cohorts born during the eradication campaign (between 1947 and 1966) by around 1.9%, compared to those born earlier. Since the highest malaria regional rate during 1929 was around 685 per 10,000 persons this translates into an increase of 0.51 additional years of education for men and women. Considering that the mean years of education between 1936 and 1966 was around 5[13], this translates into a 10.1% increase for men and women.

The point estimates, although insignificant for women, show that a one standard deviation decrease in malaria increases the literacy rate in around 1% for men and women born between 1946 and 1966. Using the highest malaria rate this means an increase of 4 additional percentage points on the literacy rate of women, and 5.5 additional percentage points for men.

The results for economic outcomes on the subsample of workers in columns (7)-(9) show that cohorts born during the eradication do not seem to have changed the amount of hours worked per week. The result is not driven by gender differences because the program had no significant effect over women or men, separately. Nor it is due to a small sample problem, because the coefficient in column (9) is not significantly different from zero. There could be two explanations for this, (i) men and women were limited either by law or by opportunities in the amount of hours they could work. And, (ii) in the absence of these restrictions, it can be due to a substitution and income effect experienced by workers. Given their higher productivity, both men and women decided to work the same amount of hours per week. In any case, the point estimates of columns (7) and (8) of row (A) imply that a reduction of one standard deviation in the malaria rate reduces the amount of hours worked by women in 0.2% and increases them in 0.3% for men.

On the other hand, the results in columns (10)-(12) show that the program did had a significant positive effect over the wage earned by men born between 1947 and 1966, but the wage of women did not increase. With the current data, I cannot discern if women do not increase their earnings because the campaign did not had any effect over their idiosyncratic productivity (which seems doubtful because their human capital did increase), due to some family composition effect, or something else. The point estimates of columns (10) and (11) of row (A) imply that a reduction of one standard deviation in the malaria rate increases the wage earned by women in 3.2% and in 5.2% for men. Since there is no statistically significant increase in the hours worked of men, the increase in their monthly wage means an increase in their wellbeing.

Altogether, the results suggest that both men and women increased their human capital, but only men were able to exploit the investment at the labor market. This result is understandable if one consider that the Costa Rican cantons were malaria was higher were mostly agricultural and hence it was difficult for women to exploit their increased productivity, which would be consistent with Pitt, Rosenzweig and Hassan (2012).

The results on years of education and wages are compared using a return to investment in education of 10.5% (Psacharopoulos and Patrinos, 2004). A one standard deviation decrease in the malaria rate increases the years of education of men by 0.09, which is translated into an increase of 1% on the wage due to return to investment. However, the result on column (11) shows that

[13] It actually was 4.94 for women and 5.08 for men, but the difference between these is relatively the same.

men increased their wage by much more, 5.2%. Hence, returns to investment of education do not account for the full increase in the wage on men; it was, instead, a combination of many factors.

The next subsection analyzes how robust these results are to the inclusion of different controls, to time trends, and if whether the benefits of the eradication campaign are due to changes on other variables or are solely due to an improvement in the conditions of malaria itself.

1.6.1 Robustness checks

The first robustness check tests if there were other things or programs whose intensity across regions was correlated with malaria before the eradication program began and that promoted positive or negative changes in the outcomes of interests. In this case the diff-in-diffs identification assumption–that without the presence of the program, the units would have had the same trend over time, i.e. there would be no differential changes in educational attainment correlated with levels of malaria before the eradication–would be violated. The malaria eradication program took place in years were many other programs were also in operation. These programs promoted improvements in health and education of children and adults, and their intensity could be related to the malaria intensity, hence changing the slopes of the lines.

There are two main programs that could have induced such changes. First, the foundation of the health insurance institution "Caja Costarricense de Seguro Social" (C.C.S.S.) near 1943, which led to the creation of the invalidity, elderly, and life insurance (known as I.V.M.) during 1947. It initially included a limited set of workers, but by 1961 the universalization of Social Security by the CCSS was approved. And, second, the *Guerra del 48* civil war that led the government of Costa Rica to dismantle his army. It is said that the additional resources were invested in the educational system. Nowadays Costa Rica has an "army of students" instead of an army of weapons. Both of them could threat any identification strategy and baseline results by inducing time and regional variation in the outcomes of interest that is somehow related to $M_j^{Pre} \times Post_c$.

Panel B of Table 2 reports the results when $\delta_p \times \delta_t$ fixed effects at the (province[14] x cohort) unit are included into equation (1.1) to control for observable and unobservable shocks and policy changes that varied by cohort and province. Row (B) shows that the coefficients remain quite similar and gain statistical power. However, literacy rate is not robust. The most stable coefficients are those for years of education, wages and hours worked, whose standard error reduces or, in the worst case, do not change almost at all. Altogether, the reading of these results is that for years of education, wages, and hours worked there are not many other effects correlated by cohort and province with $M_j^{Pre} \times Post_c$ that are driving the results, however there could be some other things that explain literacy.

The second concern is that the parallel trends assumption does not hold because the results are due to trends in the outcomes that would have continued even in the absence of the intervention. Moreover, there can be regional convergence over time between high malaria endemic zones and low endemic zones, because in Costa Rica low malaria endemic zones are located in the heart of the economy, while high endemic zones are located in the periphery of the economy and are usually less developed and have higher poverty rates. To test this, the next equation is estimated by OLS:

[14] See footnote number 5.

$$Y_{jct} = \alpha + \beta M_j^{Pre} + \sum_{c \neq 1946} \gamma_c (\delta_c M_j^{Pre}) + \delta_c + \delta_j + \delta_t + u_{jct} \tag{1.2}$$

In equation (1.2) the impact of the malaria over each cohort c is now $\beta + \gamma_c$. Since the base cohort is set at 1946, the $\hat{\gamma}_c$ measures, for a specific cohort, deviations from the regional base relation of the outcomes and malaria reflected on $\hat{\beta}$. I then plot the series $\hat{\delta}_c$. If the eradication had any impact itself I expect that the shift in the $\hat{\delta}_c$ trend coincides with childhood exposure to the eradication efforts. Once again the standard errors are clustered at the canton unit.

Each dot on the solid line in Figure 6 is the coefficient $\hat{\gamma}_c$ of the interaction between cohort dummies δ_c and the pre-campaign regional malaria intensity M_j^{Pre} from estimating equation (2) over the entire sample for years of education and literacy rate, or the subsample of workers for hours worked and wages. Figure 6 shows that the coefficients on years of education and literacy rate before the campaign kept quite stable around negative or close to zero values, but there was a clear high jump to positive values which coincides with the beginning of the eradication campaign after 1947. The coefficients on wage earned by men show a small, but not significant, increase that coincides with 1947. Altogether, these graphs show that time trends in the outcomes are not an important issue.

Third, are the baseline results driven by other improvements in health conditions rather than malaria? Table 3 presents evidence that suggests this was not the case for the main diseases that Costa Rica had during 1929–formally, tuberculosis and influenza which enter into equation (1) together to save space[15], as additional controls X_{jct} in the form of the interaction between the 1929 passive detection rate and the post-eradication cohort dummy; and hookworms[16], which enter separately.

Typically, influenza is transmitted through the air by coughs or sneezes; it can also be transmitted by direct contact with bird droppings or nasal secretions, or through contact with contaminated surfaces. Tuberculosis is spread through the air when people who have an active TB infection cough, sneeze, or otherwise transmit respiratory fluids through the air. Both of these diseases also work as placebo. Given that DDT spraying only affected mosquitoes, if it was the spraying campaign of DDT that caused the reduction of malaria through an effect over mosquitoes and not other things, then the evolution and regional dispersion of other diseases not transmitted by mosquitoes

[15] I also tried running regressions were each disease enter separately, with similar results for the coefficient of interest. However, they are not presented here to save space.

[16] An eradication campaign against hookworms began around 1914 when the government of Costa Rica extended an invitation to the Rockefeller International Health Commission (RIHC) "to cooperate with that country in work of the relief and control of uncinariasis [hookworms]" (The Rockefeller Foudation, 1913-14, pg.67). In 1915, 64% of the persons examined were found infected. The eradication program consisted on delivering direct curing treatments through pills, and indirect treatments as constructing and forcing the construction of latrines and privies. There was important heterogeneity between the hookworm cantonal rates of infection. On May 28, 1921, the RIHC officially handed over to the government of Costa Rica the program against hookworms with a rate of infection of 29.6% during 1920 (The Rockefeller Foundation, 1921). By 1924, there were 15,326 persons infected with hookworms (The Rockefeller Foundation, 1926). The consequences of infection with hookworms were important: "In Costa Rica, sixty-six laborers before being treated for hookworm disease normally cultivated 563 acres of coffee monthly. After being treated for hookworm disease they cultivated 750 acres, resulting in a net monthly increase in wages of 27 per cent after allowing for a IS per cent reduction in unit pay." (The Rockefeller Foundation, 1918, pgs. 94-5.)

should not be affected with the campaign. The hypothesis is that, if the timing of other general health conditions improved in a way related by time and region to malaria, then the inclusion of each of these diseases will cause a great alteration in our baseline estimate.

Panel B of Table 3 shows that when influenza and tuberculosis are added as controls into equation (1), their estimated coefficients are, in most of the cases, not significant from zero, and when the effect is significant it actually goes in the opposite direction. In most specifications, no significance is lost, and instead the precision of the different estimates are increased.

On the other hand, hookworms (see Mabaso et al. (2003, 2004) and Bleakley (2007)) are spread primarily by walking barefoot on contaminated soil or through the ingestion of larvae. When regression controls for this disease, the coefficient of interest actually lose significance, and change quite more than before. This was an expected effect, because in Costa Rica hookworms are positively correlated with malaria rates (with a correlation coefficient of 0.67), and they are more easily transmitted at sandy soils and warm weather, such as coastlines, where malaria was the highest. This provides more evidence that coastline cantons benefited the most from the eradication campaign.

Finally, are the baseline results driven by other improvements in the banana industry, and health or educational system rather than malaria itself? Perhaps while the UFCo was doing the DDT sprayings and monitoring of the malaria rates, they also discovered some other problems within their banana fields that led to an increased investment in physical capital, rather than human capital. Table 4 includes as controls a proxy of banana productivity, and several measures of the health and educational system. Notice that the baseline results do not change much when banana productivity is accounted for. Results also change little when they are controlled for educational variables, for example due to a school construction program. The results alleviate any concerns of the additional resources due to *Guerra del 48* from the elimination of the army are a confounding factor in the baseline estimates. The effect of the campaign is not correlated to the sample selected, because it varies little when the sample is restricted to younger persons between 15 and 45 years old; the effect also varies little when migration is taken into account, or when oligarchy variables are considered (mean *manzanas* per *finca* and mean *manzanas* per habitant). Both of these measures are interesting because the effect of the program do not depend on how migrant a canton was or how strong was the oligarchy. However, when health variables are included as controls, the baseline results varies. The health variables include the fraction of the population that were treated as patient and the number of months that the health facility remained open. Both measures could be altered by the eradication program since its implementation was very related to the town *medicatura*.

1.7 Evidence from the funding slowdown and malaria resurgence (1964-1970).

The results described in the previous section show that the first eradication campaign successfully increased the human capital stock of both men and women born during the campaign, and that men born during the campaign went on to earn a higher wage. However, what happens when there is a funding slowdown in the eradication campaign and, hence, a resurgence of malaria?

In this section cohorts born during the eradication are compared to cohorts born during the peak episode. Notice that cohorts born during the eradication were also affected by the resurgence of malaria, but not as much as cohorts born during the peak episode, which is the comparison of interest. Next is a description of the identification strategy and the results found for the effect of the funding slowdown and malaria resurgence that took place between 1964 and 1970, over the outcomes of interest.

A strategy similar to the section above is used for estimating its impact; with the difference that now the exogenous timing of the funds depletion is exploited. A diff-in-diffs is employed by estimating the next equation using OLS for males and females separately born between 1956 and 1970[17], between 15 and 65 years old:

$$Y_{jct} = \alpha + \beta(M_j^{1956} \times Peak_c) + \delta_c + \delta_j + \delta_t + v_{jct} \tag{1.3}$$

Each variable is measured as before, but now M_j^1956 refers to the pre-peak malaria intensity in 1956; and $Peak_c$ is a dummy variable that indicates membership to the "peak" cohort. Defining the peak cohort is quite complicated because cohorts born before 1964 were also injured by malaria resurgence, and this will be clear in Figure 7. In any case, I define it to take the value of one when year of birth $\geq$ 1963, the year after the funds depleted, 0 otherwise. As before, in other specifications I also check by adding γX_{jc} into equation (3), where X_{jc} is a vector of regional characteristics that vary by cohort that will be used as robustness checks. I also add $\delta_p \times \delta_t$ fixed effect to capture variation at the province x time level. A negative β means that the peak brought a reduction in each outcome for cohorts born after 1963 and in cantons where malaria was highest.

Table 5 presents the results on the coefficient of $M_j^{1956} \times Peak_c$ from estimating equation (1.3). The point estimates show that the resurgence of malaria reduced the human capital stock of both men and women. Unfortunately, the small sample size might be influencing the significance of the results. The point estimates suggest that an increase of one standard deviation in the malaria rate of 1956 reduced the years of education of women born during the peak episode by 0.20%, and of men by more, 0.68%; it also reduced the literacy rate of women and men 0.23% and 0.11% respectively. Since the highest cantonal malaria rate during 1956 was around 125 cases per 10,000 persons, this means that men lost 0.1 years of education and 0.20 percentage points of literacy rate due to the resurgence of malaria. While the results for women show that they lost 0.03 years of education and 0.43 percentage points of literacy rate.

Column (8) shows evidence that men reduced the amount of time spent at work, in about 0.36% in response to a one standard deviation increase in the malaria rate. Column (7) provide similar evidence for women, but insignificant, who reduced by 0.21% the time spent at work. This seems to suggest that reinfected men had to cut the time they spent working, perhaps due to health problems. Evaluated at the highest cantonal malaria rate during 1956, point estimates for hours worked show that women reduced the time spent at work by 0.41% and men by 0.71%.

Hence, men are more affected than women when there is a resurgence of malaria, but gain almost the same with its eradication. Moreover, in terms of years of education, the losses of men

[17] Hence, I will compare cohorts born during the eradication campaign (1956-1963) with cohorts born during the resurgence of malaria (1964-1970).

from malaria resurgence (0.1 years of education) outset a great deal of the gains accrued from the first campaign (0.5 years of education for men). When the lost years of education are evaluated using the 10.5 measure of return, the results imply that men lost 0.56% of their wage.

1.7.1 Robustness checks

Panel B of Table 5 reports results when fixed effects at the (province x cohort) unit are added to equation (3). The coefficients for hours worked by men, years of education, and literacy rate remain negative and quite stable with some gain in the precision of the estimates. The coefficient for hours worked by women is not very stable. However, this could be due to a small sample size.

Figure 7 plots the coefficients $\hat{\gamma}_c$ from estimating an equation similar to equation (1.2) of the interaction between cohort dummies δ_c and, instead, the pre-peak regional malaria intensity M_j^{Peak}. Since the base cohort is now set at 1963, the $\hat{\gamma}_c$ measures, for a specific cohort, deviations from the regional base relation of the outcomes and malaria reflected on $\hat{beta}$.

This figure shows that before 1963 the relationship between the years of education of men remained stable around zero, but after the resurgence there was a clear break in the pattern to negative values. The coefficients for women show a different story: they were stable around zero until 1961, when they jump to positive values; however there is a downward trend that began in 1962. Men and women seem to be injured in terms of literacy rate even if they were born before 1963. As to hours worked, there is a small change in the trend from positive values before 1963 to negative ones after this year. Hence, Figure 7 should be taken with caution because cohorts born before 1964 were also injured by the resurgence of malaria. Altogether, these graphs show some evidence that the change in the regional relation between the outcomes and malaria coincided with the resurgence of malaria after 1963.

Table 6, similar to Table 3, adds to equation (1.3) information on the hookworms, influenza and tuberculosis diseases infection rates during 1929 interacted with the peak cohort dummy. When influenza and tuberculosis, and when hookworms, are added as controls, the coefficients gain significance and increase in negative terms. Table 7, on the other hand, adds as controls to equation (1.3) the variables listed in each row. The coefficients in Panel B at least do not change the sign of the baseline results on row (A). Similar to Table 4, Rows (B) and (C) show some degree of correlation of the effect of the peak episode on other health variables; and there is also small correlation with banana productivity. Taking into account these two characteristics helps to increase the significance of the coefficient of interest, this means that some variation that was not captured before now is being taken into account. Education variables are not very correlated with the peak consequences, which is interesting because the school construction program (if there was any at all) was already under process after the *Guerra del 48*.

In the next section, I study how schooling and child labor market conditions affect the decision of the children to invest in education, and discuss if the benefits from eradicating malaria can be increased.

1.8 Heterogeneous effects across regions

The last section showed that both men and women born during the first eradication campaign in regions where malaria was higher had on average, a significant positive gain in years of education and literacy rates due to the campaign. As for the peak episode, the point estimates show that, on average, an increase in the malaria rate due to a funding slowdown reduced the human capital stock of men and women. However, other works in the literature have found mixed evidence of malaria eradication over human capital accumulation[18].

The hypothesis so far (Bleakley (2010) and Cutler et al. (2010)) to explain the mixed results around human capital accumulation is that malaria could affect children's productivity in both education and work. As in Grossman (1972), better health means more time available to spend between education, work and leisure. How much of this time would be invested in schooling remains an empirical question. However, there are certain scenarios that ease the prediction. In an economy where children can easily find a job the outside option of school gains value and children will prefer to use their improved health to enter into the child labor market, hence lessening the effectiveness of malaria eradication in achieving higher human capital. On the other hand, in an economy more concerned with educational attendance, the outside option lose value and health benefits may be more likely to generate higher human capital investments.

This section empirically explores these alternatives by exploiting characteristics between cantons before the campaign in terms of educational motivation (number of students per school) and the degree of development of the child labor market (fraction of children employed), and interacting the marginal effect of the malaria with these two proxies. Even though the results for Costa Rica show an important average impact of malaria eradication and resurgence over human capital and labor market outcomes, these effects could be heterogeneous across regions in Costa Rica, depending on the initial conditions of each region. Hence, even if two regions had the same malaria levels, the impact over human capital can be different depending on the schooling and child labor market conditions, due to an heterogeneous effect of the campaigns.

The next equation is estimated by OLS in order to investigate the influence of different cantonal characteristics over the outcomes:

$$Y_{jct} = \alpha + \beta(M_j^{Pre} \times During_c) + \theta(M_j^{Pre} \times During_c \times W_{j,1927}) + \delta_c + \delta_j + \delta_t + v_{jct} \tag{1.4}$$

where $W_{j,1927}$ is (i) a proxy for how interested was the canton in education, as measured by the number of children that were attending regular education during 1927 (measured by the 1927 national census) divided by the number of schools at the same canton during 1927 (measured by ProDUS); or (ii) a proxy for the child labor market, measured by the fraction of children that were

[18] Bleakley (2010) studies United States, Brazil, Colombia and Mexico. For the Latin American countries, he finds mixed evidence for the literacy and years of education of cohorts more exposed to the eradication efforts as children. But finds conclusive evidence of an increase in earnings for those countries and for the United States. On the other hand, Lucas (2010) studies Sri Lanka and Paraguay and finds that regions with the highest pre-eradication malaria rates experienced the largest gains in education as measured by years of completed primary schooling and literacy, despite the potential differences between the countries. Finally, Cutler et al. (2010) studies India, and finds no gains in literacy or primary school completion of men, and some evidence that women reduced literacy and primary schooling.

employed during 1927. Now the marginal effect of the malaria eradication campaign over an outcome $\partial Y_{jct}/\partial M_j^{Pre} \times During_c = \beta + \theta \cdot W_{j,1940}$ also depends on the pre-campaign conditions of the schooling system quality and the child labor market ($\theta \cdot W_{j,1940}$), as a result the effectiveness of the campaign could be lower or higher depending on the initial conditions. The hypothesis to be tested is that when malaria is being eradicated, cantons whose families are more interested in schooling will invest the marginal benefit of malaria in increasing the years of education, hence I expect $\hat{\theta}$ to be positive. Moreover, the marginal gains from eradicating malaria should be lessen in cantons with a more developed child labor market because the outside option of remaining at school is more expensive, hence, in this case I expect $\hat{\theta}$ to be negative for years of education.

Panel A of Table 8 shows the results for the coefficient of the interaction between $M_j^{Pre} \times During_c$ and the proxy of quality of the schooling motivation from equation (1.4). The results in this panel show important heterogeneous effects of malaria eradication over years of education and wages in response to schooling conditions. The marginal increase in years of education due to malaria eradication was higher in cantons with more students per school during 1927, and lower among regions with less students per school, but in both cases it was always positive. As for the subsample of workers, there is evidence that the marginal benefit over wage earned by male workers was higher for those born in cantons with more students per school. All the impact over wages should be coming through the impact over years of education.

Panel B also shows heterogeneous effects in response to the child labor market conditions, this time using the fraction of children employed during 1927. While on average there was a positive marginal increase of malaria eradication over the years of education of female children born in high malaria regions and during the eradication campaign, this marginal increase was bigger among female children born in regions with less children employed, than among female children born in regions with more children employed; but in any case the marginal increase of malaria eradication over the years of education was positive. On the other hand, while on average there was a positive marginal increase of malaria eradication over wages earned by male children born in high malaria regions and during the eradication campaign, this marginal increase was greater among male children born in regions with less children employed, than among children born in regions with more children employed; but in any case the marginal increase of malaria eradication over the wages earned by male workers was positive. Once again, all the impact over wages should be coming through the impact over years of education.

On the other hand, the following equation is estimated for the peak episode:

$$Y_{jct} = \alpha + \beta(M_j^{Peak} \times Peak_c) + \theta(M_j^{Peak} \times Peak_c \times W_{j,1963}) + \delta_c + \delta_j + \delta_t + v_j ct \tag{1.5}$$

Equation (1.5) uses the same measures as before, but evaluated at 1963. Panel A of Table 9 shows no evidence that the number of students per school affected the marginal increase of years of education of women and men due to a resurgence in malaria. This means that when children are dropping out of school, they do it uniformly, no matter if the level of school attendance in the canton was high or low. Column (9) of Panel A shows some evidence that the marginal decrease in the number of hours worked due to a resurgence of malaria was higher in regions with more students per school during 1963. As a result, the children were forced to work less, hence having more time available to spend in other activities, but unfortunately at the same time they were not able to increase the years of education by spending more time at the school. On the other hand, Panel B shows only weak evidence that the degree of development of the child labor market reduced the marginal decrease in years of education of men; there is no statistical evidence that the fraction of children employed during a resurgence of malaria could have any effect over the number of hours worked. Altogether, it is not very clear how heterogeneous effects work during a resurgence of malaria.

1.9 Conclusions

This work has quantified the causal effects that early-life exposure to malaria at the "during" and "peak" episodes had on subsequent economic outcomes as adults–years of education, literacy rates, hours worked per week and monthly wage. The results show that cohorts born during the first eradication campaign had significant positive gains in years of education due to the campaign. However, these cohorts do not seem to have changed the amount of hours worked per week, and there is evidence of an increase only in the weekly wage earned by men. As for the peak episode, point estimates show evidence that the resurgence of malaria reduced the human capital stock of men and women. Comparing the coefficients show that men loose more than women with a resurgence of malaria, but gain almost the same with its eradication. The comparison also suggests that these human capital gains were almost completely eliminated when funds shortage led to a resurgence of malaria emphasizing the fragility of the benefits estimated.

Second, and more important, this work had also has quantified the importance of the schooling conditions and child labor market on the marginal benefits of reducing malaria. Results show that the marginal benefits of reducing malaria over years of education were greater in cantons more concerned with schooling attendance. However, the attendance motivation does not damage the benefits accrued for men in terms of wages. Cantons with a more developed child labor market had lower marginal benefits from malaria eradication in terms of both years of education of women and wages of men.

As for the resurgence in malaria, results show no significant evidence that the schooling conditions affect the years of education neither of women nor men. This means that when children are dropping out of school, they do it uniformly, no matter how big or low is the motivation for school attendance. On the other hand, cantons with a more developed child labor market had bigger marginal benefits from malaria eradication in terms of years of education of men.

This paper suggests that malaria programs will lead to a greater increase in education when they are combined with policies that aim to reduce child labor. It would be interesting to see if similar conclusions could be drown from other health programs in different contexts. Furthermore, the results emphasize the fragility of health prevention campaigns. This is relevant in a world where many diseases that were thought to be extinct are reappearing.

Bibliography

Acemoglu, Daron, James Robinson, and Simon Johnson. "Disease and Development in Historical Perspective." *Journal of the European Economic Association* (Papers and Proceedings of the Seventeenth Annual Congress of the European Economic Association) 1, No. 2/3 (2003): 397-405.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. "Worms at work: long-run impacts of child health gains." Unpublished mimeo. `http://elsa.berkeley.edu/~emiguel/pdfs/miguel_wormsatwork.pdf` (accessed November 12, 2014) (2011) .

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. "How much should we trust differences-in-differences estimates?" *The Quarterly Journal of Economics* 119, No. 1 (2004): 249-275.

Black, Sandra E., Paul J. Devereux, and Kjell G. and Salvanes. "From the cradle to the labor market? The effect of birth weight on adult outcomes." *The Quarterly Journal of Economics* 122, no. 1 (2007): 409-439.

Bleakley, H. "Disease and development: evidence from hookworm eradication in the American South." *The Quarterly Journal of Economics* 122, no. 1 (2007): 73-117.

Bleakley, Hoyt , Dora Costa, and Adriana Lleras-Muney. "Health, Education, and Income in the United States, 1820-2000." In *Human Capital in History: The American Record*, by Leah P Boustan, Carola Frydman and Robert A. Margo. 2014.

Bleakley, Hoyt. "Health, Human Capital, and Development." *Annual Review of Economics* 2 (2010): 283-310.

Bleakley, Hoyt. "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure." *American Economic Journal: Applied Economics* 2, no. 2 (2010): 1-45.

Jiménez Castro, Wilburg. "Migraciones internas en Costa Rica." Washington: Unión Panamericana, 1956.

Centro Centroamericano de Población. *CCP: Censos.* Noviembre 12, 2013. `http://censos.ccp.ucr.ac.cr/` (accessed November 12, 2014)

Chang, Simon, Belton Fleisher, Seonghoon Kim, and Shi-yung Liu. "Long-term Effects of Early Childhood Malaria Exposure on Education and Health: Evidence from Colonial Taiwan." IZA DP No. 5526, 2011.

Costinot, Arnaud, and Dave Dolandson. "Ricardo's Theory of Comparative Advantage: Old Idea, New Evidence." *American Economic Review: Papers & Proceedings* 102, no. 3 (2012): 453-458.

Cutler, David, Winnie Fung, Michael Kremer, Monica Singhal, and Tom Vogl. "Early-life Malaria Exposure and Adult Outcomes: Evidence from Malaria Eradication in India." *American Economic Journal: Applied Economics* 2, no. 2 (2010): 72-94.

Gallup, John L., and Jeffrey D. Sachs. "The Economic Burden of Malaria." CID Working Paper No. 52, 2000.

Hong, Sok Chul . "The Burden of Early Exposure to Malaria in the United States,1850–1860: Malnutrition and Immune Disorders." *The Journal of Economic History* 67, no. 4 (2007): 1001-1035.

Jeremy Barofsky, Claire Chase, Tobenna Anekwe, Farshad Farzadfar. "The economic effects of malaria eradication: Evidence from an intervention in Uganda." Unpublished mimeo. May 2011.

Lucas, Adrienne M. "Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka." *American Economic Journal: Applied Economics* 2, no. 2 (2010): 46-71.

Mabaso, Appleton, Hughes. "Hookworm (Necator Americanus) transmission in inland areas of sandy soils in KwaZulu-Natal, South Africa." *Journal of Tropical Medicine and International Health* 9, no. 4 (2004): 471-476.

Mabaso, M. L. H., Appleton, C. C., Hughes, J. C., & Gouws, E. "The effect of soil type and climate on hookworm (Necator americanus) distribution in KwaZulu-Natal, South Africa." *Tropical Medicine & International Health* 8, no. 8 (2003): 722-727.

Maluccio, J., John Hoddinott, Jere R. Behrman, Reynaldo Martorell, Agnes R. Quisumbing, Aryeh D. Stein, "The Impact of Improving Nutrition During Early Childhood on Education among Guatemalan Adults." *The Economic Journal* 119, no. 537 (2009): 734-763.

Mendez, C. "Malaria during pregnancy: a priority area of malaria research and control." *Parasitology Today* 11, no. 5 (1995): 178-183.

Miguel, Edward, and Michael Kremer. "Worms: identifying impacts on education and health in the presence of treatment externalities." *Econometrica* 72, no. 1 (2004): 159-217.

Ministerio de Salubridad Pública. *Informe de Actividades, año 1954.* San José, Costa Rica: Imprenta Nacional, 1954.

—. *Memoria.* San José, Costa Rica: Imprenta Nacional, 1950, 1955, 1956, 1957.

Ministerio de Salubridad Pública, Servicio Nacional Erradicación de la Malaria. *Plan de erradicación de la malaria, Costa Rica.* San José, Costa Rica, 1956.

Ministerio de Salud. *Diagnostico situacional de la malaria y el uso del DDT en Costa Rica.* San José, Costa Rica, 2001.

Murillo Delgado, Dagoberto. *Megabases de datos georreferenciados para la educación primaria y secundaria de los centros educativos de Costa Rica (2000-2009).* San José, Costa Rica: Estado de la Nación, 2011.

Pitt, Mark M., Mark R. Rosenzweig, and Nazmul Hassan. "Human Capital Investment and the Gender Division of Labor in a Brawn-Based Economy." *The American Economic Review*, 2012 (Dec.), 102(7): 3531–3560.

Programa Erradicación de la Malaria. "Información Histórica." San José: Ministerio de Salud, Archivo Central, 1940-1956, 1963, 1967.

Psacharopoulos, George, and Harry Anthony Patrinos. "Returns to investment in education: a further update." *Education economics* 12, no. 2 (2004): 111-134.

Purizaca, Manuel. "La malaria en la gestación." *Revista peruana de ginecología y obstetricia.* 2008: 131-142.

Secretaría de Salubridad Pública y Protección Social. *Memoria.* San José, Costa Rica, 1930, 1938.

Ruggles, Steven; Alexander, J. Trent; Genadek, Katie; Goeken, Ronald;. Schroeder, Matthew B; and Sobek, Matthew. *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

UN. *The Millennium Development Goals Report 2013.* New York: United Nations, 2013.

Venkataramani, Atheendar S. . "Early life exposure to malaria and cognition in adulthood: Evidence from Mexico." *Journal of Health Economics* 31 (2012): 767-780.

W. Kumm, Henry, and Horacio Ruiz S. "A Malaria Survey of the Republic of Costa Rica, Central America." *The American Journal of Tropical Medicine and Hygiene* 1, no. 19 (September 1939): 425-445.

W. Kumm, Henry, and Horacio Ruiz S. "The Mosquitoes of Costa Rica." *The American Journal of Tropical Medicine and Hygiene* 20, no. 3 (May, 1940): 385-422.

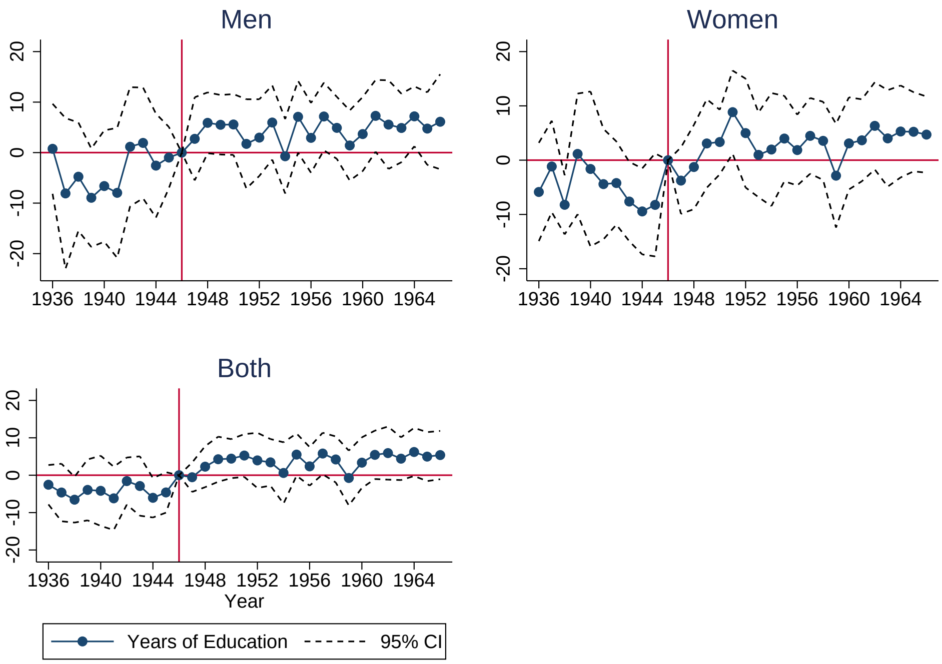
WHO. *A strategic framework of malaria prevention and control during pregnancy in the African Region.* 2004.

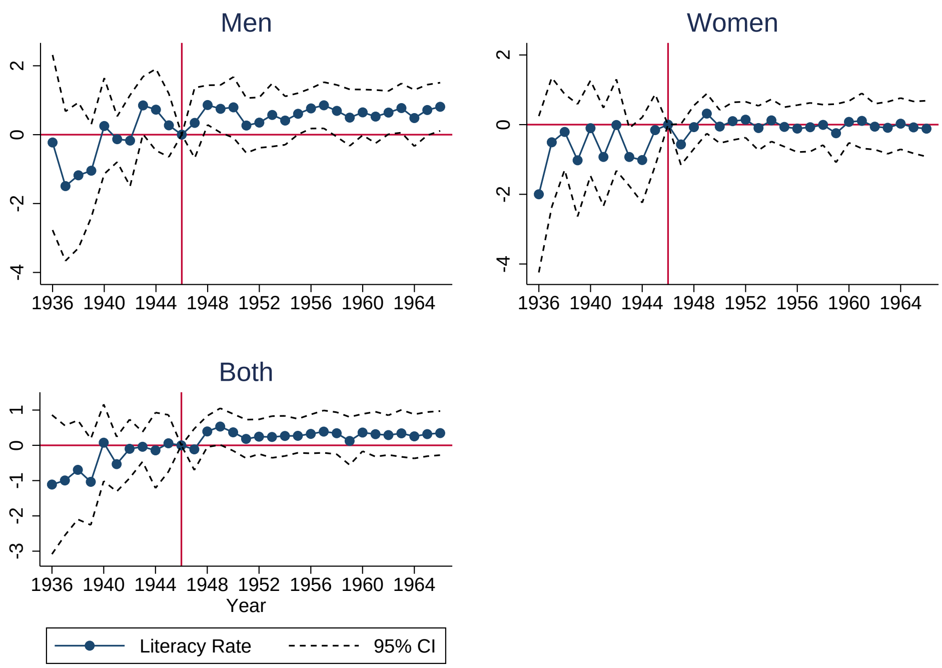
—. "WHO Global Malaria Programme. World Malaria Report 2012." Edited by WHO Press. 2012. `http://www.who.int/malaria/publications/world_malaria_report_2012/wmr2012_no_profiles.pdf` (accessed November 12, 2014)

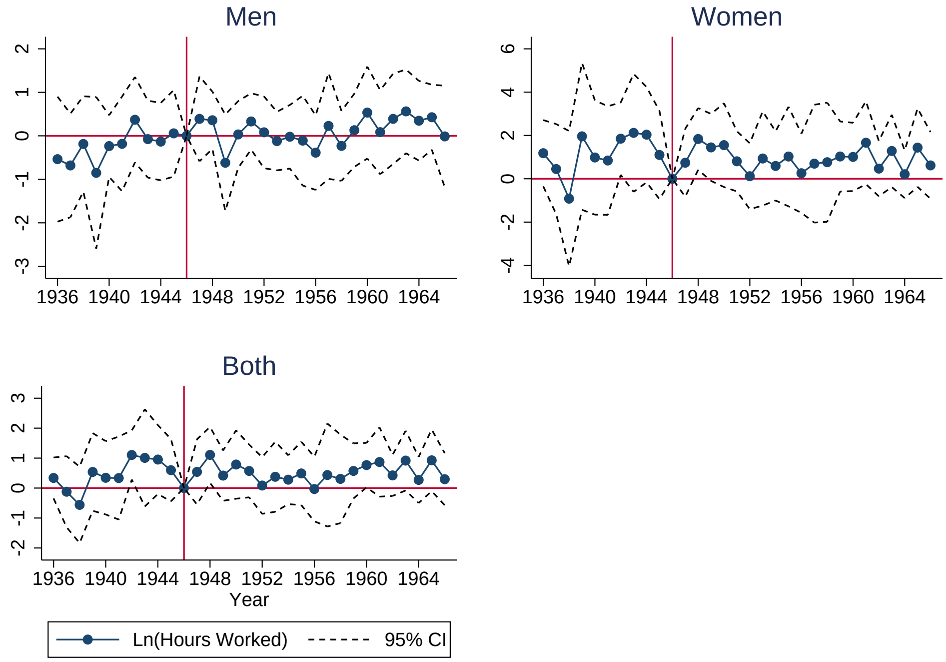
WHO. *World Malaria Report 2010.* Geneva, Switzerland: World Health Organization (WHO), 2010.

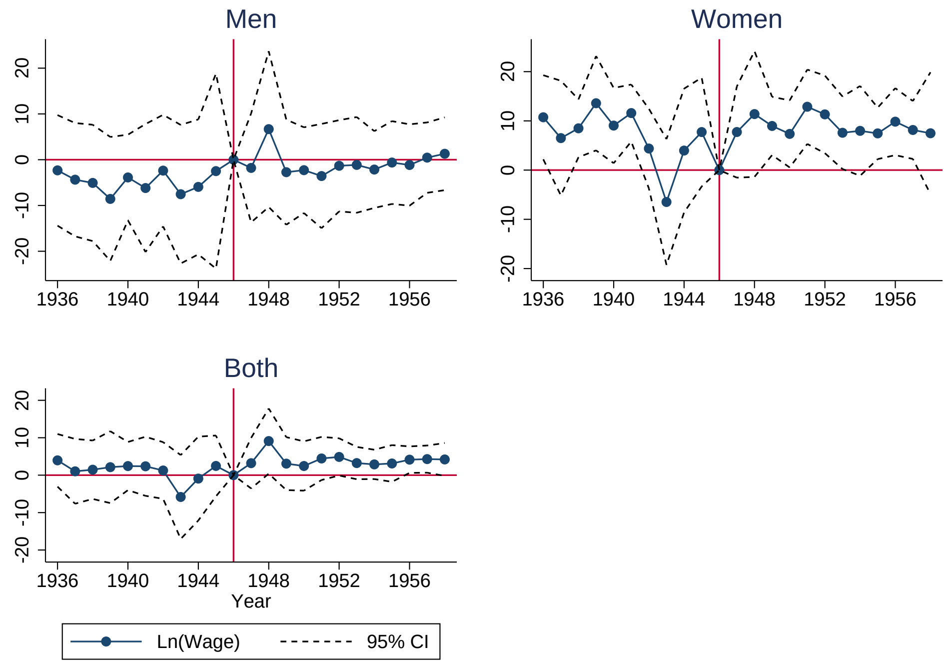
—. "World Malaria Report 2012 FACT SHEET." *WHO Malaria.* December 2012. `http://www.who.int/malaria/publications/world_malaria_report_2012/wmr2012_factsheet.pdf` (accessed November 12, 2014).

Figure 6: Coefficients of the interactions of cohort dummies and malaria rate in the canton of birth in equation (1.2) using pre-eradication campaign malaria infection rate in 1929.



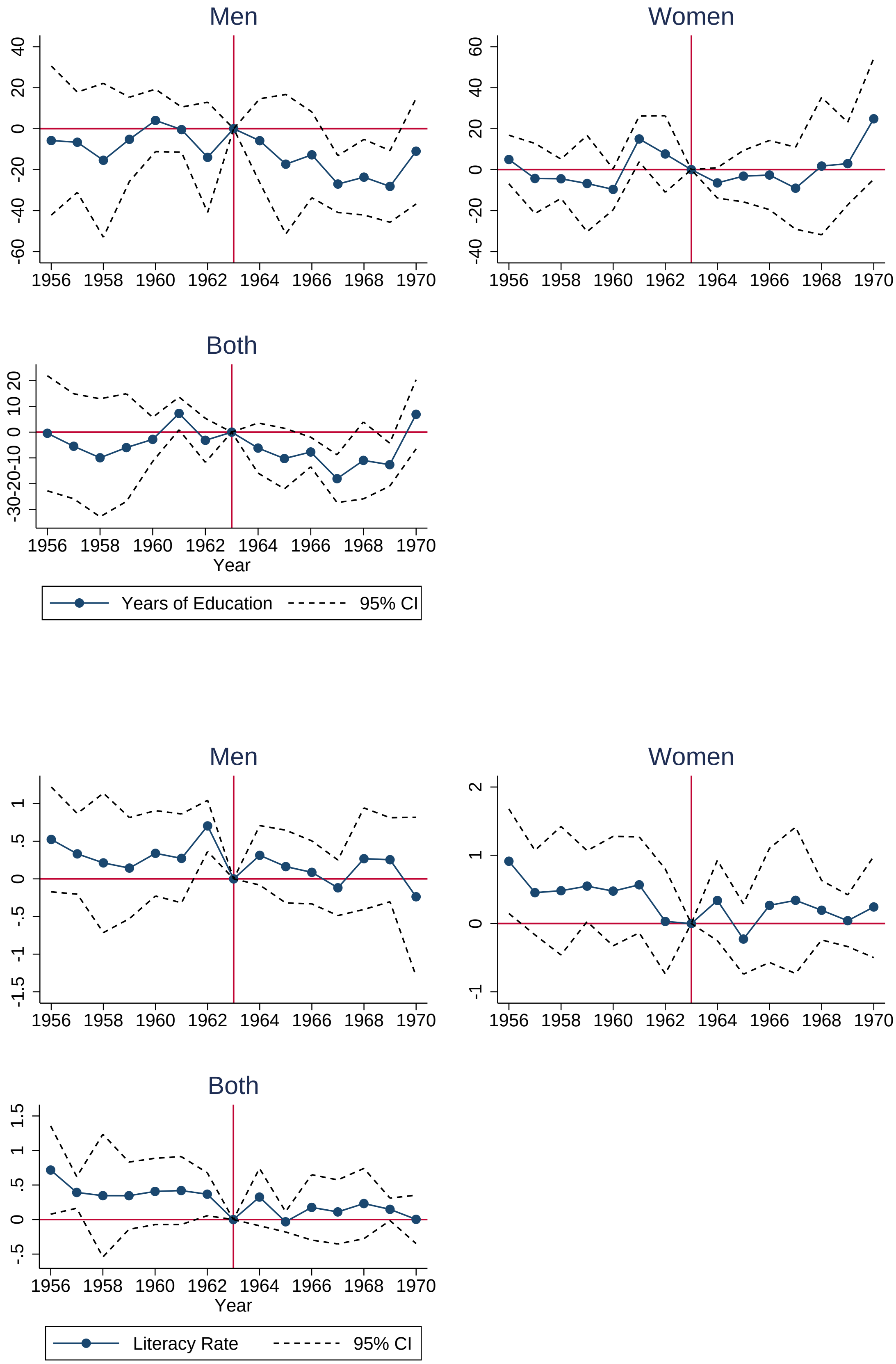


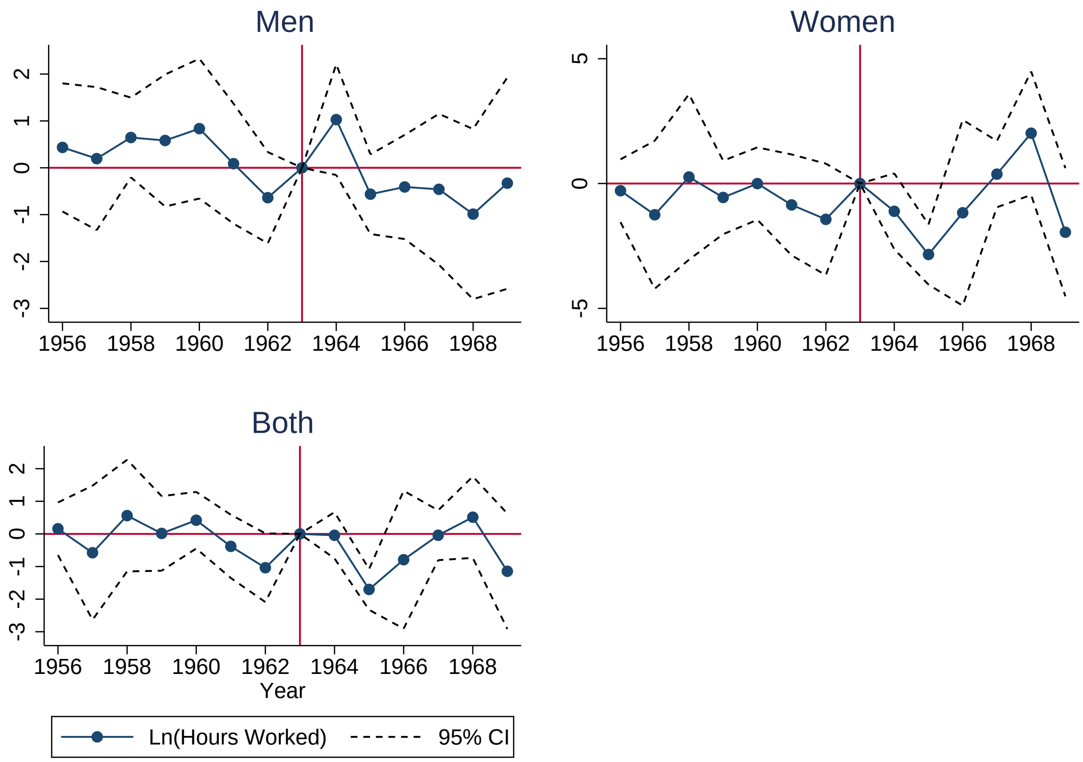




Notes: Each dot on the solid line of these graphs shows coefficients $\hat{\gamma}_c$ of the interaction between cohort dummies δ_c and the pre-campaign regional malaria intensity M_j^{Pre} from estimating equation (2) over the entire sample for years of education and literacy rate, and over the subsample of workers for hours worked and wages. All regressions include cohort, cantonal and census year fixed effects, with standard errors clustered at the canton level. The dotted lines on each graph are the 95% confidence interval, with standard errors clustered at the canton level. The horizontal axis measures the year of birth of each cohort and the vertical axis measures the coefficients.

Figure 7: Coefficients of the interactions of cohort dummies and malaria rate in the canton of birth in a modified equation (1.2) using malaria rate in 1956.





Notes: Each dot on the solid line of these graphs shows coefficients $\hat{\gamma}_c$ of the interaction between cohort dummies δ_c and the pre-campaign regional malaria intensity M_j^{1956} from estimating equation (1.2) over the entire sample for years of education and literacy rate, and over the subsample of workers for hours worked and wages. All regressions include cohort, cantonal and census year fixed effects, with standard errors clustered at the canton level. The dotted lines on each graph are the 95% confidence interval, with standard errors clustered at the canton level. The horizontal axis measures the year of birth of each cohort and the vertical axis measures the coefficients.

Table 1: Descriptive statistics.

	Whole Sample Mean	By Malaria Infection during 1929 High Malaria >Mean	Low Malaria <Mean	Source
Sex	0.50	0.50	0.50	CCP
	[0.50]	[0.50]	[0.50]	
Age	39.79	39.73	39.83	CCP
	[14.68]	[14.68]	[14.68]	
Years of Schooling	6.00	5.66	6.16	CCP
	[2.62]	[2.52]	[2.66]	
Literacy rate	0.89	0.87	0.91	CCP
	[0.16]	[0.18]	[0.15]	
Weekley wage	410.96	410.08	411.41	CCP
	[357.12]	[353.47]	[358.95]	
Weekley hours worked	46.32	46.42	46.27	CCP
	[4.90]	[5.00]	[4.85]	
Pre-eradication malaria infection	74.09	201.44	14.15	Health Ministry
rate per 10,000 hbts, 1929	[125.63]	[159.79]	[12.82]	Archives
Sexual diseases infection rate	155.00	339.55	68.15	Health Ministry
per 10,000 hbts during 1929	[203.83]	[263.70]	[73.88]	Archives
Influenza infection rate	17.57	25.91	12.09	Health Ministry
per 10,000 hbts during 1929	[30.99]	[36.74]	[25.66]	Archives
Tuberculosis infection rate	8.09	6.32	9.29	Health Ministry
per 10,000 hbts during 1929	[11.73]	[5.37]	[14.52]	Archives
Hookworms infection rate	24.15	49.23	9.73	Health Ministry
per 10,000 hbts during 1929	[40.25]	[56.27]	[14.30]	Archives
Mortality, 1929	258.13	250.69	262.35	Health Ministry
	[235.96]	[221.26]	[245.98]	Archives
GAEZ banana potential yield	4.93	4.84	4.98	GAEZ-FAO
	[1.10]	[.81]	[1.23]	
Number of ”stays” at health	110,728.8	27,608.6	160,601.0	Health Ministry
facilities, 1937	[417,382.9]	[24,345.8]	[526,222.4]	Archives
Number of schools, 1927	4.97	5.06	4.92	ProDUS
	[4.66]	[5.26]	[4.31]	
Number of schools, 1925	9.29	9.21	9.33	Health Ministry
	[7.58]	[7.19]	[7.83]	Archives
Child emplyment rate, 1927	0.36	0.35	0.37	CCP
	[0.12]	[0.13]	[0.11]	
Number of children per school,	230.10	109.45	295.24	CCP and ProDUS
1927	[396.27]	[69.87]	[478.14]	
Number of children per school,	100.02	71.83	113.82	CCP and Health
1925	[88.25]	[60.52]	[96.62]	Ministry Archives

Notes: Standard deviations displayed in parentheses below mean. All variables means are calculated using men and women born in Costa Rica between 15 and 65 years old. See the data section for more information on sources and variable construction.

Table 2: Effect of the first malaria campaign on human capital and economic attainment, by sex.

DEPENDENT VARIABLE:	YEARS OF EDUCATION			LITERACY RATE			LN(HOURS WORKED)			LN(WAGE)		
SEX:	Women	Men	All	Women	Men	All	Women	Men	All	Women	Men	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						PANEL A: Baseline Results						
(A) During eradication cohort x	7.1780 **	7.6955 **	7.4389 **	0.5714	0.8004 *	0.6857	-0.1598	0.2707	0.0690	2.5487	4.1393 ***	3.3446 *
1929 malaria infection rate	[3.3475]	[3.2105]	[3.1740]	[0.4165]	[0.4579]	[0.4249]	[0.2517]	[0.1895]	[0.1467]	[3.3452]	[1.1101]	[1.7635]
R-squared	*0.80*	*0.81*	*0.80*	*0.35*	*0.35*	*0.34*	*0.31*	*0.29*	*0.18*	*0.88*	*0.88*	*0.88*
Effect of a 1 s.d. decrease in malaria rate	*1.8%*	*1.9%*	*1.9%*	*0.8%*	*1.1%*	*1.0%*	*-0.2%*	*0.3%*	*0.1%*	*3.2%*	*5.2%*	*4.2%*
						PANEL B: additional (province $\times$ cohort) fixed effects						
(B) During eradication cohort x	9.1597 ***	10.0910 ***	9.6166 ***	-0.8963 ***	-0.7748 ***	-0.8371 ***	0.1945	0.1071	0.1371	2.3843	3.2592 **	2.7431
1929 malaria infection rate	[2.7993]	[3.1298]	[2.8183]	[0.3135]	[0.2511]	[0.2639]	[0.1950]	[0.0837]	[0.0904]	[2.7662]	[1.5886]	[1.7037]
R-squared	*0.81*	*0.82*	*0.81*	*0.38*	*0.40*	*0.37*	*0.36*	*0.34*	*0.20*	*0.89*	*0.89*	*0.88*
Observations	*7,585*	*7,596*	*15,181*	*8,714*	*8,723*	*17,437*	*3,936*	*4,007*	*7,943*	*2,126*	*2,416*	*4,542*

Notes: Each cell in this table shows estimates on the coefficient of the cohort dummy and the malaria rate at the canton of birth, from equation (1.1). The dependent variables are the variables denoted in the column headings, and the independent variables are the variables denoted in each row. All specifications include canton of birth dummies, year of birth dummies, and census year dummies. Clustered standard errors at the canton unit are in parenthesis. *** denotes statistical significance at 1 percent level, ** significance at 5 percent level, * significance at 10 percent level. Panel A shows baseline results, Panel B additionally adds interactions between province of birth dummies and year of birth dummies. Columns (1)-(6) uses information from censuses 1963-2011, columns (7)-(9) uses information from 1973-1984, and columns (10)-(12) from 1963-1973.

Table 3: Sensitivity checks for the first malaria eradication campaign estimates on human capital and economic attainment with additional diseases, by sex.

DEPENDENT VARIABLE:	YEARS OF EDUCATION			LITERACY RATE			LN(HOURS WORKED)			LN(WAGE)		
SEX:	Women	Men	All	Women	Men	All	Women	Men	All	Women	Men	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PANEL A: Baseline Results												
During eradication cohort x 1929												
Malaria Rate	7.1780 **	7.6955 **	7.4389 **	0.5714	0.8004 *	0.6857	-0.1598	0.2707	0.0690	2.5487	4.1393 ***	3.3446 *
	[3.3475]	[3.2105]	[3.1740]	[0.4165]	[0.4579]	[0.4249]	[0.2517]	[0.1895]	[0.1467]	[3.3452]	[1.1101]	[1.7635]
PANEL B: Specifications with other diseases as controls												
During eradication cohort x 1929												
Malaria Rate	12.2240 ***	10.7710 **	11.4910 ***	1.2898 **	1.5227 **	1.4053 **	-0.2136	0.1608	-0.0176	-0.6470	5.5282 ***	2.5907 *
	[3.4248]	[4.0864]	[3.6557]	[0.5742]	[0.7227]	[0.6384]	[0.2809]	[0.2100]	[0.1666]	[2.5310]	[1.2386]	[1.4299]
Influenza Rate	6.5778	7.1599	6.7729	-6.0124 ***	-5.7435 **	-5.8881 **	-0.0978	0.8553	0.3544	8.6183	-3.9228	1.7365
	[16.525]	[17.230]	[16.477]	[2.0240]	[2.6006]	[2.2721]	[1.0474]	[0.7367]	[0.5095]	[12.346]	[5.5646]	[6.2044]
Tuberculosis Rate	45.2600	62.6030	53.5500	-20.4620 ***	-16.1940 ***	-18.3530 ***	-4.1281 *	-1.7985	-2.9356	-5.1795	22.7760	10.0430
	[40.092]	[48.183]	[39.910]	[4.2722]	[4.8062]	[4.4282]	[2.3561]	[2.2954]	[1.8492]	[27.346]	[18.691]	[18.097]
Observations	*4,509*	*4,519*	*9,028*	*5,406*	*5,414*	*10,820*	*2,429*	*2,495*	*4,924*	*1,316*	*1,466*	*2,782*
R-squared	*0.78*	*0.79*	*0.78*	*0.35*	*0.34*	*0.34*	*0.27*	*0.26*	*0.17*	*0.88*	*0.86*	*0.86*
During eradication cohort x 1929												
Malaria Rate	0.8015	2.4662	1.6487	1.5021 **	1.5896 **	1.5468 **	-0.0898	0.5421 *	0.2271	-2.0242	7.2916 ***	2.7587
	[5.6164]	[4.4642]	[4.9190]	[0.6853]	[0.7766]	[0.7221]	[0.3107]	[0.3096]	[0.1659]	[3.9183]	[1.9725]	[2.2343]
Hookworms Rate	25.1950 *	16.4560	20.7780	-5.9445 ***	-5.2532 ***	-5.6062 ***	0.4031	-0.9994	-0.2891	19.8350	-15.8480 **	1.3435
	[14.854]	[11.215]	[12.609]	[1.8185]	[1.9146]	[1.8148]	[0.8690]	[0.8519]	[0.4828]	[13.747]	[6.4917]	[7.5712]
Observations	*6,327*	*6,336*	*12,663*	*7,309*	*7,316*	*14,625*	*3,292*	*3,359*	*6,651*	*1,773*	*2,021*	*3,794*
R-squared	*0.81*	*0.82*	*0.81*	*0.36*	*0.36*	*0.35*	*0.29*	*0.28*	*0.17*	*0.88*	*0.88*	*0.87*

Notes: Panel A rewrites the results from Panel A of Table 3. Panel B adds other diseases as additional controls. The other diseases are the diseases denoted in the first column. Each of these diseases enters into equation (1) as the interaction between the 1929 passive detection rate and the post-eradication cohort dummy. The dependent variables are the variables denoted in the column headings. Clustered standard errors at the canton unit are in parenthesis. *** denotes statistical significance at 1 percent level, ** significance at 5 percent level, * significance at 10 percent level. All specifications include canton of birth dummies, year of birth dummies, and census year dummies.

Table 4: Additional controls for the first malaria eradication campaign, by sex.

DEPENDENT VARIABLE:	YEARS OF EDUCATION			LITERACY RATE			LN(HOURS WORKED)			LN(WAGE)		
SEX:	Women	Men	All	Women	Men	All	Women	Men	All	Women	Men	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PANEL A: Baseline Results												
(A) During eradication cohort x 1929 malaria infection rate	7.6283 **	7.6955 **	7.6987 **	0.6066	0.8002 *	0.6708	-0.0982	0.2707	0.1055	2.7271	4.1393 ***	3.0672 *
	[3.4824]	[3.2105]	[3.1909]	[0.4236]	[0.4581]	[0.4216]	[0.2147]	[0.1895]	[0.1539]	[3.0561]	[1.1101]	[1.5888]
PANEL B: additional controls												
(B) Banana productivity (GAEZ-FAO)	6.9445 **	6.6907 **	6.8203 **	0.5324	0.7319	0.6317	-0.0763	0.2736	0.1029	2.2957	4.0754 ***	3.1860 *
	[3.0206]	[2.7965]	[2.8014]	[0.4133]	[0.4515]	[0.4226]	[0.2203]	[0.1897]	[0.1345]	[3.4794]	[1.1427]	[1.8537]
(C) Health	1.2143	1.4850	1.3665	2.2716 ***	2.6609 ***	2.4657 ***	0.2213	0.6139 **	0.4238 ***	-0.4849	4.9265 **	2.3513
	[4.3524]	[3.8520]	[3.8942]	[0.7926]	[0.8096]	[0.7770]	[0.3263]	[0.2640]	[0.1175]	[5.0521]	[1.8457]	[2.4308]
(E) Education	6.5373 *	7.4744 **	7.0159 **	0.9397 *	0.9968 *	0.9688 *	-0.1478	0.1585	0.0101	1.1856	3.7985 ***	2.5573
	[3.4371]	[3.3657]	[3.2995]	[0.4939]	[0.5340]	[0.5060]	[0.2679]	[0.1486]	[0.1430]	[3.0868]	[1.3787]	[1.7092]
(G) Younger sample (between 15 and 45 years old)	5.5586 *	6.7596 **	6.1708 **	0.5438	0.8154 *	0.6795 *	-0.1116	0.3208	0.1115	2.5487	4.1393 ***	3.3446 *
	[2.9394]	[2.9761]	[2.7585]	[0.3651]	[0.4290]	[0.3732]	[0.3210]	[0.1942]	[0.1957]	[3.3452]	[1.1101]	[1.7635]
(J) Fraction of movers to total population	7.1398 **	8.3967 **	7.7757 **	0.8414 **	0.9948 **	0.9183 **	-0.0288	0.1561	0.0681	1.8623	4.0818 ***	2.9654 *
	[3.3549]	[3.4123]	[3.2981]	[0.4107]	[0.4743]	[0.4335]	[0.2511]	[0.1840]	[0.1611]	[3.1530]	[1.2084]	[1.6303]
(K) Mean *manzanas* per *finca*	7.9039 **	7.6435 **	7.7806 **	0.1532	0.2673	0.2101	-0.0417	0.2555	0.1093	4.0486	3.1334 ***	3.4410 *
	[3.6162]	[3.3447]	[3.3675]	[0.3907]	[0.4141]	[0.3888]	[0.2444]	[0.1880]	[0.1389]	[3.6883]	[1.0390]	[1.9034]
(L) Mean *manzanas* per habitants	7.6982 **	7.3807 **	7.5465 **	0.2839	0.3859	0.3347	-0.0547	0.2550	0.1032	4.0473	3.3447 ***	3.5538 *
	[3.5392]	[3.2765]	[3.3035]	[0.4028]	[0.4256]	[0.4020]	[0.2469]	[0.1939]	[0.1421]	[3.6426]	[1.0566]	[1.8886]

Notes: Each cell in this table shows estimates on the coefficient of post-eradication cohort dummy and the 1929 malaria rate at the canton of birth, from equation (1). The dependent variables are the variables denoted in the column headings. Each additional control is specified in each row of the first column. All specifications include cohort, canton and census year fixed effects. GAEZ banana potential yield is interacted with a time trend. The health controls include the yearly rate of treated patients and the number of months a health centers remained open. Education controls include the number of schools during 1927 and the number of schools during 1925. Fraction of movers to total population includes the fraction of population that is immigrant and the fraction of population that is emigrant. The last two rows are self-explanatory. All these controls enters into equation (1) interacted with the post-eradication cohort dummy. For more information on data sources and construction see Section 3. The other schools construction programs is the cantonal mean number of schools that are open six years after the person was born. Clustered standard errors at the canton unit are in parenthesis. *** denotes statistical significance at 1 percent level, ** significance at 5 percent level, * significance at 10 percent level.

Table 5: Effect of a malaria resurgence on human capital and economic attainment, by sex.

DEPENDENT VARIABLE:	YEARS OF EDUCATION			LITERACY RATE			LN(HOURS WORKED)		
SEX:	Women	Men	All	Women	Men	All	Women	Men	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				PANEL A: Baseline results, peak episode					
During peak cohort	-2.4242	-8.3022	-5.3632	-0.3471	-0.1633	-0.2551	0.3251	-0.5686 **	-0.1594
x 1956 malaria infection rate	[9.8144]	[7.4265]	[8.1335]	[0.3495]	[0.1996]	[0.2432]	[0.4751]	[0.2777]	[0.2666]
R-squared	*0.92*	*0.91*	*0.90*	*0.62*	*0.50*	*0.54*	*0.46*	*0.58*	*0.30*
Effect of a 1 s.d. increase in malaria rate	*-0.20%*	*-0.68%*	*-0.44%*	*-0.23%*	*-0.11%*	*-0.17%*	*0.21%*	*-0.36%*	*-0.10%*
				PANEL B: additional (province x cohort) fixed effects					
During peak cohort	-13.3650 *	-16.6380 **	-15.0020 **	-0.7105 *	-0.2880 *	-0.4992 *	0.5542	-0.5677 ***	-0.0020
x 1956 malaria infection rate	[7.6648]	[6.9407]	[6.5873]	[0.4052]	[0.1460]	[0.2523]	[0.3561]	[0.1925]	[0.2126]
Observations	*1,627*	*1,627*	*3,254*	*2,059*	*2,062*	*4,121*	*736*	*740*	*1,476*
R-squared	*0.90*	*0.90*	*0.89*	*0.64*	*0.51*	*0.54*	*0.52*	*0.59*	*0.35*

Notes: Each cell in this table shows estimates on the coefficient of the cohort dummy and the malaria rate at the canton of birth, from equation (3). The dependent variables are the variables denoted in the column headings, and the independent variables are the variables denoted in each row. All specifications include canton of birth dummies, year of birth dummies, and census year dummies. Clustered standard errors at the canton unit are in parenthesis. *** denotes statistical significance at 1 percent level, ** significance at 5 percent level, * significance at 10 percent level. Panel A shows baseline results, Panel B additionally adds interactions between province of birth dummies and year of birth dummies. Columns (1)-(6) uses information from censuses 1963-2011, and columns (7)-(9) uses information from 1973-1984.

Table 6: Sensitivity checks for the malaria peak episode estimates on human capital and economic attainment with additional diseases, by sex.

DEPENDENT VARIABLE:	**YEARS OF EDUCATION**			**LITERACY RATE**			**LN(HOURS WORKED)**		
SEX:	Women	Men	All	Women	Men	All	Women	Men	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					PANEL A: Baseline results, peak episode				
Peak cohort x 1956 malaria rate	-2.0848	-8.7697	-5.4273	-0.3768	-0.2583	-0.3177	-0.1673	-0.6263 **	-0.4365
	[9.8805]	[6.9988]	[7.7662]	[0.3520]	[0.1872]	[0.2271]	[0.6319]	[0.2793]	[0.3221]
					PANEL B: additional diseases controls				
Peak cohort x									
1956 malaria rate	-11.4960 ***	-12.8950 ***	-12.1960 ***	-0.4488 ***	-0.1442	-0.2962 ***	0.3002	-0.5372 ***	-0.1545
	[2.4119]	[3.0813]	[2.4802]	[0.1293]	[0.1184]	[0.1025]	[0.2674]	[0.1595]	[0.1567]
1929 influenza rate	1.7495	-3.6910	-0.9708	-0.7318 *	-0.5656	-0.6476 *	-1.9591 *	-0.1119	-1.1862 **
	[7.1564]	[12.106]	[9.1035]	[0.4110]	[0.4368]	[0.3392]	[0.9395]	[0.5190]	[0.4947]
1929 tuberculosis rate	-18.0830	45.5920	13.7540	-3.1034	6.5656 **	1.7300	-18.2400 ***	5.9262	-6.2572 **
	[73.346]	[65.514]	[65.718]	[3.7005]	[2.3767]	[2.0911]	[6.3445]	[4.1246]	[2.2107]
Peak cohort x									
1956 malaria rate	-0.3582	-7.0541	-3.7062	-0.1594	-0.1647	-0.1615	-0.1023	-0.3553	-0.2810
	[13.595]	[8.6035]	[10.790]	[0.4985]	[0.2119]	[0.3280]	[0.6640]	[0.4054]	[0.2706]
1929 hookworms rate	-10.0290	-13.6370	-11.8330	-0.6365 **	-0.7618 **	-0.6982 ***	-0.4885	0.6954	0.0126
	[8.0681]	[10.714]	[8.9542]	[0.2518]	[0.3461]	[0.1895]	[0.9273]	[0.9280]	[0.7893]

Notes: Panel A rewrites the results from Panel A of Table 3. Panel B adds other diseases as additional controls. The other diseases are the diseases denoted in the first row. Each of these diseases enters separately into equation (3) as the interaction between the 1956 active detection rate and the peak cohort dummy. The dependent variables are the variables denoted in the column headings. Clustered standard errors at the canton unit are in parenthesis. *** denotes statistical significance at 1 percent level, ** significance at 5 percent level, * significance at 10 percent level. All specifications include canton of birth dummies, year of birth dummies, and census year dummies.

Table 7: Sensitivity checks for the malaria peak episode estimates on human capital and economic attainment with additional controls, by sex

	DEPENDENT VARIABLE:	YEARS OF EDUCATION			LITERACY RATE			LN(HOURS WORKED)		
	SEX:	Women	Men	All	Women	Men	All	Women	Men	All
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
						PANEL A: Baseline results, peak episode				
(A)	During eradication cohort x	-2.0848	-8.7697	-5.4273	-0.3768	-0.2583	-0.3177	-0.1673	-0.6263 **	-0.4365
	1956 malaria infection rate	[9.8805]	[6.9988]	[7.7662]	[0.3520]	[0.1872]	[0.2271]	[0.6319]	[0.2793]	[0.3221]
						PANEL B: additional controls				
(B)	Banana productivity	-8.6314	-14.2170 *	-11.4240	-0.4126	-0.2228	-0.3177	0.3227	-0.5908 *	-0.1311
	(GAEZ-FAO)	[7.7607]	[7.3692]	[6.8941]	[0.3584]	[0.1433]	[0.2318]	[0.3680]	[0.3096]	[0.2224]
(C)	Health	-9.4191	-18.2090 **	-13.8140 **	-0.4642	-0.4401 **	-0.4517 **	0.0332	-0.5492	-0.3078
		[7.3878]	[7.1727]	[6.0691]	[0.4067]	[0.1717]	[0.2182]	[0.6241]	[0.3442]	[0.3158]
(E)	Education	-0.4288	-5.2678	-2.8483	-0.2320	-0.0224	-0.1271	-0.2314	-0.6843 *	-0.5051
		[13.016]	[10.176]	[11.431]	[0.4890]	[0.2891]	[0.3744]	[0.6105]	[0.3742]	[0.3167]
(G)	Younger sample (between	-2.6380	-14.1560	-8.3972	-0.4085	-0.1456	-0.2761	-0.1514	-0.6059 *	-0.4212
	15 and 45 years old)	[11.999]	[8.8258]	[9.7891]	[0.3650]	[0.2581]	[0.2882]	[0.6442]	[0.3293]	[0.3210]
(J)	Fraction of movers to total	-5.1818	-9.7512	-7.4665	-0.5864 *	-0.1253	-0.3558	-0.1414	-0.8238 **	-0.4995
	population	[10.646]	[7.6175]	[8.7381]	[0.3342]	[0.1790]	[0.2465]	[0.6936]	[0.3999]	[0.4230]
(K)	Mean *manzanas* per *finca*	-3.8030	-12.0170	-7.9102	-0.3791	-0.2252	-0.3018	-0.1965	-0.5883 *	-0.4356
		[10.838]	[8.4604]	[8.9666]	[0.3586]	[0.2117]	[0.2458]	[0.6158]	[0.3305]	[0.3118]
(L)	Mean *manzanas* per habitants	-3.7125	-11.9260	-7.8190	-0.3762	-0.2173	-0.2964	-0.2234	-0.5772 *	-0.4443
		[10.877]	[8.5516]	[9.0278]	[0.3585]	[0.2172]	[0.2480]	[0.6149]	[0.3344]	[0.3105]

Notes: Each cell in this table shows estimates on the coefficient of peak-cohort dummy and the 1963 malaria rate at the canton of birth, from equation (3). The dependent variables are the variables denoted in the column headings. Each additional control is specified in each row of the first column. All specifications include cohort, canton and census year fixed effects. Included controls are the same as those in Table 4, see the Notes. Clustered standard errors at the canton unit are in parenthesis. *** denotes statistical significance at 1 percent level, ** significance at 5 percent level, * significance at 10 percent level.

Table 8: Results on the dependence of the marginal benefit of the first eradication campaign and the cantonal concerns with educational attainment and development of the child labor market, by sex.

DEPENDENT VARIABLE:	**YEARS OF EDUCATION**			**LN(WAGE)**		
SEX:	Women	Men	All	Women	Men	All
	(1)	(2)	(3)	(4)	(5)	(6)
				PANEL A: schooling motivation		
During eradication cohort	0.8741	1.2179	1.0499	1.0932	3.1889 **	2.2130
x 1929 malaria infection rate	[4.2677]	[3.6899]	[3.8515]	[3.6022]	[1.3879]	[1.8681]
During eradication cohort						
x 1929 malaria infection rate	0.0384 **	0.0427 *	0.0405 *	0.0064 *	0.0020	0.0040
x students per schools during 1927	[0.01862]	[0.02296]	[0.02062]	[0.003593]	[0.006218]	[0.004048]
Observations	*6,077*	*6,089*	*12,166*	*2,041*	*2,318*	*4,359*
R-squared	*0.83*	*0.84*	*0.83*	*0.88*	*0.88*	*0.88*
				PANEL B: child labor market		
During eradication cohort	35.0600 ***	3.6448	9.7625	4.9304	14.2190 ***	13.7100 **
x 1929 malaria infection rate	[12.700]	[14.076]	[8.7854]	[16.599]	[5.3006]	[6.6929]
During eradication cohort						
x 1929 malaria infection rate	-62.4370 **	6.4299	-7.5013	-5.1391	-22.8590 *	-22.9310
x fraction of children employed during 1927	[25.351]	[27.534]	[14.871]	[35.442]	[11.587]	[14.200]
Observations	*6,331*	*6,342*	*12,673*	*2,126*	*2,416*	*4,542*
R-squared	*0.83*	*0.83*	*0.83*	*0.88*	*0.88*	*0.88*

Notes: Panel A shows estimates of $\hat{\beta}$ the during eradication dummy and a de-mean malaria rate in 1929 coefficient, and of the interaction coefficient $\hat{\theta}$ between the number of students per school during 1927 and the during eradication dummy and the 1929 malaria rate at the canton of birth, from equation (4). Panel B shows the same coefficient but using instead the fraction of children employed during 1927. The dependent variables are the variables denoted in the column headings, and the independent variables are the variables denoted in each row. All specifications include canton of birth, year of birth, and census year fixed effects. Clustered standard errors at the canton unit are in parenthesis. *** denotes statistical significance at 1 percent level, ** significance at 5 percent level, * significance at 10 percent level. Panel A shows results for the concern with educational attainment, and Panel B shows results for the development of the child labor market. Columns (1)-(3) uses information from censuses 1963-2011, and columns (4)-(6) from 1963-1973.

Table 9: Results on the dependence of the marginal cost of the malaria resurgence and the cantonal concerns with educational attainment and development of the child labor market, by sex.

DEPENDENT VARIABLE:	YEARS OF EDUCATION			LN(HOURS WORKED)		
SEX:	Women	Men	All	Women	Men	All
	(1)	(2)	(3)	(7)	(8)	(9)
			PANEL A: schooling motivation			
During eradication cohort						
x 1956 malaria infection rate	9.5947	6.1853	7.8900	1.9766	-0.0068	1.1592 *
During eradication cohort	[13.762]	[13.464]	[12.183]	[1.4403]	[0.8647]	[0.6789]
x 1956 malaria infection rate	-0.0759	-0.0950	-0.0854	-0.0106	-0.0031	-0.0079 ***
x students per schools during 1963	[0.0518]	[0.0709]	[0.0531]	[0.0070]	[0.0041]	[0.0028]
Observations	*1,627*	*1,627*	*3,254*	*736*	*740*	*1,476*
R-squared	*0.899*	*0.89*	*0.883*	*0.462*	*0.572*	*0.305*
			PANEL B: child labor market			
During eradication cohort						
x 1956 malaria infection rate	-9.2176	-83.0880 *	-13.7870 *	-2.4848	-3.1686	-0.6898
During eradication cohort	[22.338]	[40.815]	[6.9181]	[1.9690]	[2.3745]	[1.0162]
x 1956 malaria infection rate	25.3980	18.5330 *	2.9399	116.0300	6.8924	1.3092
x fraction of children employed during 1963	[84.953]	[10.262]	[2.0207]	[92.752]	[6.3757]	[4.3450]
Observations	*1,627*	*1,627*	*3,254*	*736*	*740*	*1,476*
R-squared	*0.898*	*0.89*	*0.883*	*0.461*	*0.572*	*0.304*

Notes: Panel A shows estimates of $\hat{\beta}$, the during eradication dummy and a de-mean malaria rate in 1956 coefficient, and of the interaction coefficient $\hat{\theta}$ between the number of students per school during 1956 and the during eradication dummy and the 1956 malaria rate at the canton of birth, from equation (4). Panel B shows the same coefficient but using instead the fraction of children employed during 1956. The dependent variables are the variables denoted in the column headings, and the independent variables are the variables denoted in each row. All specifications include canton of birth, year of birth, and census year fixed effects. Clustered standard errors at the canton unit are in parenthesis. *** denotes statistical significance at 1 percent level, ** significance at 5 percent level, * significance at 10 percent level. Panel A shows results for the concern with educational attainment, and Panel B shows results for the development of the child labor market. Columns (1)-(3) uses information from censuses 1963-2011, and columns (4)-(6) from 1963-1973.

Chapter 2

Who Matters? Peer Effects in the Adoption of a New Employment Subsidy for Vulnerable Youths

Abstract

Take-up of many social programs is very low. This paper studies if peer interactions can increase this in the case of the Youth Employment Subsidy in Chile. It estimates the causal impact of two types of peer adoption on one's own adoption, focusing on high-school classmates and co-workers. The subsidy is analyzed since the beginning of its implementation in July 2009, using a unique database that combines four different administrative records. Identification comes from the discreteness in the eligibility rule, where only workers with a continuous vulnerability score below a cut-off point were eligible. I find that high-school classmates strongly influence one's adoption of the subsidy while coworkers do not. Peer effects are more important among younger youths that graduated from a *técnico profesional* high school, and within small high-schools and firms. However, these peer effects disappear six months after. Simulations on the econometric model show that, if the government wants to increase SEJ take-up, it should target small schools with advertisement campaigns.

2.1 Introduction

Many social programs around the world have low take-up rates. For example, in the United States, take-up varies a great deal across means and non-means tested programs (Currie, 2004). The State Children's Health Insurance Program (SCHIP) has a take up rate around 8.1% and 14% (LoSasso and Buchmueller, 2002), the Child Care Subsidy Programs has a take up of 15% (Administration for Children and Families, 1999), while the Unemployment Insurance had a take-up rate of 83% between 1980 and 1982, and the Medicade had a take-up rate of 96% during 2002[1]. While some

[1]Something similar happens in the United Kingdom, where the Working Families' Tax Credit of the United Kingdom has a take up rate of 72% (Currie, 2004), and the Income Support has a take up of 64% (Cuclos, 1995). In Norway, Dahl,

have argued that high costs of learning, application and stigma, low benefits from applying, and behavioral issues on the profile of the payments can explain the different take-up rates, I study a less common explanation which is the impact of peers, potentially through information sharing. I do this in the context of a youth employment subsidy in Chile, whose take up rate has remained below 20% (similar to the take up calculated by Bravo and Rau (2014)).

The Youth Employment Subsidy (SEJ) was implemented in July 2009. It consists in a monetary incentive for employed and self-employed youths and their employers. During the application process, all information is verified internally and automatically without the need of paperwork, ensuring that the application cost is near zero. Moreover, if accepted, payments begin at most 90 days after the application is submitted. Workers can decide how to receive SEJ payments, monthly or annually, and via a private bank deposit or cash. For some workers, depending on the monthly gross income, an annual payment of SEJ could represent more than an extra monthly wage each year and corresponds to a subsidy of up to 23% of the monthly wage. Altogether, this social program has an extremely simple and private application process, that must be carried out on the internet at any time and place. This further decreases the application cost, by providing an easy and accessible way to apply[2]. However, despite of advertising campaigns at the beginning of the program implementation, and the low application costs, the percent of admissible workers that are taking up SEJ has stayed very low, beginning at less than 4% and never growing beyond 20% initially as shown in Figure 2.1.

What could explain the low take up rate of the SEJ welfare program? This work studies whether an alternative explanation lies in the lack of information and how this could be solved through peers. The hypothesis is that during the first months of SEJ implementation, informational barriers made the experience of peers very useful. Given the particular case of SEJ, this information would translate into adoption, so peer effects could be importantly high at the beginning. At the beginning there were many isolated groups with no early SEJ adopters, and this would explain the low take up and its slow growth.

This work also estimates the causal impact of different types of peers' adoption decision on one's own decision. It focuses on coworkers and former classmates given that in Chile schools have been shown to be highly relevant (e.g. Zimmerman, 2015) in determining labor market outcomes. But before drawing conclusions regarding the impact of peer's adoption, first the endogeneity problem of one's peers decision must be solved. Specifically, an OLS could be confounding the effect of correlated unobservables at the group level.

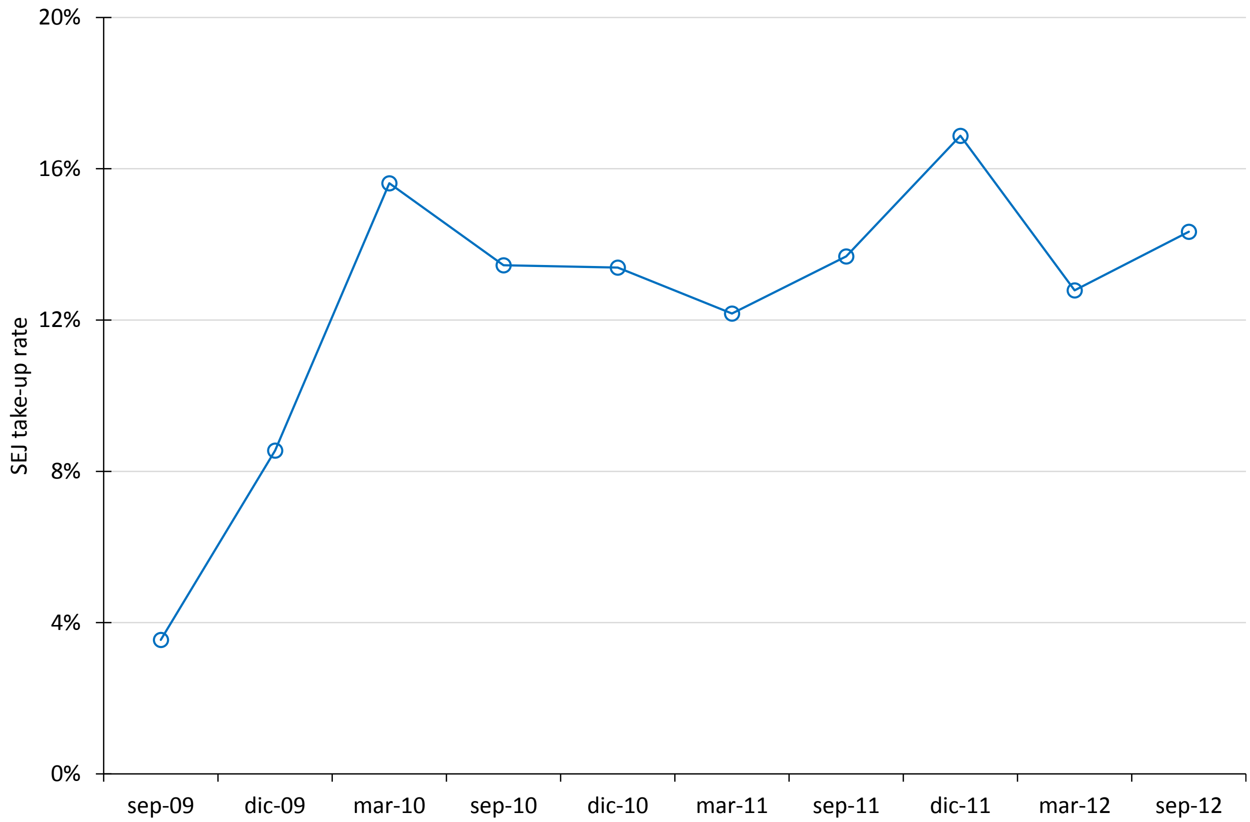
The identification strategy employed in this work exploits a very sharp eligibility rule that SEJ followed, in order to identify causal peer effects. In particular, in order to be eligible, a worker's vulnerability score known as *Ficha de Protección Social* (FPS) must be equal to or below 11,734 points[3].

Loken and Mogstad (2014) calculate that after a reform made to promote gender equality, the take up of governmental paid paternity leave by eligible fathers jumped from 3% prior to the reform to approximately 35% in 1993 after the reform was implemented.

[2]According to SUBTEL (2014), in 2009, 31.1% of Chilean households had any kind of access to the internet from the house; during 2014, this amount has increased to 61.6%. According to CASEN 2011, almost 80% of youth Chileans between 18 and 25 years old had any access to the internet, and the average among the lower quintiles of income was around 65%.

[3]This FPS is private information of the worker and is not necessarily known by the employer or other coworkers.

Figure 2.1: SEJ take up rate



Notes: This graph shows the take up rate of the Youth Employment Subsidy (SEJ) by date. Take up is defined as the number of people which received any SEJ payment (monthly or annual) at the specified date divided by the number of people admissible to apply to SEJ at the same date.

This rule induced quasi-experimental variation that comes close to an ideal experiment. I then use as an IV the fraction of the peers at the left of a very small window around the cut off. The idea is that admissibility to SEJ was quasi-randomly distributed inside this window. I show that networks with a higher fraction of peers at the left of this small window were more likely to have adopted SEJ. This instrumental variable is used to identify a causal peer effect. Results show a strong first stage, and also documents that no other observables changed near this cut-off, except the fraction of peers that are taking up SEJ. I also demonstrate that the IV is orthogonal to several variables and quasi-randomly distributed among the population, so it is very likely to also be independent of common unobservable shocks at the group level (correlated unobservables).

To measure this network, this work exploits a unique database that comes from merging together four different administrative datasets using a unique identification number. Together, they provide information on individual socioeconomic characteristics for all individuals between 18 and 25 years old and with a vulnerability score, employment history for employed workers in private firms, history of school enrollment by grade and year between 2002 and 2012, and a record of the monthly and annual SEJ payments, allowing me to focus on a very small window around the cut-off. In order to avoid problems from endogenous group membership, classmates and coworkers networks are defined in a point in time before the introduction of SEJ, tracking peer effects among peers who were within one's network before SEJ was implemented. Hence, any changes in group formation are exogenous to the introduction of SEJ.

First-stage results show that, at the beginning of SEJ implementation during September 2009, having a 10% larger fraction of classmates at the left of the window increased the fraction of classmates with SEJ by 0.19%, and having a 10% larger fraction of coworkers at the left of the window increased the fraction of coworkers with SEJ by 0.20% (both significant at the 1%). The reduced form estimates show that having a 10% larger fraction of classmates at the left of the window increased the individual probability of adopting SEJ by 0.17% (significant at the 5% level), and having a 10% larger fraction of coworkers at the left of the window decreased the individual probability of adopting SEJ by -0.05% but statistically indistinguishable from zero. It is thoroughly argued that the only channel through which the instrument affected individual adoption was through adoption by the peers, so almost all the adoption by the classmates was translated into a proportional increase of the individual probability of adopting. Hence, a 10% higher fraction of classmates with SEJ implies a 9.1% higher individual probability of adopting SEJ (significant at the 5%); while a 10% larger fraction of coworkers with SEJ implies a reduction of -25.2% in the individual probability of adopting SEJ but statistically equal to zero. I conclude that the classmates played a vital role in determining participation in SEJ program during the early months, but the coworkers were not as relevant.

Other works have studied the importance of classmates for privileged persons in Chile. According to Zimmerman (2015), admission to an elite university program raises the number of leadership positions students hold by 50%, but gains are larger for students who attended one of nine elite private high schools and near zero for students who did not. In the context of this work, classmates are not only important for privileged people but also for more vulnerable students by providing valuable and important information (perhaps about the existence, costs, benefits, etc.) of newly social programs.

In analyzing the mechanisms that drive the peer effects, a second set of results shows that an informational channel is one of several good candidates to explain the results. Peer effects are more important among youths under 21 years old, because these are youths that recently graduated from high school and the interaction with other classmates is likely to be higher compared to older youths. Peer effects are also higher among graduates from *técnico profesional* high schools, who are expected to be the most benefited from this information. Finally, peer effects are higher among youths within smaller networks of classmates and coworkers, where it is easier to realize if someone has adopted SEJ. And peer effects are smaller among among youths within larger networks of classmates and coworkers, but the first stage here is weaker.

Third, the peer effect of the classmates network was very important at the beginning, but became irrelevant six months after. This has many possible interpretations. First, as a trendy talking phenomenon, where youths began talking about a "viral" topic at the same time that advertising campaigns where undergoing, but whose importance decayed over time. Second, during the early stages information is very scarce about the existence of SEJ, and other details of its application procedure and costs and benefits, but afterwards many people learn about it and peers become less relevant. Third, peers shift in importance and get depreciated over time, and since the classmates definition is fixed here, they can become less important as people get older. Fourth, Bravo and Rau (2013) finds that admissibility to SEJ had an impact over the economic cycle of employment but this effect decays over time, as a result adopting SEJ several months after its implementation because someone told me so becomes less relevant. Finally, this can also suggest that peer effects can depend on external characteristics as, for example, advertising campaigns that triggers them. Altogether, classmates help to solve a lack of information, but once the information is available, they lose importance.

Finally, under the experiment of changing the initial distribution of the people with SEJ, and assigning them within groups that meet certain criteria, the econometric model predicts that advertisement campaigns targeted randomly among the universe of workers achieves a take up rate lower than the take up rate under the real initial distribution. Targeting small and big firms implies an initially higher take up, but there is no gain over time. Targeting small schools allows the take up rate to increase over time and reach a steady state take up rate that is higher than the initial take up rate. This initial assignment of people with SEJ can be interpreted in many ways, but an interesting one is to see it as a targeted advertising campaign where the government targets people within groups that meet certain criteria and successfully increases take up among them, with the purpose of influencing the decision of their peers.

This work makes several contributions to the current state of knowledge. First, this is the first work to study a welfare program during the early stages of its implementation, two months after it was implemented. Many have examined the take up of social programs and have provided explanations for the low take up rates, but using well established programs. According to Currie (2004), a basic cost/benefit framework has remained the basis for empirical investigations of non-participation in social programs[4]. Individuals eligible for means-tested programs face a "stigma" cost (Moffit, 1983); and costs of learning about, and applying for the programs. These costs may be sufficient to deter some individuals from using them. However, SEJ payments are private and the application cost is near zero, and so this is not a likely explanation. From a behavioral economics perspective (O'Donoghue and Rabin, 1999), a time inconsistent behavior might put off a person from enrolling in a program because many of the costs of enrolling in social programs are borne immediately, whereas the benefits are in the future. But payments from SEJ begin at most 90 days after the application was sent, and this does not seem a likely explanation either.

Though there are many potential explanations for the low take-up rates, none of this literature has examined a social program since the beginning of its implementation. It is during this phase that peer effects can provide a better explanation. According to Bertrand, Luttmer and Mullainathan (2000), social networks are important due to an information channel, where a person's knowledge depends on the behavior of others. This channel operates mainly by reducing the cost of applying for welfare or increasing its benefits. And due to a social norm channel, where a person's preferences may depend on the behavior of others.

Recently, a small amount of works have empirically confirmed the importance of social networks in determining participation in already undergoing social programs, but with different estimates.

[4]See, for example, the works from Anderson and Meyer (1997); Riphahan (2001); Kleven and Kopczuk (2011)

Duflo and Saez (2001) find that randomly giving information to individuals in a group increases the probability of enrollment in a specific retirement plan by that individual and the other control members of the same group, but in a small amount. Dahl, Loken and Mogstad (2014) finds that coworkers and brothers are 11 and 15 percentage points, respectively, more likely to take paternity leave if their peer father was induced to take up leave by a Norwegian reform. The most likely mechanism is information transmission about costs and benefits, including increased knowledge of how an employer will react, because stigma costs in this paid paternity leave program where unknown. Finally, Aizer and Currie (2004) argue that an informational channel cannot account for the peer effects that they find in the take up of a publicly-funded prenatal care program. Instead they could be due to a different way in which the relevant institutions treat members of different groups; or alternatively, how members of different groups respond to similar treatments. But, once again, these papers fail to study social programs in the implementation phase and peer effects during this phase could behave differently due to information sharing.

In Chile, Carneiro, Galassoy, and Ginja (2014) showed that information sharing is important in defining social program participation. They find, among other things, that *Chile Solidario* participants, a program that provides information on social programs, increased their take-up of a family allowance for poor children (the *Subsidio Unico Familiar*, SUF) by 11%, relative to an average take-up of 65% among comparable non-participants. But Carneiro, Galassoy, and Ginja (2014) do not analyze how peer effects influence program participation.

Second, the present paper has the advantage of using a database with information on several kinds of networks, but focuses on classmates and coworkers, so this allows to compare the importance of the peer effects between them. The literature above is limited on the information from the peers, and as a result it has not studied the effect of take up by former classmates. Dahl, Loken and Mogstad (2014) study coworkers and family members, Aizer and Currie (2004) study neighbors at the zip code level, and Duflo and Saez (2001) study university departments. In the present paper I have enough information to show that the relevant network is the classmates and not the schoolmates.

Third, the present paper is one of the first to track the evolution of peer effects for several months afterwards, until September 2011. The above literature has documented peer effects in a point in time.

Finally, this paper exploits the network structure and the econometric model to simulate advertisement campaigns that target people within specific groups with the objective of increasing the take up rate of SEJ. While this approach has not been exploited in the literature of take-up rates, it is similar to the one used by Carrell, Sacerdote and West (2013) but in a very different context: academic performance and peer groups of entering freshmen at the United States Air Force Academy.

The present work provides empirical evidence on a theoretically informational channel described elsewhere. When payoffs from different actions are unknown, agents use their own past experience as well as the experience of their neighbours to guide their decision making (Bala and Goyal, 1998). Others study the way that word-of-mouth communication aggregates the information of individual agents (Ellison and Fundenberg, 1995). Agents learn in a bayesian (Acemoglu, Dahleh, Lobel and Ozdaglar, 2011) or non-bayesian way (Jadbabaie et al., 2012). Being homophily (Golub and Jackson, 2012) and the network structure, especially what is known as a "key player"

(Ballester, Calvó-Armengol and Zenou, 2006), among the factors that determine social learning.

Other literature have studied referral based models where coworkers provide information about available jobs and unknown productivity of workers (Glitz, 2013). But it have not made many attempts at studying what kind of other information coworkers provide. Dahl, Loken and Mogstad (2014) find important peer effects among coworkers in social program participation.

The paper is organized as follows. Section 2.2 describes SEJ program in detail. Section 2.3 explains the identification strategy formally, describes the empirical strategy and explains how are the variables empirically measured and defined. Section 2.4 describes the data used by this work. Section 2.5 analyzes the main results for September 2009, provides robustness checks, and explains the possible mechanisms that explains the peer effects. It also provides results for other dates. Section 2.6 describes the targeted advertising simulation. Finally, section 2.7 presents the main conclusions.

2.2 SEJ subsidy

This section describes the SEJ program, its admissibility criteria, the postulation process and the payments scheme. Because data at hand allows to study only dependent workers, this section focuses on them, and less so in employers and independent workers.

The Youth Employment Subsidy (SEJ) is an initiative oriented to employed and self-employed youths and their employers with independent application processes, improving their wages and supporting those who hire them, by providing a monetary contribution. SEJ was proposed by the advisory council of work and equity in May 2008, but it did not become a bill until April 2009 and implemented in July 2009[5]. The execution and administration of SEJ is in charge of the Service of Training and Employment (SENCE) from the Ministry of Labor and Social Welfare, and additionally involves the Social Security Institute (IPS) and the Internal Revenue Service.

According to Huneeus (2010), SEJ advertising was supplemented by field visits of the minister and the entire team of the Ministry of Labor and Social Welfare, some of them also had the participation of the Finance Minister. Advertising was placed in television, national and regional radio, national and regional newspapers, subway, public transportation in Santiago and their bus stops.

Admissility Criteria. To be considered as admissible to apply to SEJ, a person must meet the following criteria:

1. be between 18 and less than 25 years old;

2. be part of a family group belonging to the poorest 4th decile of the population, which is equivalent to having a vulnerability score equal or less than 11,734 points in the FPS;

3. certify an annual gross income less than CLP$4,320,000[6] in the calendar year in which the

[5] Figure B.1 in appendix B.1 shows line trends for the traffic of Google searches of words related to SEJ by date. It shows that related searches began in July 2009, most of the using the correct keywords “subsidio empleo joven”. This lessens worries about anticipation to the politic.

[6] According to the law, all quantities in Chilean pesos are adjusted each year on the 100% annual variation experimented by the Chilean CPI.

benefit is requested (or a monthly gross income less than CLP$360,000 for those who request provisional monthly payments);

4. not be working in a state institution or a company with a state contribution higher than 50%; and

5. have social security contributions paid up to date.

The benefit is maintained as long as the recipient continues to meet the above criteria. Beginning in April 2011, a worker who is 21 years or older must have obtained a high school diploma to access or keep getting the subsidy.

Application. The application process is extremely simple and can be carried out at any time from any place because applications are made electronically via the internet. Workers and employers who want to apply must follow independent processes that require independent information. They must log into the website `www.subsidioempleojoven.cl`. If they enter their ID number, the person is immediately reported whether or not the requirements are met. If the person clicks on the application, the system displays a "registration form" which must be completed and submitted electronically. Once the form is successfully submitted, SENCE has a limit of up to 90 days to respond on the result of the application. The application status can be verified at any time by accessing the website of SEJ. If the person believes the system is displaying incorrect information, the person can contact SENCE fromt he same webpage.

In the event that the website refuses the application option on the grounds of an FPS higher than the cut-off score, the worker can approach his municipality and ask for her FPS to be recalculated. However, this does not guarantee that the FPS score will be updated.

Paperwork. This application process means that paperwork is reduced to the minimum. Moreover, several procedures are simplified to the minimum, for example, it is not necessary to submit documents establishing the employment relationship of the worker with his employer because the information is verified internally. Nor is it necessary to submit documentation in case of medical license because that information is internally verified by the Social Security Institute (IPS).

Paperwork is necessary in very specific cases, but in all of them the required documentation may be electronically uploaded to the website. For example, when the worker wants to appeal for suspended payments, or when there is a suspension of payment of the benefit due to unpaid health contributions. Appeals due to suspended payments can be entered electronically from the SEJ website.

Payment. If the application is successful, the grants will begin to be paid within a maximum period of 90 days after filing. The benefit will be accrued from the first day of the month following the date of the submission of the application.

Payment to the employee will be, as a rule, annual, being carried out during the second half of the next calendar year on which wages and employment income were earned. The dependent worker may opt for monthly payments upon submitting the application, and is able to change his choice only once during the year. In both cases, the payments are made by the medium indicated by the worker at the time of completing the application, either in cash or a bank deposit account.

The kind of payment may be changed from the SEJ website, from bank account deposit to cash payments and vice versa.

For some workers, SEJ means receiving more than an extra monthly wage. Table 2.1 and Figure 2.2 shows the amount of the subsidy to which the worker is entitled to. The amount of the annual subsidy depends on three tracts: first, for workers whose annual gross income is equal to or less than \$1,920,000, the subsidy amounts to 20% of the sum of wages and taxable income. Second, for workers whose annual gross income is greater than \$1,920,000, but not exceeding \$2,400,000, the benefit will amount to \$384,000 (20% of \$1,920,000). And, third, for workers whose annual gross income is greater than \$2,400,000 but less than \$4,320,000, the benefit will amount to \$384,000 less the 20% of the difference between the sum of wages and annual taxable income and \$2,400,000.

Table 2.1: Payment scheme for workers, by annual gross income and monthly gross income.

Annual Gross Income (AI)	Subsidy (SEJ)
$AI \leq 1,920,000$	$20\% \cdot (AI)$
$1,920,000 < AI \leq 2,400,000$	$384,000$
$2,400,000 < AI \leq 4,320,222$	$384,000 - 20\% \cdot (AI - 2,400,000)$
Monthly Gross Income (MI)	Subsidy (SEJ)
$MI \leq 160,000$	$20\% \cdot (MI)$
$160,000 < MI \leq 200,000$	$32,000$
$200,000 < MI \leq 360,000$	$32,000 - 20\% \cdot (MI - 200,000)$

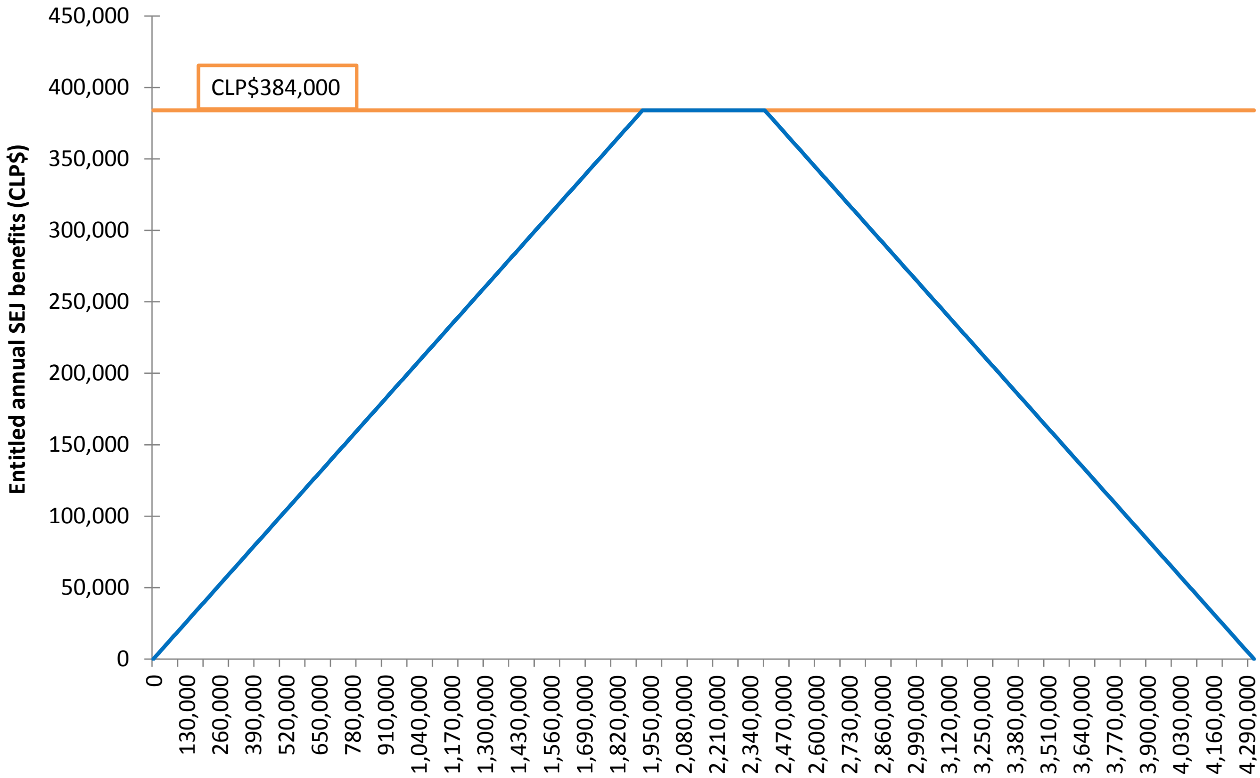
If the worker opted for monthly payments, then the amount of the subsidy will be almost similar. First, for workers with a monthly gross income less than or equal to \$160,000 the subsidy is 30% of taxable monthly remuneration. Second, if the income exceeds \$160,000 but does not exceed \$200,000, the subsidy is \$32,000 (20% of \$160,000). Third, incomes in excess of \$200,000 but less than \$360,000, the grant amount to \$32,000 less 20% of the difference between gross monthly salary and \$200,000. Monthly payments also imply retention of 75% of what the workers deserves as monthly allowance, the difference is paid when he receives the annual allowance. Figure 2.2 shows the entitled SEJ benefits for dependent workers according to his annual gross income and when the monthly gross income is annualized assuming uniform payments over the twelve months (annualized monthly income).

End of Payments. The worker ceases to be entitled to the subsidy for the months that the employer failed to pay or payed late social security contributions for the worker, or if the worker turns 21 years old and at that date has not received a high school diploma. Another reason is failure to fulfill at least one of the eligibility criteria.

Two specific groups of workers have the right to request additional time to access SEJ. First, workers who have completed regular studies in a higher education institution and are between 18 and before 25 years old. And second, women can request electronically an extension for each children born alive within the 3 months prior to the fulfillment of 25 years of age.

The workers whose employment relationship ends while they were receiving the subsidy do not need to submit a new application at the time their new employment relationship begins. To the extent that they meet the requirements, they will keep the benefit and it will be paid as soon as the

Figure 2.2: Annual SEJ benefits entitled for dependent workers according to his annual gross income in Chilean pesos (CLP$).



pension contributions payments by the new employer are verified.

2.3 Identification Strategy

2.3.1 The Problem of Interest and a Solution

The problem of interest lies in estimating an endogenous peer effect (Manski, 1993), in the sense that take up of individual i varies with the mean of the take up among those persons in the reference group g,

$$Y_{ig} = \alpha + \beta \overline{Y}_{-ig} + e_{ig} \tag{2.1}$$

where $Y_{ig} \in \{0,1\}$ is the take up decision of individual i who belongs to group g, $\overline{Y}_{-ig} \in [0,1]$ is the leave out fraction of peers with SEJ, $\overline{Y}_{-ig} = \frac{\sum_{j \in N(-i)} Y_j}{n(-i)}$, $N(-i)$ is the set of i's peers, and $n(-i) \equiv \#N(-i)$ is the number of peers leaving i outside this group. However, equation (2.1) has three problems as cited in Manski (1993) and Moffit (2001). First, there are correlated unobservables at the group level, basically because $e_{ig} = u_g + \varepsilon_{ig}$, which are unobserved shocks at the group level. Second, there are endogenous group membership and, for example, given the introduction of SEJ, workers could have moved to other employments or high schools were adoption of social programs was bigger or smaller. And finally, there is a reflection or simultaneity problem in which the decision of person i affects the group g, but the group's decisions also affects individual i, as a result the above equation is a system of equations and both decisions are made simultaneously.

This section describes the strategies used to solve these concerns, and explains the empirical strategy used to quantify a causal peer effect. To simplify the problem and explain the identification strategy, assume each network is composed of only two persons. Denote by $Y_i(d, z)$ the potential take-up decision of individual i. Treatment status $D_{-i} = d$ is defined as the take-up decision of i's peer (who I call $-i$), hence $d \in \{0, 1\}$. The instrument value $Z_{-i} = z$ is defined as being admissible to apply to SEJ or not, so $z \in \{0, 1\}$. Hence, $Y_i(d, z)$ tells me what the take-up decision of i would be given alternative combinations of peer take-up and peer eligibility. This work is interested in identifying the causal effect of peer adoption of SEJ over the take-up decision of individual i, given $-i$'s realized admissibility status, $Y_i(1, Z_{-i}) - Y_i(0, Z_{-i})$. What would i do if his peer adopts or not? The observed outcome, Y_i, can be written in terms of potential outcomes as $Y_i = Y_i(0, Z_{-i}) + [Y_i(1, Z_{-i}) - Y_i(0, Z_{-i})]D_{-i}$. Random coefficient notation for this is $Y_i = \alpha_0 + \rho_i D_{-i} + \eta_i$ with $\alpha_0 \equiv E[Y_i(0, Z_{-i})]$ and $\rho_i \equiv Y_i(1, Z_{-i}) - Y_i(0, Z_{-i})$.

However, the D_{-i} is subject to selection bias and correlated unobservables. A randomization over the peer adoption decision $D_{-i} = d$ would allow to reduce the specification or selection bias concerns. This randomization is denoted by the instrumental variable $Z_{-i} = z$. If the randomization is well done, it ensures that differences in the adoption of SEJ are not systematic.

The construction of the instrument Z_{-i} deserves a detailed explanation. As explained in Section 2.2, admissible people belong to the 4th quintile of the population or lower, which means having a vulnerability score of *FPS* equal or less than 11,734 points. In practice, I restrict myself to high schools and firms with at least one person inside a neighborhood of a very small size Δ around this vulnerability score cut off, $[x_0 - \Delta, x_0 + \Delta], x_0 = 11,734$. For people inside this small window, admissibility was as good as randomly distributed[7]. The instrument value is defined as non-missing only for the individuals whose peer vulnerability score $FPS_{-i} \in [x_0 - \Delta, x_0 + \Delta]$, otherwise the instrument's value will be missing. Then the instrument is defined as a dummy variable that indicates admissibility in terms of i's peer vulnerability score $Z_{-i} = \mathbf{1}\{FPS_{-i} \in [x_0 - \Delta, x_0]\}$. This is not the same as the compliant subpopulation, the compliers are those peers whose vulnerability score is within the interval and adopt when they are admissible. Hence, the causal effect of *FPS* eligibility status given $-i$'s take-up rate is $Y_i(D_{-i}, 1) - Y_i(D_{-i}, 0)$.

As is well known (Imbens and Angrist, 1994), under a set of assumptions, the Wald estimand can be interpreted as a local average treatment effect (LATE), the average causal effect of peer adoption on those whose treatment status can be changed by the instrument (compliers). The reduced-form

[7]This is almost a fuzzy regression discontinuity, because the decision of i's peer $(-i)$ to adopt SEJ or not is not a deterministic function of FPS_{-i} alone; but also of other variables unobserved by the econometrician and that determine assignment to treatment. This means that some admissible peers $Z_i = 1$ still do not adopt SEJ $D_i = 0$. What the empirical strategy exploits is the jump in the probability of treatment at the $x_o = 11,734$ cut-off. However, the control $f(FPS_{-i})$ is not obvious since I only have one observation of FPS for each peer. This function can be interpreted as a potential adoption function $f_{-i}(FPS)$ which tells what i's peer would do for any value of his vulnerability score of FPS. A similar problem has been analyzed previously by Dahl, Loken and Mogstad (2012). Their solution was to look at networks where there is a single peer father in the reform window. A different solution is used in this work, which is to focus on a small neighborhood around x_0.

relation $E[Y_i|Z_{-i} = 1] - E[Y_i|Z_{-i} = 0]$ measures the average causal effect of peer's admissibility over i's adoption decision. And the first-stage relation $E[D_{-i}|Z_{-i} = 1] - E[D_{-i}|Z_{-i} = 0]$ measures the average causal effect of peer's admissibility over the peer's adoption.

The structural and first-stage equations are respectively,

$$Y_i = \alpha_1 + \rho D_{-i} + \eta_i \tag{2.2}$$

$$D_{-i} = \alpha_2 + \pi_1 Z_{-i} + \xi_{-i1} \tag{2.3}$$

And the reduced form can be obtained by plugging-in (2.3) into (2.2):

$$Y_i = \alpha_3 + \beta Z_{-i} + \xi_{2i} \tag{2.4}$$

To lift up the assumption of one peer only, assume that each individual i has a specific peer's group $N(-i)$ of size $n(-i)$. This reference group or network contains individuals whose adoption may affect i's adoption decision, and vice versa. Unless otherwise stated, I assume that i is excluded from his network, so $i \notin N(-i)$. This is equivalent to taking leave-out means, and it corresponds to the usual empirical formulation when there is more information than in a survey data (*e.g.* Sacerdote (2001), Soetevent and Kooreman (2007), and Bramoullé *et al.* (2009)). An individual is isolated if his network is empty. The next step is to sum equations (2.2), (2.3) and (2.4) among the members of the network and divide by $n(-i)$:

$$Y_{ig} = \alpha_1 + \rho \overline{Y}_{-ig} + \eta_{ig} \tag{2.5}$$

$$\overline{Y}_{-ig} = \alpha_2 + \pi \overline{Z}_{-ig} + \overline{\xi}_{-i1} \tag{2.6}$$

$$Y_{ig} = \alpha_3 + \beta \overline{Z}_{-ig} + \xi^*_{2i} \tag{2.7}$$

Where $\overline{Y}_{-ig} = \frac{\sum_{j \in N(-i)} D_j}{n(-i)}$ is the fraction of i's peers that have adopted SEJ, $\overline{Z}_{-ig} = \frac{\sum_{j \in N^*(-i)} Z_{-i}}{n^*(-i)}$ is the fraction of i's peers that are at the left of the $[x_0 - \Delta, x_0 + \Delta]$ window, where $N^*(-i) \equiv \{j : FPS_j \in [x_0 - \Delta, x_0 + \Delta\}$ and $n^*(-i) \equiv \#N^*(-i)$, and $\overline{\xi}_{-i1} = \frac{\sum_{j \in N(-i)} \xi_{j1}}{n(-i)}$ is an unobservable error inside network $N(-i)$. The equations before can be modified to include a vector $\mathbf{X}_i$ of individual and groups controls.

2.3.2 Empirical Strategy

This paper is interested in comparing the effect of two different networks. This work assumes that the correct reference groups are the coworkers and the classmates. It also assumes that both networks are independent (the validity of this assumption will be assessed in Subsection 2.5.1). The peer effect used here assume that the relevant measure is the mean behavior of the reference group, "but it could be the 90th percentile, or the 10th percentile, or possibly not just the mean, but perhaps also lower variance aids in enhancing individual achievement" (Boozer and Cacciola, 2001). The rich dataset allows myself to restrict to individuals that can be seen before the introduction of SEJ, and to measure the coworkers and classmates before the introduction of SEJ. This also allow me to fix the groups of coworkers and classmates in a point in time and to track those groups and the individuals that integrate them several months (and sometimes even years) before the introduction of the SEJ (depending on how old are they, of course). The idea with defining networks this way, is that it allows to isolate any changes induced by the introduction of the SEJ over the composition of the groups. This lessens concerns with endogenous group membership given the introduction of SEJ.

The coworkers of an individual i are defined according to the last employer before the introduction of SEJ. Two individuals i and j are coworkers if they share the same last employer, i.e. if they were working for the same employer at the same time. On the other hand, the classmates of person i are defined only for those who graduated from twelve grade (known as *cuarto medio*) in the year of 2009 or before. Two individuals i and j are classmates if they were registered in twelve grade, in the same educational institution and in the same year, as a result they are classmates of the same generation, not only schoolmates.

Now $\overline{Y}_{-ig}$ will have to be modified in order to account for both networks. Define the group of i's coworkers as $C(-i)$ of size $c(-i)$, and the group of i's classmates as $S(-i)$ of size $s(-i)$. Since I have assumed that the correct reference groups are composed of only coworkers and classamtes and with enough independence between them, then $N(-i) = \{C(-i), S(-i)\}$ of size $n(-i) = c(-i) + s(-i)$. As a result, $\overline{Y}_{-ig} = \frac{\sum_{j \in N(-i)} D_j}{n(-i)} = \frac{\sum_{j \in C(-i)} D_j}{n(-i)} + \frac{\sum_{j \in S(-i)} D_j}{n(-i)}$ because $\sum_{j \in N(-i)} D_j = \sum_{j \in C(-i)} D_j + \sum_{j \in S(-i)} D_j$. In the empirical setting I instead use $\overline{Y}_{-ig} = \frac{\sum_{j \in C(-i)} D_j}{c(-i)} + \frac{\sum_{j \in S(-i)} D_j}{s(-i)}$.

Define $\text{l\_net}_{-i} \equiv \frac{\sum_{j \in C(-i)} D_j}{c(-i)}$ and $\text{s\_net}_{-i} \equiv \frac{\sum_{j \in S(-i)} D_j}{s(-i)}$, then the structural equation (2.5) becomes:

$$Y_{ils} = \alpha + \rho_1 \text{l\_net}_{-il} + \rho_2 \text{s\_net}_{-is} + \delta_1 \mathbf{X}_i + \eta_{ils} \tag{2.8}$$

Where l_net_{-il} and s_net_{-is} measures, respectively, the fraction of i's coworkers and classmates with SEJ. This means that now there are two endogenous variables and two instruments, and so two equations in the first stage equation (2.6):

$$\text{l\_net}_{-il} = \gamma_1 + \pi_{11} \text{z\_lnet}_{-il} + \pi_{12} \text{z\_snet}_{-is} + \delta_{21} \mathbf{X}_i + \overline{\xi}_{-ils1} \tag{2.9}$$

$$\text{s\_net}_{-is} = \gamma_2 + \pi_{21} \text{z\_lnet}_{-il} + \pi_{22} \text{z\_snet}_{-is} + \delta_{22} \mathbf{X}_i + \overline{\xi}_{-ils2} \tag{2.10}$$

Where $\text{z\_lnet}_{-i} \equiv \frac{\sum_{j \in C^*(-i)} Z_j}{c^*(-i)}$ and $\text{z\_snet}_{-i} \equiv \frac{\sum_{j \in S^*(-i)} Z_j}{s^*(-i)}$ and they measure, respectively, the number of coworkers and classmates whose vulnerability score is at the left of the window divided by the number of coworkers and classmates whose vulnerability score is inside the window, i.e. the fraction of coworkers and classmates whose vulnerability score is at the left of the window. Notice that $N^* = \{C^*(-i), S^*(-i)\}$, where $C^*(-i) \equiv \{j \in C(-i) : FPS_j \in [x_0 - \Delta, x_0 + \Delta]\}$ and $S^*(-i) \equiv \{j \in S(-i) : FPS_j \in [x_0 - \Delta, x_0 + \Delta]\}$. Equation (2.8) is estimated by 2SLS, with first stages given by (2.9) and (2.10). The vector $\mathbf{X}_i$ are several controls at both the individual and group level, with leave-out mean characteristics at the group level, and will also work to carry out robustness checks.

Finally, the reduced form equation (2.7) is given by:

$$Y_{ils} = \alpha + \beta_1 \text{z\_lnet}_{-il} + \beta_2 \text{z\_snet}_{-is} + \delta_3 \mathbf{X}_i + \xi^*_{3ils} \tag{2.11}$$

The next section explains where the database comes from and variables available, but it is important to mention here that the individual controls are age, age^2, sex, a dummy if the person is migrant, FPS score, Ln(wage), potential SEJ payment, years of education, and a dummy for each of the four admissibility criteria explained at the end of section 2.4. The group level controls are the mean of each individual controls at both group levels, except for the fraction of peers with an FPS lower than 11,734 because this is linearly correlated with the instrumental variable. Hence, the estimating sample is restricted to people with no missing values in these variables, and is composed of 68,589 individuals. Appendix B.2 display Kernel density estimates and other descriptive statistics of the number of classmates and coworkers in this sample. In the construction of the instrument, I restrict myself to workers whose vulnerability score is inside a window of size $\Delta = 500$ around the cutoff $x_0 = 11,734$. I also center the vulnerability scores around 0 by subtracting $11,734$ from them. Since any shock common to the group creates spurious peer effects, standard errors are clustered using multi-way clustering at the (grade-school-year x workplace-year) according to Cameron, Gelbach and Miller (2006).

2.4 Data

This section describes the database used in this work and how it was constructed. Information was merged using a unique identification number and it comes from four different administrative datasets: data from the Unemployment Insurance (AFC) and the vulnerability score (FPS), administrative records from the Chilean Student Registration (RECH), and administrative records from the Youth Employment Subsidy (SEJ).

The FPS database contains information on socioeconomic characteristics of all individuals in Chile with a vulnerability score, independently of the economic status of the individuals (employed or unemployed, etc.) and nationality. The database is updated from time to time due to the application of surveys by municipalities as requested by the families in order to update the socioeconomic characteristics of the family, which can alter their FPS score, or to obtain a vulnerability score. Data used in this work contains information from December 2007 to September 2013, with a periodicity of March, September and December in most of the years. In this database, I obtained an individual's date of birth, sex, an indicator if the person was born in Chile, educational level, *comuna* of residence[8], and the vulnerability score and decile, but only for individuals between 18 and 25 years old.

The AFC database contains information of all the workers in Chile who have ever contributed to the unemployment insurance (UI) system since the law was implemented in October 2002. In Chile these are formal employed workers not hired in government institutions. This database allows me to identify if someone is employed or not at a given firm because an anonymous but unique ID is provided for each employer, as well as total taxable income in the last 6 and 12 months, and the number of months with taxable income in the last 6 and 12 months. Since the AFC database contains information of the labor history of only formal dependent workers, then information on the networks could be limited if the workers maintain a stronger relationship to other dependent informal or independent workers.

[8]Information on the *comuna* of residence is coded in a way that does not allow to know where each person lives. I can only know that two persons live in different areas because their *comuna* of residence codes are different between each other.

At the time the database was delivered, the AFC database was already merged to the FPS dataset described above using a nameless and masked ID. Notice that this "FPS+AFC" database provides very valuable and disaggregated information on both employed and unemployed individuals, before and after the SEJ program began, which allows me to know if a worker meets SEJ eligibility criteria or not, the labor network of formal dependent workers, individual and mean peers characteristics, and characteristics at the level of the employer.

The RECH database contains nationwide information on the academic history of children, youths and adults, enrolled in any educational establishments officially recognized by the Ministry of Education. The database is updated annually and contains information on more than 3.5 million students. Teaching modalities involved are preschool, primary and middle school for children, high school for youths, special (differential) and primary, middle and high school for adults (except the *Chile Califica* program). The two processes that feed information to the RECH require participation of all officially recognized educational establishments, both public and private. The variables requested in this work includes the student's nameless and masked ID, a masked and nameless ID of the educational establishment where the student is registered (RBD), study plan[9] , grade and classroom codes. Data provided is between 2002 and 2012, for all individuals registered in those years. The RECH database allows identifying, very precisely, the classmates of an individual, at different educational institutions, years and grades.

Finally, SEJ databases are annual administrative records of all monthly and annual payments of SEJ made by SENCE. Variables requested in this work only include a nameless and masked ID for SEJ receiver, an indicator of monthly or annual payment, and month and year of the subsidy. It does not matter if this database includes payments made to dependent or independent workers or to employers, because SEJ database is merged afterwards with the AFC+FPS database that includes labor information of only dependent workers. This database allows me to know when a worker begins to receive SEJ and if at any time t the worker is receiving or not SEJ. Unfortunately the SEJ database used here does not contain information on the amount of SEJ payments, nor the application date. It is also a database of workers who are getting paid, not a database of applicants, hence it limits the scope of the conclusions that can be drawn.

The admissibility criteria was explained in Section 2.2; however, with the data at hand, there is no way of telling if a person is working at a company with a state contribution higher than 50%; and if a person has social security contributions paid up to date. Hence, in this work, a person is considered admissible to receive SEJ if he:

1. is between 18 and 25 years old;
2. has a vulnerability score equal or less than 11,734 points in the FPS;
3. is working [10]; and
4. earns an annual gross income less than CLP$4,320,000

[9]The study plan refers to Educación Parvularia, Enseñanza Básica, Educación Especial, Enseñanza Media Humanista Científica, Enseñanza Media Técnico Profesional and Enseñanza Media Artística

[10] Note that these workers can be employed by a company with state contributions higher than 50%, so estimates are actually a lower bound because non-admissible people are taken as admissible.

Table 2.2 presents descriptive statistics on the main variables used in the work over different samples. Column (1) uses the full sample, column (2) uses a sample of workers with SEJ at some time, and column (3) uses a sample of workers without SEJ at some time. The sample consists of youths with a mean age of 21 years old, a small fraction are migrants, and it is well balanced between men and women with a mean *FPS* of 8,256 points (there is no comparison point at the national level). It contains youths with a mean wage of CLP$223,053 wich is lower than the mean national wage of CLP$355,771 according to CASEN (2011), and the occupancy rate is 51% which is lower than the mean national occupancy rate of 55% for the period 2010-2012. The mean years of education are 11 and they are higher than the national mean of 9.5 years of education during 2012 (INE, 2012). Only 76% had an FPS score lower than 11,734, most of them are between 18 and 25 years old, and only 45% have a wage lower than the $360,000 threshold; as a result only 33% are eligible to apply to SEJ. Column (4) shows that the sample of people with SEJ is, on average, quite different from the sample of people without SEJ on several characteristics, providing evidence of potential self-selection into SEJ participation and making it nonrandom. But the most interesting fact of all, is that the fraction of classmates with SEJ among people with SEJ is 13%, making it 3% higher than the fraction of classmates with SEJ among people without SEJ which is 10%. The same thing happens with the fraction of coworkers with SEJ, which is 17% among people with SEJ and 10% among people without SEJ, 6% higher. This suggests that there could be some social multiplier in the adoption of SEJ, and this is precisely what the next section attempts to quantify by removing potential concerns for a simultaneity and correlated unobservables bias.

2.5 Results

This section describes the results obtained using the empirical strategy described in subsection 2.3.1. Table 2.3 shows the main results, during September 2009. All regressions in this table include controls at the individual and group levels. Column (1) shows the coefficients from an OLS regression of equation 2.8 in a linear probability model (LPM). It shows that the individual probability of adopting SEJ increases by 12.9% and 32.8%, given a 10% increase in the fraction of classmates and coworkers with SEJ inside his network, respectively. However, these peer effects are subject to several known problems such as correlated unobservables and a reflection problem. The next columns try to solve these problems and to establish causality.

Columns (2) and (3) show first stage estimates of equations (2.10) and (2.9), respectively. They show the causal effect of the fraction of classmates and coworkers with a vulnerability score lower than the cut off and inside the 500 points window on the fraction of classmates and coworkers with SEJ, during September 2009. The F statistics shown follow the Angrist and Pischke (2009) suggestion because both first stages are used simultaneously in the second stage. Both first stages present an F-statistics higher than 10. These columns show that having a 10% higher fraction of classmates and coworkers at the left side of the small window increases the fraction of classmates with SEJ in 0.18% and coworkers with SEJ in 0.20%. Though not very large in magnitude, it is very precisely estimated.

Column (4) shows reduced form intention to treat (ITT) estimates. It shows that having 10%

more classmates with a vulnerability score at the left of the window had an average causal effect of increasing the individual probability of adopting SEJ by 0.17%, a point estimate that is significant at the 5% level. However, having a bigger fraction of coworkers with a vulnerability score at the left of the window had no impact over the individual decision to adopt SEJ. A test shows that the sum of the coefficients on the classmates and coworkers network equals zero (p-value 0.17).

The coefficient from a dummy that indicates if the person has an FPS score lower than the 11,734 cut off is 0.160 (s.e. 0.005)—not shown on this table. This coefficient measures a take up and it means that being admissible in FPS increases by itself the probability of adopting SEJ in 16%. But this effect is magnified because having all classmates at the left of the window increases the individual probability of adopting SEJ by 1.7%. Hence, the total increase in the probability of adoption is 17.7%[11]. For this reason, the ITT coefficients are social multipliers.

Finally, column (5) shows the Wald 2SLS estimates of equation (2.8) using columns (2) and (3) as first stages. These coefficients measure the average causal peer effect of the coworkers and classmates' SEJ adoption over individual adoption of SEJ. Unlike column (1), this column shows no evidence that the coworkers peer effect is different from zero. However, it also shows a positive average causal peer effect of working classmates over individual SEJ adoption during September 2009. Having 10% additional classmates with SEJ increases the individual probability of adopting SEJ by 9.1%. This is a very important peer effect when compared to other works, for example it is almost eight times as those calculated in Dahl, Loken and Mogstad (2014), as was expected from SEJ whose application process is so easy and cheap. Hence, admissible workers have a lot to gain from finding out information concerning the existence of SEJ and its application process.

2.5.1 Robustness Checks

This section describes some robustness checks performed to the baseline results. The first robustness check includes and excludes several controls, and changes the model specification from a linear probability model to a probit. These results are shown in Table 2.4.

The first two robustness checks in Table 2.4 adds fixed effects at the high school level. A *comuna* is the lowest and most basic administrative division of Chile, and corresponds to what in other countries is known as a municipality. The *comuna* is a division with local government purposes only, because in Chile the state government only extends at the regional and provincial level. This is interesting because despite the fact that the SEJ program was managed by the state government institution SENCE, some municipalities could have characteristics that made them more or less interested in participating in SEJ. The reason to add these fixed effects is to control for that unobserved heterogeneity at both the high school and *comuna* levels. On the other hand, high schools with a higher fraction of classmates at the left of the window could also have many unobserved characteristics—they have worst classmates, teachers, infrastructure, etc.—that makes them more

[11] In separate regressions (not shown) with only individual controls (and no group controls) and a linear trend in the FPS in both sides of the discontinuity, the coefficient from a dummy that indicates if the person has an FPS score lower than the 11,734 cut off is 0.165 (s.e. 0.005), this coefficient measures the individual take-up rate near the discontinuity. When this dummy is excluded and the two IV are included, the coefficient from $Z(S(-i))$ is 0.015 (s.e. 0.006) and from $Z(C(-i))$ is -0.005 (s.e. 0.005). These coefficients are robust to the use of a quadratic or third degree polynomial in the FPS.

susceptible to apply to SEJ and this could violate the exclusion restriction. Rows B and C show that, despite of the inclusion of the fixed effects, the first stage remains very strong and unchanged; moreover, the second stage coefficients do not deviate too much from the baseline results.

Another concern could be that the classmates network is no more than a proxy for an already existing geographic network, since people usually do not live very far from the high school they attend to, and these geographic regions could have unobserved characteristics that also makes them more susceptible to apply to SEJ. Row D controls for the percentage of classmates living in the same *comuna*. The inclusion of this additional control do not change the first stage coefficients, and the second stage coefficients remain almost unchanged.

Row E considers only admissible youths in the estimating sample, peer effects are stronger among this sample because admissible youths are the main winners from having more classmates with SEJ. Row F, considers running the baseline regression without any control variables. Notice that the coefficients from the first stages change extremely little. This confirms that the instrumental variable was quasi-randomized in the window around the cut off and hence, this assures that it is independent from other covariates.

Row G considers a LPM with one endogenous and one instrument, it shows that the coefficients from the first stages change very little; hence both instrumental variables are quite independent from each other. This means that the classmates and coworkers networks were quite independent from each other, so someone's classmates are not at the same time his coworkers. Notice that the 2SLS coefficient of the classmates in row G is almost equal to one. Boozer and Cacciola (2001) show that the instrumental variables estimator is algebraically exactly equal to 1 for the empirical endogenous peer effects regression, without covariates, where the researcher uses characteristics of the *full* group in the sample as either an instrument or regressor[12]. Row F should in part reduce this concern because none of the coefficients in column (5) row F equal 1, since not everybody is somebody's coworker or classmate. Remember that the sample where regressions are calculated is restricted to people with information on the fraction of his classmates and coworkers with SEJ, and with information on both the individual and group controls. Hence, due to sample restriction, the sample of 68,589 persons is composed of a subgroup of the group used to construct l_net$_{-i}$ and s_net$_{-i}$. But, moreover, if Boozer and Cacciola (2001) was the driving force in these results then the reduced form and first stage coefficients from row H should be the same, and they are not. Row H shows row G with no controls and no restriction on the sample. The idea here is that, in this case, the coefficients from the first stage and reduced form would be the same if all the Y_{ils} are used to construct Y_{-ils}. The main reason why these coefficients are not equal to one is that the instrument is constructed as a "leave-out mean", and it uses only a sub-group of the full group in the sample as an instrument.

Finally, row I shows coefficients form a non-linear probit model. The reduced form follows a probit model, the first stages are OLS, and the second stage is an instrumental variables probit model calculated following Newey (1987).

Another potential concern is causality or the reflection problem; if everyone inside a network is adopting SEJ at the same time then a very important peer effect could be identified even though

[12] See Appendix B.2.

peers could not necessarily be talking about SEJ adoption or interacting between them in the application process, it just happens that all of them are applying at the same time. If the robustness checks and the placebo tests did not convince the reader, the next test in Table 2.5 tries to reduce this concern by looking at the effect of early adopters during July 2009 over the individual probability of adopting SEJ in September 2009. The results in this table show two strong first stages, and a second stage where a 10% increase in the fraction of classmates with SEJ during July 2009 increased the individual probability of adopting SEJ during September 2009 by almost 14%.

Altogether, the results are quite robust to the inclusion of different variables and do not depend on the linearity of the selected model. Moreover, the identified peer effect is not just a product of a reflection problem, since there are also important peer effects when early adopters are studied.

2.5.2 Testing the exclusion restriction

Identification requires the instrument, the fraction of peers at the left of the window, to operate though a single known causal channel, and to be as if randomly assigned. In other words, in the absence of differences in the fraction of peers at the left of the window, firms and high-schools with a low fraction of peers at the left of the window would not have been different on average from firms and high-schools with a high fraction of peers at the left of the window.

To shed light on the plausibility of this assumption, Table 2.6 regresses a variety of characteristics at the classmates and coworkers levels on both instrumental variables. These regressions are similar to equations (2.9) and (2.10) but instead of using s_net_{-is} and l_net_{-il} at the left side, I take the mean of characteristics specified in each row at the level of the classmates in column (1) and coworkers in column (2). Overall, this Table shows that nothing else changed around the cut-off. High-schools and firms are not any different around the discontinuity in terms of age, fraction of women, years of education, fraction of repeating students, fraction of migrants, fraction employed, number of classmates and coworkers, and number of classmates and coworkers with SEJ. There are some differences in monthly wage and fraction of peers employed, both at the 10% level, but fortunately with a very small coefficient as to be of any concern, under 0.16% and 0.12%, respectively, in the worst of the cases where all the coworkers are at the left of the window. But, moreover, the fraction of coworkers with a wage under CLP$360,000 is not significant.

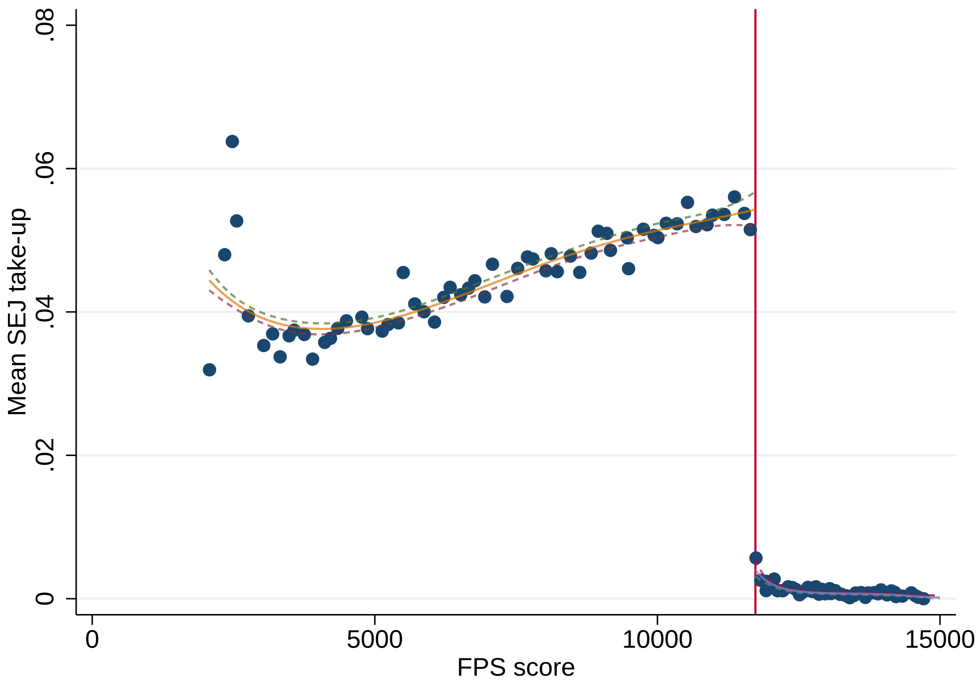
Repeating students are those who appear twice or more in the RECH as registered in *cuarto medio*. This could proxy for test grades, which I lack information from, because repeating students are expected to score several times lower than students who successfully pass the grade at the first shot. The idea is to test if the exclusion restriction is violated because students with a higher fraction of classmates and coworkers at the left of the window could also be more or less susceptible to apply to SEJ. Fortunately there is no effect of z_snet and z_lnet.

Although the last check is very good for very low test scores at left tail of the distribution, it is not so for test scores where the distribution of scores is more concentrated. However, if test scores proxy for some unobserved ability of the students, it is expected that they be biasing downward the results. If students with worst unobserved characteristics are at the left of the window, then they would also be less able to apply to SEJ and the peer effect should be biased downwards and not upwards.

2.5.3 Mechanisms

Before explaining the potential mechanisms through which the peer effects are operating, it is important to mention what is driving the first stage. Figure 2.3 shows individual mean SEJ take-up as a function of the FPS score. As can be seen from the figure, the probability of adopting SEJ is quite uniform across the different vulnerability scores, and is not concentrated in just a small interval. As a result, the fraction of peers at the left of the small window is a good predictor inside and outside this window of who is applying or not to SEJ. The dataset shows some individuals that were receiving SEJ and had a vulnerability score higher than 11,734. This is more likely to be due to the assumptions at the time of construction the database than due to flexibility of SENCE at the time of deciding if someone was able to receive SEJ or not.

Figure 2.3: Take up rate



Notes: This figure plots individual SEJ take-up in September 2009 against the FPS score . Each circle is the average SEJ take-up within 60 bins of width 161 FPS points at the left and 54 bins of width 60 FPS points to the right of the cutoff. Solid lines are fitted values using four-order polynomial logit regression on either side of the discontinuity. Dotted lines are 95% confidence intervals.

But what is driving the peer effects? Table 2.7 shows results using subsamples of small and large firms and schools. Small (large) enterprises are networks composed with less (more) than 100 coworkers, and small (large) schools are networks composed with less (more) than 100 classmates. The 100 threshold was selected by analyzing detailed descriptive statistics of the number of members in each network[13] and a rule of thumb. See the footnote at the end of the table for a detailed description on how this table was constructed. Notice that the peer effects are mostly driven by groups of small size, such as small schools and firms. First stage results are not shown, but they are

[13]The mean number of peers in each group was 77 coworkers and 66 classmates, the medians were 8 coworkers and 57 classmates

stronger in small size groups, and quite weak among larger groups. However, the reduced form is not biased and it also shows no significant peer effects among large schools and firms.

On the other hand, the importance of classmates found in section 2.5 is especially likely in this context to be due to informational benefits that get translated into adoption. There are two potential informational benefits or channels of information, first about the existence of the program, and second about the benefits and costs of participation.

One channel of information is about the existence of the social program, which is very likely given that SEJ had just been introduced. However, knowledge about the existence of the program would not automatically translate into adoption due to its unknown application procedure, admissibility rules, and other details of how it works, which are points that remain very unknown for a recently launched welfare program. For example, during November 2014 the Minister of Labor did a media campaign where she encouraged the 122,000 persons enrolled at SEJ and women employment subsidy to claim USD$37,029,382 of unclaimed payments. The accumulation of unclaimed payments are sometimes due to ignorance of how SEJ works. As explained above, a person does not lose the benefit when he changes work, but the payments are held until he becomes eligible once again, then payments are resumed automatically. In light of this view, payments where left unclaimed by many workers, that in some individual cases summed up to USD$2,067[14]. In total, during this date, 1,800 youth workers may claim amounts higher than USD$1,378[15].

Another channel of information is knowledge about the benefits and costs of participation, including a stigma cost. In this setting, the initial information set about these is quite limited. But the admissibility requirements allowed some networks to quasi-randomly have more/less peers around a small window of the cut off while leaving everyone else intact. The results show that this exogenous increase in the fraction of admissible peers that were allowed to apply translated into a higher adoption of SEJ inside each network, hence reducing uncertainty about the costs and benefits.

With the current data, it is very hard to tell what these youths are talking about, whether about the benefits, costs, application procedure, etc, and the type of information that is being transmitted inside each network—which can be different within each network. Classmates could be talking more about SEJ existence while the coworkers can provide more details about its costs and benefits. Despite this, there are several subsamples where the explanation is more likely to hold. The first one is among younger workers, who have not yet or have recently graduated from high-school and so the network of their classmates is more recent, hence they are more likely to keep a closer interaction or appeal to them when they need information about social programs. Table 2.8 tests if peer effects among young workers are stronger. The results show that the classmates are important in both young and old workers, but the magnitude of the peer effect is larger for workers under 21 years old.

There are other subsamples where peer effects are expected to be smaller or larger. Panel B shows results for students that come from two different kinds of high schools. In Chile, there are three modalities of high schools. The first one are the *técnico profesional* high schools whose main

[14]"Bono Trabajo Mujer y Subsidio al Empleo Joven: llaman a cobrar más de 21 mil millones", 24horas.cl, November 3, 2014; and "Ministra de Trabajo llama a cobrar bonos Trabajo Mujer y Empleo Joven", 24horas.cl, November 2, 2014

[15] "Bonos y subsidios Sence: llaman a cobrar $21 mil millones no reclamados", 24horas.cl, October 24, 2014

objective, according to OIE (1993), is to meet the needs of the labor market, and the educational and business policies, as well as the students' interests. Overall, it seeks to facilitate the exchange and dialogue between teachers and labor market institutions. These high schools, in constrast from the *científico humanista* high schools, are more oriented in constructing abilities that help students find a job quickly after graduating, they are hired more easily because they can signal that they are more productive. But also, students that attend these schools are usually more interested in finding a job after graduating than students from *científico humanista* high schools, creating a selection bias. According to Cáceres and Bobenrieth (1993), though not causal, they also receive a higher wage than students graduating from *científico humanista*s. As a result, peer effects among students graduating from *técnico profesionales* should be higher than peer effects among students graduating from *científico humanistas*. This is precisely the result from panel B. The third modality are the *Artístico* high schools, but there are no people enrolled in *cuarto medio* in those high schools at the RECH database until 2011, because the *Artísticos* were not implemented until 2006.

Panel C shows that the point estimates of the peer effects among those with a monthly wage lower than CLP$200,000 is just marginally larger than the point estimates of the peer effects among those with a monthly wage higher than CLP$200,000. Notice that the benefit represent a higher fraction of the monthly wage for those with a wage lower than CLP$200,000, which are those whose monthly wage is between the increasing part and the plateau of Figure 2.2. This suggests that peer effects were equally important for people whose potential payment from SEJ would had been high and for those whose payment would had been low.

Finally, peer effects are also expected to be stronger among small groups, because adoption of SEJ by one member could be more noticeable and could provide information more easily. There is little evidence in this area of knowledge, of how a peer effect changes as the group size changes. Levine and Moreland (1990) commented that "as a group grows larger, it also changes in other ways, generally for the worse. People who belong to larger groups are less satisfied, participate less often, and are less likely to cooperate with one another" (p. 593)[16]. Table 2.7 show results when the networks are divided by group size, it shows that the peer effects are several times larger for individuals that comes from small enterprises and schools, compared to the baseline results. The results for small enterprises and schools are also around half time larger than the baseline results.

Table 2.9 shows results using networks with weaker ties. Panel B expands the definition of a classmate to schoolmates, all those students who attended the same school in *cuarto medio* regardless of the year. For example, if person A was enrolled in twelve grade in 2001 and person B was enrolled in twelve grade during 2006 and in the same school as person A, then A and B are classmates. Column (5) shows no schoolmates peer effect. This provides evidence that the relevant classmates are those of the same generation. Panel C changes the definition of a coworker to all those employees who worked in the same firm before SEJ irrespective of the year. For example, if the last employment of person A was during March 2008 and the last employment of person B

[16]There is some evidence in the education literature that group size has an important effect, almost always negative, over academic results. Krueger and Whitmore (2001) showed that attending a small class in the early grades, as Project STAR did, was associated with higher performance on standardized test, and an increase in the likelihood that students take a college-entrance exam.

was during March 2009 and in the sample firm as A, then A and B are coworkers. I do not find any coworkers peer effect in this weaker network either, and the classmates peer effect remains very similar to the baseline results in Panel A.

There are several reasons why a coworker peer effect could not be relevant. First, there could be high turnover in employment, these youths could also spend many periods unemployed, and since the SEJ is only for workers hired in the formal sector, people could be shifting from informal to formal labor market.

2.5.4 Other dates

The results form Section 2.5 show how classmates's take up of SEJ had an important effect over individual take up of SEJ during September 2009. But, does this effect hold over time? There is no reason to believe that peer effects can remain constant over time, they can very well increase or decrease over time, and they can even shift in importance between one network or another. Table 2.10 shows the peer effects six months after SEJ was implemented. These regressions are calculated independently using a same sample of 47,373 persons. This sample is different from the baseline sample for several reasons[17], but the main objective of using this new subsample is to track the sample people several months after SEJ was implemented and to have a uniform sample to calculate the regressions. It is important to mention that the results for September 2009 using this new subsample are quite similar to the baseline results reported in Table 2.3. As can be seen in Table 2.10 and column (5), a 10% increase in the fraction of classmates with SEJ increased the individual probability of adopting SEJ by 8%, but the coworkers peer effect is not statistically different from zero. The knowledge effect of the social multiplier disappears six months after in March 2010, a 10% increase in the fraction of admissible classmates or coworkers at the left of the window had no impact over the individual probability of adopting SEJ, even when the first stages remain very strong. Column (5) shows that a 10% increase in the fraction of classmates and coworkers with SEJ did not change the individual probability of adopting SEJ. This variation over time provide interesting evidence on the behavior of the peer effects over time and how they can quickly change in just a matter of a few months. These hypothesis deserves further research in future work.

Peers could be more important during the early stages of SEJ implementation for several reasons. First, information about the existence of SEJ, and other details of its application procedure and costs and benefits was very scarce during the initial period, but afterwards many people learned about it and peers became less relevant. Second, peer effects could be similar to a herd behavior of a trend topic, where the peer effect is very important at the beginning when advertising is undergoing, but as the advertising looses terrain, the amount of time devoted to SEJ in the conversations also does so. Third, peers shift in importance and get depreciated over time, and since the classmates definition is fixed here, they can become less important as people get older, but this does not necessarily means that people stopped learning about the existence of SEJ and other details of its application procedure and its costs and benefits. It is just that they are no longer talking with their classmates about SEJ. Fourth, Bravo and Rau (2013) finds that admissibility to SEJ had an impact over the

[17] For example, there is aging out, where people over 25 years old leave the sample; and there are also people going out and coming into the labor market.

economic cycle of employment but this effect decays over time, as a result, adopting SEJ several months after its implementation just because someone told me so becomes less relevant. Finally, all those who were interested in adopting SEJ have already adopted it. Altogether, classmates help to solve a lack of information, but once the information is available, they lose importance.

2.6 Targeted Advertising

This section exploits the results from an econometric model with very similar coefficients to those in Table 2.3. Results from this model showed important peer effects among the classmates, but not as big among the coworkers. Hence, the initial distribution of people with SEJ among the networks could determine the take up rate not only initially, but also over time. The initial assignment of people with SEJ can be interpreted in many ways, but an interesting one is to see it as a targeted advertising campaign where the government targets people within groups that meet certain criteria and successfully increases take up among them, with the purpose of influencing the decision of their peers. Assume the government is interested in increasing SEJ take up. Would there be any increase in the take up rate[18] if the government had initially exploited the structure of the network and targeted specific groups with SEJ advertisement? What happens when the initial persons with SEJ are assigned among the schools and firms according to different conditions? This section attempts to answer this question by exploiting the econometric model described in section 2.3.

Appendix B.4 shows a detailed description of the algorithm used to calculate the take up rates, but some important comments are at hand. During September 2009 there were 47,047 early adopters with SEJ and 323,919 were admissible to apply to SEJ, which implies a take-up rate of 14.5%. The first step in this section is to recalculate an econometric model without covariates. This econometric model has similar coefficients than the econometric model in Table 2.3. The structural equation is given by:

$$y_i = \alpha + \rho_1 \text{s net}_i + \rho_2 \text{l net}_i + u_i$$

Here, the constant α plays an important role because it measures the probability when s_net$_{-is}$ or l_net$_{-il}$ are near zero. The first stages are:

$$\begin{aligned} \text{s net}_{-is} &= \alpha + \gamma_{11} \text{z snet}_{-is} + \gamma_{12} \text{z lnet}_{-il} + \varepsilon_i \\ \text{l net}_{-il} &= \alpha + \gamma_{21} \text{z snet}_{-is} + \gamma_{22} \text{z lnet}_{-il} + \varsigma_i \end{aligned}$$

The variables s_net, l_net, z_snet and z_lnet are constructed using information from 399,728 persons, but the econometric model is calculated using only a sample of 89,440 persons with just enough information. The second step is to distribute the 47,047 early adopters among the sample according to five different criteria[19]. According to a random advertising campaign, where they

[18] Take up in this section will be understood as the number of people with SEJ divided by the total number of people, instead of the formal definition of take up as the number of admissible people with SEJ divided by the number of admissible people.

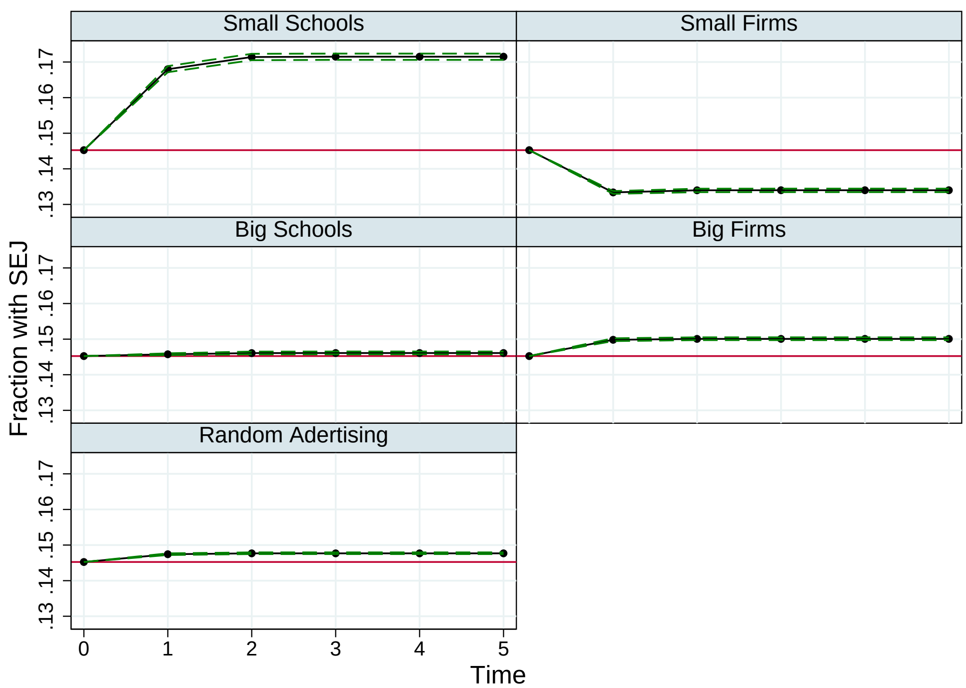
[19] In order to keep things simple, the assignment of SEJ is made regardless of the admissibility criteria.

are randomized among the 323,919 eligible; or according to targeted advertising where they are randomized among all the people within small or big schools; or randomized among all the people within small or big firms. Small and big sizes are defined as groups with less than 50 persons[20], but results using different thresholds are available upon request from the author.

When the econometric model is used to predict the individual probabilities of adopting SEJ, it can be applied to only 129,563 persons with both a school and labor networks (hence with no missing values in both s_net and l_net variables). If the predicted probability is higher than a certain threshold p* and the person did not have SEJ before, then it is assumed that the person adopts the subsidy. Notice that the randomization of who gets SEJ is made among 323,919 admissible persons, but the take up rates are calculated among only 129,563. Then, the amount of people with SEJ in the 129,563 sample could be different in each replication.

Figure 2.4 shows results when this exercise is calculated for 5 periods and 100 replications. The threshold p* is established at 0.3 (results for other values are available upon request), but it is important to mention that the optimal cut-off according to Hanley and McNeil (1982) ROC analysis is 0.14. The vertical axis measures the fraction of people with SEJ among the 129,563. For comparison, the red line in this figure is drawn at 14.5%, which is the initial fraction of people with SEJ in reality (47,047 of the 323,919).

Figure 2.4: Evolution of take up rates over time using targeted advertising.



Notes: Fraction with SEJ refers to the number of people with SEJ divided by the number of people. The dashed lines represent 95% confidence intervals. The red line refers to 12.4%, the initial fraction of people with SEJ; p*=0.3, n=100.

There are several fascinating results. First, as can be seen from comparing the five different graphs, the highest increase in take up is achieved when the early adopters are distributed only

[20]Choosing a bigger cut off would exhaust available data.

among small schools, in this case the take up increases from 14.5% to 17.1%. The main reason is that, at the margin, one additional classmate with SEJ is more important inside a small school than one additional classmate with SEJ inside a large school. As a result, s_net increases more inside a small school and since the driver of the peer effects are the school networks, then the individual probability of adopting SEJ next period increases substantially.

Second, distributing the early adopters among small firms has a negative effect over the take up. The reason is similar to the previous one. One additional adopter inside a small firm increases l_net more than one additional adopter inside a large firm. However, since the coworkers peer effect is negative, this increase in l_net reduces the individual probability of adopting SEJ. This is the reason why the take up decreases over time in the case of small firms.

Third, notice that some people were inside both a classmates and a coworkers networks. This is the reason why in the take up do not decrease when the early adopters are distributed among big firms. In this case it is easier to find someone who belong to someone else's classmate network. And since the marginal adopter in a big firm do not weigh too much, then the classmate peer effect will be more prevalent.

The reasons described above help explaining why distributing the early adopters among big schools do not increase the take up more than distributing them among big firms. On one hand, the marginal adopter inside a big school increases s_net only by little; but on the other hand, these are someone else's coworkers; the s_net has a positive coefficient while the l_net has a negative one, so both effects are net out.

Finally, randomly distributing the early adopters among the full population achieves a small increase in the take up. This increase is so small because under this scenario, the early adopters are distributed among people without a network and people inside a network. This explains the small increase, because there are people who is getting the benefit but since they are isolated, then there is no social multiplier that increases take up through time.

The main conclusion from this exercise is that, if the government had chosen to target schools with advertising, and especially small schools, instead of firms, the take up rate of SEJ nowadays could had been higher.

2.7 Conclusions

This work has quantified an endogenous peer effect in the decision to adopt a Chilean subsidy known as the Youth Employment Subsidy (SEJ). It did so by using an instrumental variable that is quite independent and quasi-randomly distributed among the population, and so it is very likely to also be independent of common unobservable shocks at the group level. Networks are defined before the introduction of the SEJ, as to deal with endogenous group membership, as a result all changes in the groups' composition can be taken as orthogonal to the introduction of the SEJ.

The results show that during September 2009, the classmates played a vital role in determining participation in the SEJ program, but the coworkers were not as relevant. The results show that a 10% higher fraction of classmates with SEJ implies a 9.1% higher individual probability of adopting SEJ (significant at the 5%); while a 10% larger fraction of coworkers with SEJ implies a reduction of

-25.2% in the individual probability of adopting SEJ but statistically equal to zero.

There are several potential explanations for the existence of peer effects, but the most likely one is due to an informational channel about the existence and unknown application costs, benefits and other details of SEJ. Given the particular case of SEJ, were costs are extremely low, this information translated into adoption, and this explains the very important peer effects during the early months of its implementation. However, they decay six months after and become insignificant. The reason is that the classmates help to solve a lack of information, but once the information is available, they lose importance.

This hypothesis is sustained by a second set of results. They show that peer effects are more important among younger workers who recently graduated from high school, especially if they attended *cuarto medio* at a *técnico profesional* high school, and within smaller networks of classmates and coworkers. These results are interpreted as providing evidence that an informational channel is the most important driver of the peer effects, because younger workers interact more often with classmates compared to older workers, and in small size groups it is easier to realize if someone has adopted SEJ. However, they deserve further research because the first stage is weak.

When the early adopters of SEJ are distributed according to several criteria the results show that the highest increase in take up is achieved when the early adopters are distributed only among small schools. As a result, if the government had chosen to target small schools with advertising, and especially small schools, instead of firms, the take up rate of SEJ nowadays could had been higher.

There are better methodologies to solve the reflection problem than the one used in this work and this leaves a lot of work to be done. The problem is that the estimating sample includes both eligible and ineligible individuals; when the sample is restricted to individuals outside the window used to construct the IV, then I would be allowing a subpopulation to become eligible while this does not affect the second group (Dahl, Loken, and Mogstad, 2014). In this case, the reflection problem is solved because the two equations that make up the simultaneous equation system that determines social program participation now includes an IV in only one of the two equations. Future work can exploit this other methodology to study peer effects. For example, program eligibility in the *Chile Solidario* (CS) program is a also a discontinuous function of the FPS. Carneiro, Galassoy, and Ginja (2014) find that CH participants increased take-up by CS recipients of a family allowance for poor children (the *Subsidio Unico Familiar*, SUF) because CS provided information of other social programs. This allows to study take up rates of SEJ and other social programs among ineligible peers from recipients of CS. On the other hand, Bravo and Rau (2013) finds that employment and participation rates increased among the eligible population of SEJ in the first six months of implementation. Once again, this result can be exploited to study what happens with the employment and participation rates inside networks who had at least one peer inside a very small window around the 11,734 cut off.

Table 2.2: Descriptive Statistics

	Full Sample (1)	With SEJ (2)	Without SEJ (3)	Difference (2) - (3) (4)
Age	21.4	22.0	21.4	0.64
	[2.22]	[1.71]	[2.23]	[0.004]
Women	0.50	0.51	0.50	0.01
	[0.50]	[0.50]	[0.50]	[0.001]
Migrant	0.004	0.003	0.004	-0.001
	[0.06]	[0.052]	[0.06]	[0]
Vulnerability Score	8,256	6,733	8,325	-1,592
	[3,890]	[3,034]	[3,910]	[6.4]
N	*8,826,468*	*381,963*	*8,444,505*	
Wage	223,053	213,414	223,936	-10,522
	[143,000]	[75,600]	[147,000]	[242]
Potential SEJ Payment	30,484	36,294	29,952	6,342
	[16,900]	[12,800]	[17,100]	[28.5]
N	*4,510,280*	*378,532*	*4,131,748*	
Years of Education	11.3	11.4	11.3	0.18
	[2.35]	[2.13]	[2.36]	[0.005]
N	*3,485,751*	*260,445*	*3,225,306*	
1(FPS≤11,734)	0.76	1.00	0.74	0.74
	[0.43]	[0.06]	[0.44]	[0]
1(18<age<25)	0.95	1.00	0.95	0.95
	[0.21]	[0.06]	[0.21]	[0]
Employed	0.51	0.99	0.49	0.50
	[0.5]	[0.09]	[0.5]	[0.001]
1(w<$360,000)	0.45	0.96	0.43	0.54
	[0.5]	[0.19]	[0.49]	[0.001]
Admissible to apply to SEJ	0.33	0.96	0.30	0.66
	[0.47]	[0.21]	[0.46]	[0.001]
N	*8,826,469*	*381,963*	*8,444,506*	
Fraction of classmates with SEJ	0.10	0.13	0.10	0.03
	[0.07]	[0.07]	[0.07]	[0]
N	*2,999,945*	*2,701,282*	*298,663*	
Fraction of coworkers with SEJ	0.11	0.17	0.10	0.06
	[0.16]	[0.19]	[0.16]	[0]
N	*891,344*	*791,627*	*99,717*	

Notes: Standard deviation in square brackets

Table 2.3: Effect of peer's adoption over individual SEJ adoption, September 2009.

	September 2009				
	OLS	First Stage		Reduced	Second
		Classmates	Coworkers	Form	Stage (2SLS)
	(1)	(2)	(3)	(4)	(5)
Classmates	0.1292***	0.0184***	0.0003	0.0167***	0.9113***
	[0.0224]	[0.0033]	[0.0020]	[0.0063]	[0.2718]
Coworkers	0.3281***	0.0002	0.0201***	-0.0048	-0.2521
	[0.0189]	[0.0007]	[0.0039]	[0.0049]	[0.2622]
F-statistic (AP)		*32.2*	*26.0*		
N	*68,589*	*68,589*	*68,589*	*68,589*	*68,589*

Notes: Standard errors are clustered using multi-way clustering at the (grade-school-year x workplace-year) according to Cameron, Gelbach and MIller (2006). Individual controls include age, age squared, sex, a dummy if not Chilean, vulnerability score, Ln(Wage), potential SEJ payment, years of education, number of coworkers and classmates; group controls include the mean individual controls, except age squared. Controls also include four dummies, one for each of the four admissibility criteria. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.4: Robustness checks for the peer effects, September 2009.

		First Stage			Reduced	Second	
		Classmates	Coworkers	*F-AP*	Form	Stage (2SLS)	*N*
		(1)	(2)	(3)	(4)	(5)	(6)
A	Baseline						
	Classmates	0.0184***	0.0002	*32.2*	0.0167***	0.9113***	*68,589*
		[0.0033]	[0.0007]		[0.0063]	[0.2718]	
	Coworkers	0.0003	0.0201***	*26.0*	-0.0048	-0.2521	
		[0.0020]	[0.0039]		[0.0049]	[0.2622]	
B	School fixed effects						
	Classmates	0.0162***	0.0001	*21.7*	0.0166**	1.0230***	*68,411*
		[0.0035]	[0.0006]		[0.0069]	[0.3646]	
	Coworkers	0.0000	0.0197***	*88.6*	-0.0053	-0.2757	
		[0.0022]	[0.0021]		[0.0041]	[0.2213]	
C	Comuna + School FE						
	Classmates	0.0163***	-0.0001	*21.9*	0.0162**	0.9945***	*68,411*
		[0.0035]	[0.0006]		[0.0068]	[0.3586]	
	Coworkers	-0.0003	0.0199***	*90.0*	-0.0046	-0.2289	
		[0.0022]	[0.0021]		[0.0041]	[0.2162]	
D	Percentage of classmates living in the same comuna						
	Classmates	0.0183***	0.0002	*32.1*	0.0167***	0.9168***	*68,589*
		[0.0033]	[0.0007]		[0.0064]	[0.2732]	
	Coworkers	0.0003	0.0201***	*26.0*	-0.0048	-0.2516	
		[0.0020]	[0.0039]		[0.0049]	[0.2620]	
E	Only elegibles included						
	Classmates	0.0169***	0.0021	13.7	0.0213**	1.3330*	38,177
		[0.0046]	[0.0027]		[0.0108]	[0.5397]	
	Coworkers	0.0006	0.0194***	16.2	-0.0104	-0.5737	
		[0.0012]	[0.0048]		[0.0078]	[0.4921]	
F	No Controls						
	Classmates	0.0203***	0.0014	*20.2*	0.0227***	1.1167***	*68,589*
		[0.0045]	[0.0011]		[0.0072]	[0.2522]	
	Coworkers	0.0003	0.0213***	*21.9*	-0.0017	-0.149	
		[0.0022]	[0.0045]		[0.0054]	[0.2561]	
G	One endogeouns, one instrument						
	Classmates	0.0184***		*32.0*	0.0167***	0.9067***	*68,589*
		[0.0033]			[0.0060]	[0.2783]	
	Coworkers		0.0201***	*27.0*	-0.0048	-0.2404	
			[0.0039]		[0.0043]	[0.2498]	
H	LPM excactly identified and no controls						
	Classmates	0.0169***		*21.0*	0.0072*	0.4287***	*326,153*
		[0.0037]			[0.0037]	[0.1268]	
	Coworkers		0.0206***	*25.2*	-0.0017	-0.0812	*132,709*
			[0.0041]		[0.0042]	[0.2178]	
I	Probit model						
	Classmates	0.0156***		*17.0*	0.0910*	5.4336**	*68,550*
		[0.0038]			[0.0377]	[1.7895]	
	Coworkers		0.0178***	*21.0*	-0.0346	-2.6004	

Notes: a/ the F statistic in rows B and C are not Angrist and Pischke (2009) but just an F test that the excluded instruments equal zero. In rows A-F, the standard errors are double clustered at the firm and schooling network. In rows G and H, the standard errors are clustered at whether the firm or the schooling network. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.5: Peer effects by early adopters during July 2009

	OLS	First Stage		Reduced	Second	N
		Classmates	Coworkers	Form	Stage (2SLS)	
	(1)	(2)	(3)	(4)	(5)	(6)
A BASELINE, SEPTEMBER 2009						*68,589*
Classmates	0.1292***	0.0184***	0.0003	0.0167***	0.9113***	
	[0.0224]	[0.0033]	[0.0020]	[0.0063]	[0.2718]	
Coworkers	0.3281***	0.0002	0.0201***	-0.0048	-0.2521	
	[0.0189]	[0.0007]	[0.0039]	[0.0049]	[0.2622]	
F-statistic (AP)		*32.2*	*26.0*			
B ADOPTION DURING JULY 2009						*68,589*
Classmates	0.1796***	0.0119***	-0.0001	0.0167***	1.3976***	
	[0.0304]	[0.0024]	[0.0014]	[0.0063]	[0.4554]	
Coworkers	0.4136***	-0.0002	0.0096***	-0.0048	-0.482	
	[0.0263]	[0.0005]	[0.0029]	[0.0049]	[0.5673]	
F-statistic (AP)		*24.2*	*10.7*			

Notes: Specifications mirror those in Table 2.3. Early adopters during July 2009 are those who received SEJ in July 2009, and they are used to construct the numerator of the variables s_net and l_net, number of peer with SEJ. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.6: Placebo Checks

Mean peers characteristics	Classmates IV	*R-squared*	Coworkers IV	*R-squared*
	(1)		(2)	
Age	-0.014	0.86	-0.0036	0.06
	[0.017]		[0.020]	
Sex	0.012	0.26	0.0129	0.32
	[0.012]		[0.008]	
Years of Education	0.000	0.12	-0.0002	0.02
	[0.0003]		[0.0004]	
Migrant	-0.021	0.45	-0.0101	0.13
	[0.013]		[0.021]	
Repeting student	-0.001	0.06	-0.0005	0.02
	[0.001]		[0.001]	
Ln(Wage)	-0.010	0.47	-0.0155*	0.34
	[0.008]		[0.009]	
$1(w \leq 360{,}000)$	0.003	0.60	-0.0006	0.03
	[0.004]		[0.003]	
Employed	-0.006	0.29	-0.0121*	0.08
	[0.007]		[0.006]	
Number of peers	-1.557	0.13	4.6557	0.03
	[2.532]		[9.978]	
Number of peers	0.529	0.24	1.5391	0.03
with SEJ	[0.418]		[2.084]	

Notes: This table shows results of a regression with the dependent variable in the left column and explanatory variables z_lnet and z_snet and a set of individual controls that include age, age squared, sex, a dummy if the person is not Chilean, Ln(wage), years of education, vulnerability score, potential SEJ payment, and dummies for each admissibility criteria. Column (1) uses mean characteristics from classmates and presents only the estimated coefficient for z_snet; column (2) uses mean characteristics from coworkers and presents only the estimated coefficient for z_lnet. Standard errors clustered by grade-school-year for column (1), and by workplace-year before SEJ-date for column (2). * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.7: Subsamples by group size

		Reduced Form (1)	Second Stage (2)
A.	SMALL SCHOOLS		
	Classmates	1.1402**	1.2089***
		[0.4036]	[0.3532]
	Coworkers	-0.3348	-0.3429
		[0.2770]	[0.3208]
	LARGE SCHOOLS		
	Classmates	-0.6903	-0.2588
		[2.0482]	[1.0989]
	Coworkers	-0.1845	-0.1879
		[0.4990]	[0.5111]
B.	SMALL FIRMS		
	Classmates	1.1873**	1.1501**
		[0.4534]	[0.3848]
	Coworkers	-0.2906	-0.2698
		[0.2534]	[0.2736]
	LARGE FIRMS		
	Classmates	0.5475	0.1644
		[0.7519]	[1.1447]
	Coworkers	-1.4445	2.1514
		[2.1226]	[4.1259]

Notes: Specifications mirror the baseline specification described in Table 2.3 and have the same first stage estimates, but do not include *comuna* fixed effects and, instead, includes predetermined group characteristics. *N=68,589* in both panels. There are only two regressions in this table and it is constructed as follows. The first step is to estimate a first stage using the sample of 68,589 similar to the one reported in Table 2.3. Then fitted values for s_net and l_net are predicted using this first stage. Those fitted values and each individual and group controls are then interacted with two dummy variables, D_i and its complement $D_i^c = D_i - 1$. For example, $D_i = 1$ if the individual i belongs to a small schools. Then the small firms panel show coefficients that accompany the D_i dummy, while the large firms show coefficients that accompany the D_i^c dummy. This is the reason why the number of observations remains at 68,589.

Table 2.8: Other subsamples

	First Stage			Reduced	Second	
	Classmates	Coworkers	*F-AP*	Form	Stage	N
	(1)	(2)	(3)	(4)	(5)	(6)
A. UNDER 21 YEARS OLD						*13,886*
Classmates	0.0261**	-0.0013	*10.1*	0.0287*	1.0512**	
	[0.0084]	[0.0023]		[0.0173]	[0.5339]	
Coworkers	0.0051	0.0207***	*12.9*	0.0035	0.2367	
	[0.0044]	[0.0057]		[0.0104]	[0.4582]	
OVER 21 YEARS OLD						
Classmates	0.0128**	0.0009	*10.2*	0.0118*	0.8740**	*54,465*
	[0.0040]	[0.0009]		[0.0067]	[0.4420]	
Coworkers	-0.0013	0.0171***	*19.0*	-0.0073	-0.4729	
	[0.0022]	[0.0039]		[0.0048]	[0.3554]	
B. TECNICO PROFESIONAL						*48,136*
Classmates	0.0176***	0.0004	*11.8*	0.0215**	1.2591***	
	[0.0051]	[0.0010]		[0.0088]	[0.3808]	
Coworkers	0.0021	0.0193***	*21.5*	-0.0061	-0.3417	
	[0.0027]	[0.0041]		[0.0054]	[0.3176]	
HUMANISTA CIENTIFICO						*20,215*
Classmates	0.0120**	0.0011	*5.6*	0.0073	0.5635	
	[0.0052]	[0.0019]		[0.0092]	[0.6949]	
Coworkers	-0.0027	0.0139**	*7.6*	-0.0022	-0.199	
	[0.0027]	[0.0050]		[0.0079]	[0.5956]	
C. WAGE $\leq$ CLP$200,000						*26,730*
Classmates	0.0215***	-0.0012	*20.7*	0.0228**	1.0390**	
	[0.0046]	[0.0015]		[0.0094]	[0.4126]	
Coworkers	-0.0007	0.0143**	*7.0*	-0.0097	-0.5924	
	[0.0031]	[0.0053]		[0.0071]	[0.6290]	
WAGE $>$ CLP$200,000						*41,621*
Classmates	0.0116**	0.0016	*8.0*	0.0111	0.9699*	
	[0.0041]	[0.0011]		[0.0076]	[0.5703]	
Coworkers	0.0005	0.0199***	*22.8*	-0.0025	-0.2055	
	[0.0023]	[0.0041]		[0.0056]	[0.3136]	

Notes: $N = 68,589$. Specifications mirror the baseline specification described in Table 2.7. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.9: Peer effects in networks with weaker ties

		OLS	First Stage		Reduced Form	Second Stage (2SLS)	*N*
			Classmates	Coworkers			
		(1)	(2)	(3)	(4)	(5)	(6)
A	BASELINE /a						*68,589*
	Classmates	0.1292***	0.0184***	0.0003	0.0167***	0.9113***	
		[0.0224]	[0.0033]	[0.0020]	[0.0063]	[0.2718]	
	Coworkers	0.3281***	0.0002	0.0201***	-0.0048	-0.2521	
		[0.0189]	[0.0007]	[0.0039]	[0.0049]	[0.2622]	
	F-statistic (AP)		*32.2*	*26.0*			
B	SAME SCHOOL, INDEPENDENT OF GRADUATION YEAR /b						*68,589*
	Classmates	0.1895***	0.0382***	0.0019	0.004	0.1173	
		[0.0381]	[0.0102]	[0.0062]	[0.0169]	[0.4395]	
	Coworkers	0.3273***	0.0003	0.0201***	-0.0049	-0.2433	
		[0.0146]	[0.0004]	[0.0039]	[0.0041]	[0.2104]	
	F-statistic (AP)		*14*	*90.5*			
C	SAME FIRM, INDEPENDENT OF YEAR /c						*68,589*
	Classmates	0.1263***	0.0184***	0.0009	0.0167***	0.8996***	
		[0.0210]	[0.0033]	[0.0017]	[0.0057]	[0.3139]	
	Coworkers	0.4054***	0.001	0.0247***	0.0041	0.1274	
		[0.0211]	[0.0009]	[0.0046]	[0.0054]	[0.2018]	
	F-statistic (AP)		*145.5*	*29.0*			

Notes: Specifications mirror those in Table 2.3, but changing the definition of networks. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 2.10: Other dates

	OLS	First Stage		Reduced	Second	N
		Classmates	Coworkers	Form	Stage (2SLS)	
	(1)	(2)	(3)	(4)	(5)	(6)
			SEPTEMBER 2009			
Classmates	0.1172***	0.0218***	0.0012	0.0173**	0.8072**	*47,373*
	[0.0270]	[0.0037]	[0.0025]	[0.0079]	[0.3314]	
Coworkers	0.3985***	-0.0002	0.0188***	-0.0049	-0.2545	
	[0.0252]	[0.0009]	[0.0046]	[0.0063]	[0.3782]	
F statistic (AP)		*36.1*	*15.9*			
			MARCH 2010			
Classmates	0.0545*	0.0129***	0.0052**	-0.0097	-0.7646	*47,373*
	[0.0320]	[0.0035]	[0.0025]	[0.0077]	[0.7178]	
Coworkers	0.3254***	0.0011	0.0265***	0.0000	0.0324	
	[0.0214]	[0.0009]	[0.0052]	[0.0057]	[0.2129]	
F statistic (AP)		*12.6*	*25.7*			

Bibliography

Abulkadiroglu, A., Angrist, J., and Pathak, P. (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, 82(1):137–196.

Acemoglu, D., Dahleh, M. a., Lobel, I., and Ozdaglar, A. (2011). Bayesian learning in social networks. *Review of Economic Studies*, 78(4):1201–1236.

Aizer, A. and Currie, J. (2004). Networks or neighborhoods? Correlations in the use of publicly-funded maternity care in California. *Journal of Public Economics*, 88(12):2573–2585.

Anderson, P. M. and Meyer, B. D. (2012). Unemployment Insurance Takeup Rates and the After-Tax Value of Benefits. *The Quarterly Journal of Economics*, 112(3):913–937.

Angrist, J. and Imbens, G. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30(October):98–108.

Bala, V. and Goyal, S. (1998). Learning from Neighbours. *Review of Economic Studies*, 65(3):595–621.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. Wanted: The key player. *Econometrica*, 74(5):1403–1417.

Bertrand, M., Luttmer, E. F. P., and Mullainathan, S. (2000). Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115(3):1019–1055.

Blasco, S. and Fontaine, F. (2009). A structural model of the unemployment insurance take-up.

Borjas, G. J. and Hilton, L. (1996). Immigration and the Welfare State: Immigrant Participation in Means-Tested Entitlement Programs. *The Quarterly Journal of Economics*, 111(2):575–604.

Box, P. O., Haven, N., and Boozer, M. a. (2001). Inside the 'Black Box' of Project STAR: Estimation of Peer Effects Using Experimental Data.

Bravo, D. and Rau, T. (2012). Effects of Large-scale Youth Employment Subsidies: Evidence from a Regression Discontinuity Design.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust Inference With Multiway Clustering.

Carneiro, P., Galasso, E., and Ginja, R. (2013). Tackling social exclusion: Evidence from Chile.

Cingano, F. and Rosolia, A. (2012). People I Know: Job Search and Social Networks. *Journal of Labor Economics*, 30(2):291–332.

Currie, J. M. (2004). The Take Up of Social Benefits.

Dahl, G. B., Lø ken, K. V., and Mogstad, M. (2012). Peer Effects in Program Participation.

Duflo, E. and Saez, E. (2002). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly Journal of Economics*, 118(3):815–842.

Dustmann, C., Glitz, A., and Schönberg, U. (2012). Referral-based Job Search Networks.

Ellison, G. and Fudenberg, D. (1995). Word-of-mouth communication and social learning. *Quarterly Journal of Economics*, 110(1):93–125.

Final, I. (2014). Estudio Quinta Encuesta Sobre Acceso, Usos, Usuarios y Disposición de Pago por Internet en Zonas Uranas y Rurales de Chile.

Geroski, P. (2000). Models of technology diffusion. *Research Policy*, 29(4-5):603–625.

Glitz, A. (2013). Coworker Networks in the Labour Market.

Gloria, M., García, A., Del, J., and Anabalón, R. (2005). Registro de los Estudiantes de Chile : un sistema de información masiva en extranet. In *VII Simpósio Internacional de Informática Educativa SIIE05*, pages 369–374.

Golub, B. and Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *Quarterly Journal of Economics*, 127(3):1287–1338.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1):201–209.

Hellerstein, J. K., McInerney, M., and Neumark, D. (2011). Neighbors and Coworkers: The Importance of Residential Labor Market Networks.

Huneeus, C. (2010). Balance de los avances y desafíos de las políticas de empleo para jóvenes. Technical report.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.

Jadbabaie, A., Molavi, P., Sandroni, A., and Tahbaz-Salehi, A. (2012). Non-Bayesian social learning. *Games and Economic Behavior*, 76(1):210–225.

Kleven, H. J. (2008). Transfer Program Complexity and the Take Up of Social Benefits. *American Economic Journal: Economic Policy*, 3(1):54–90.

Lee, D. S. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2):281–355.

Manski, C. F. (1993). Identification of Social Endogenous Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531–542.

Moffitt, R. (1983). An economic model of welfare stigma. *The American Economic Review*, 73(5):1023–1035.

Moffitt, R. A. (2001). Policy Interventions, Low-Level Equilibria, and Social Interactions. In *Social Dynamics*, pages 45–82.

Newey, W. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics*, 36(3):231–250.

O'Donoghue, T. and Rabin, M. (1999). Doing It Now or Later. *The American Economic Review*, 89(1):103–124.

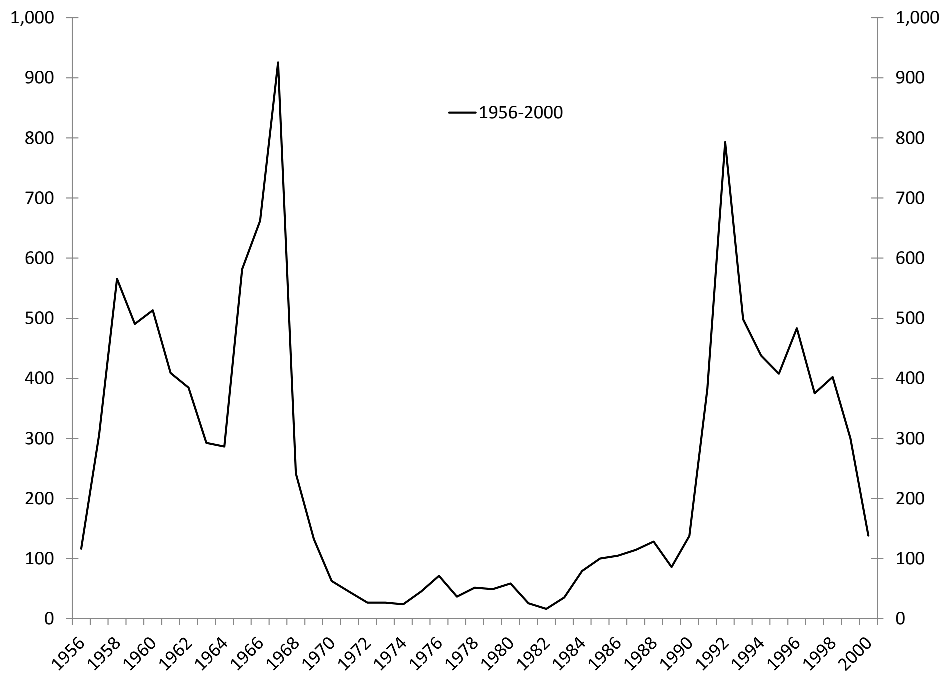
Riphahn, R. (2001). Rational poverty or poor rationality? The take-up of social assistance benefits. *Review of income and wealth*, 1(3):379–398.

Zimmerman, S. D. (2015). Making Top Managers : The Role of Elite Universities and Elite Peers.

Appendix A

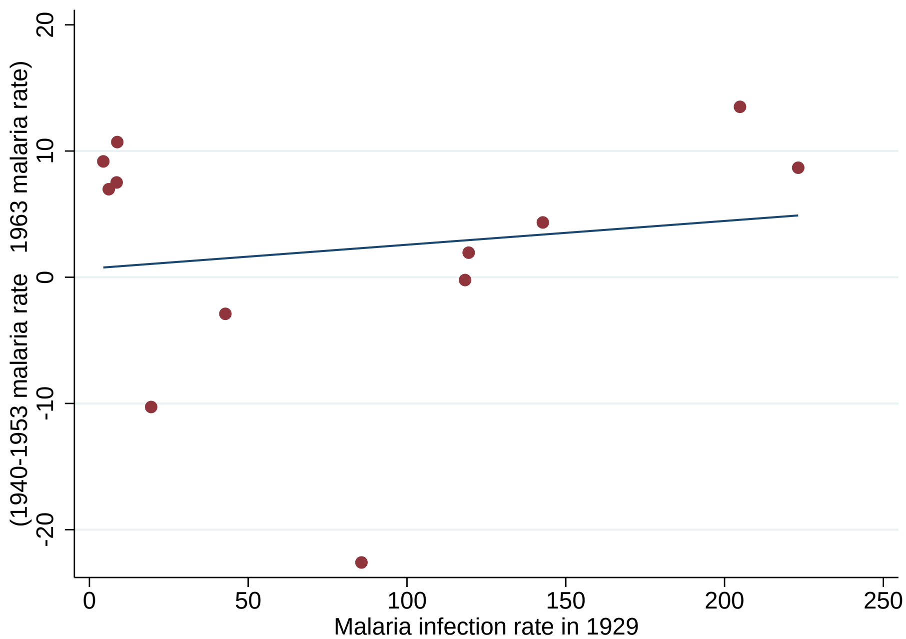
Appendix to Chapter 1

Figure A.1: Evolution of the malaria rate per 10,000 habitants in Costa Rica between 1956 and 2000.



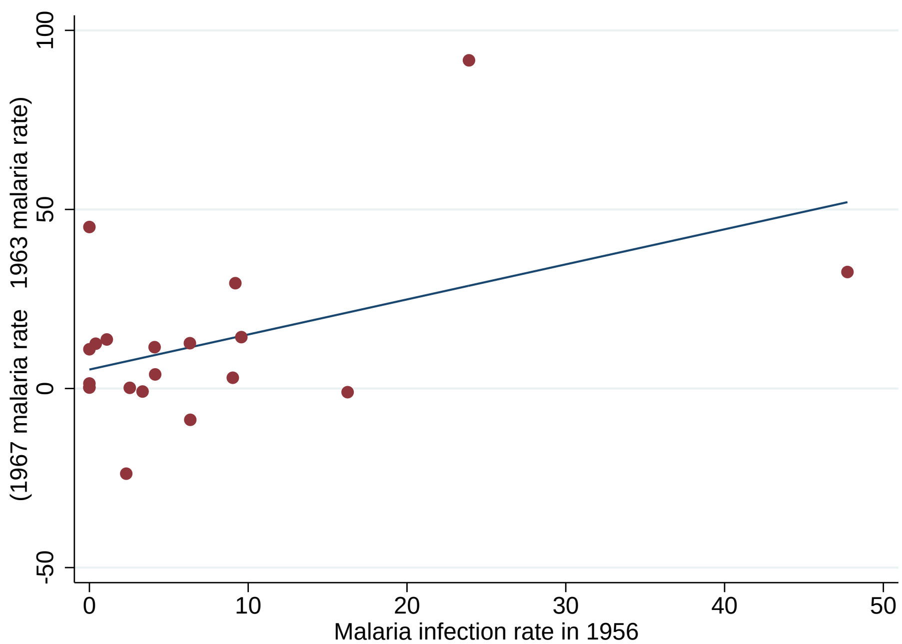
Source: author's calculations based on WHO.

Figure A.2: Figure 5a excluding outliers.



Notes: see Figure 5a. β =0.02, t =0.47, R^2 =0.02

Figure A.3: Figure 5b excluding outliers



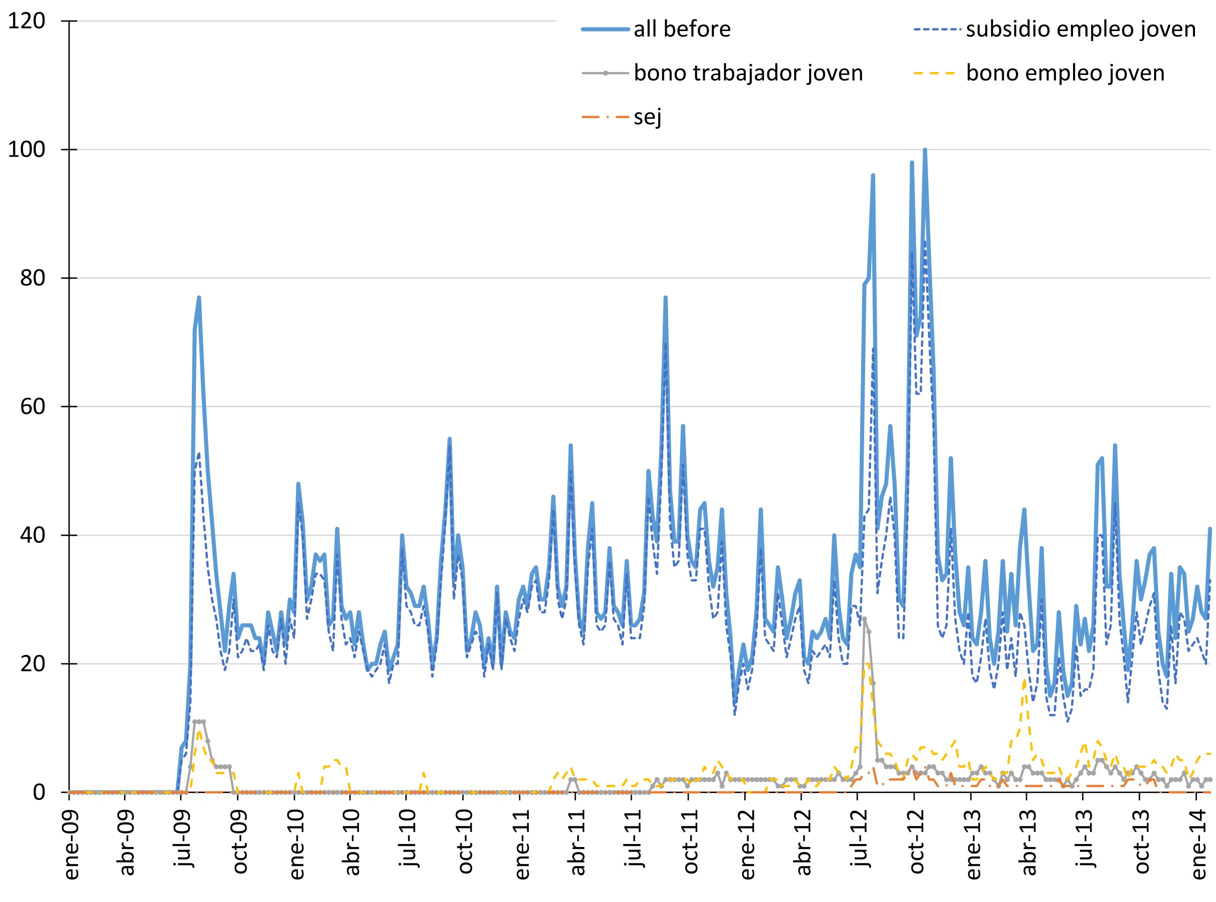
Notes: see Figure 5b. β =0.98, t =2.22, R^2 = 0.22

Appendix B

Appendix to Chapter 2

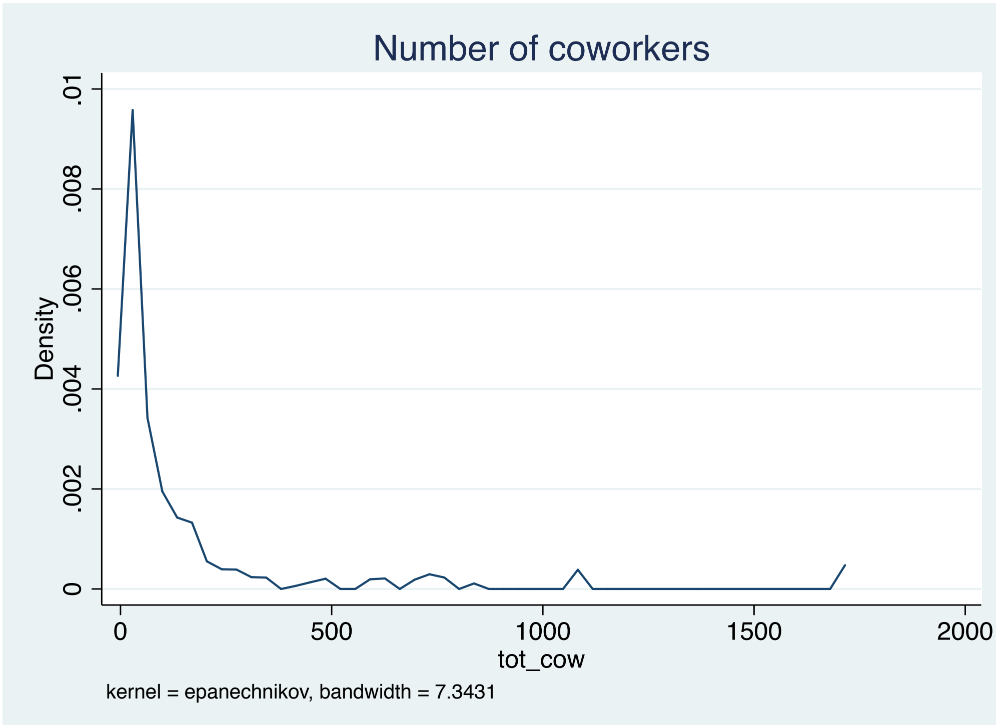
B.1

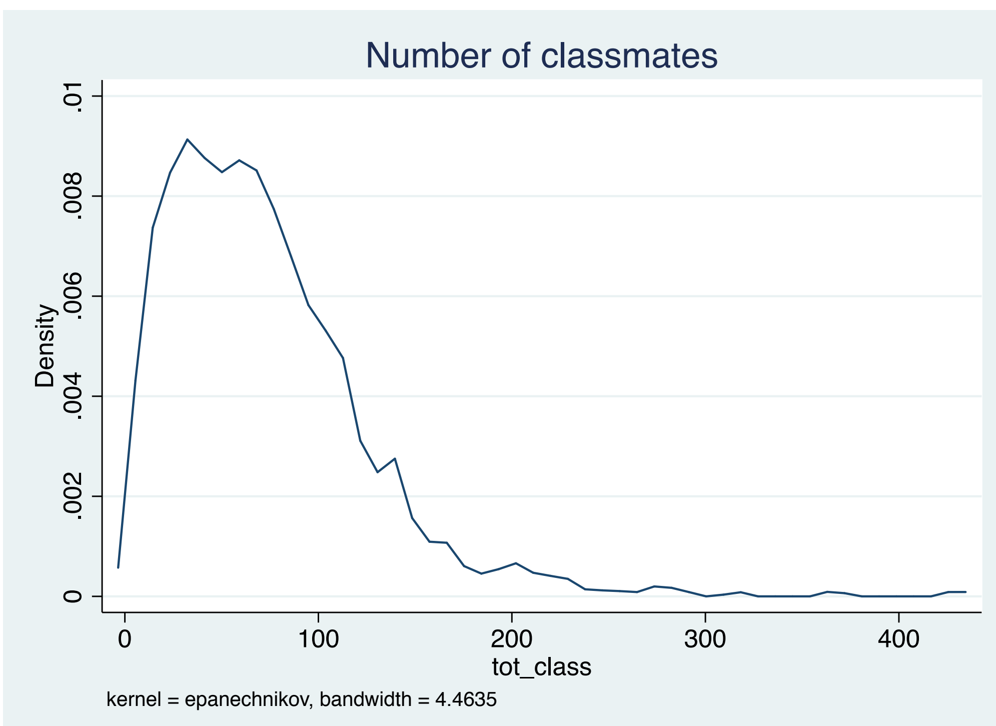
Figure B.1: Google searches of words related to SEJ, Google Trends, by date.



Notes: "all before" refers to an index that combines the four topics below. Data were taken from Google Trends.

B.2 Kernel Density Estimates and Descriptive Statistics





These figures show Kernel Density Estimates using the estimating sample of 68,589

Table B.1: Descriptive Statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
Number of coworkers	68,589	136.0	281.0	1	1,708
Number of classmates	68,589	72.5	51.7	1	430

Notes: This table shows descriptive statistics for the number of coworkers and classmates using the 68,589 estimating sample.

B.3 Appendix C

Assume that the endogenous peer effect model $y_{ij} = \bar{y}_{-i,j}\beta + \epsilon_{ij}$ is instrumented with a variable z_j which is constructed using the full sample as $z_j = \frac{1}{N_j}S_j$, where $S_j \equiv \sum_j d_{ij}$ where $d_{ij} = \mathbf{1}\{FPS_i \leq 11734\}$ and assume that $N_j = N$, hence $z_j = \frac{1}{N}S_j$. The instrumental variables estimator for β in a sample of NJ youths in J groups is:

$$\hat{\beta} = \frac{\sum_j \sum_i S_j y_{ij}}{\sum_j \sum_i S_j \bar{y}_{-i,j}} \tag{B.1}$$

Boozer and Cacciola (2001) shows that this expression is equal to 1 in the absence of other covariates. The reason is that $\bar{y}_{-i,j}$ in equation B.1 can be rewritten as $N\bar{y}_j - y_{ij}$ and then

$$\hat{\beta} = \frac{\sum_j \sum_i S_j y_{ij}}{\sum_j \sum_i S_j [\frac{1}{N-1}(N\bar{y}_j - y_{ij})]} \tag{B.2}$$

Notice that the S_j is not affected by the sum over the i subscripts, and the only terms affected are the y_{ij}, then

$$\hat{\beta} = \frac{\sum_j S_j \bar{y}_j}{\sum_j S_j [\frac{1}{N-1}(N\bar{y}_j - \bar{y}_j)]} \tag{B.3}$$

"This expression is easily seen to equal 1..." (pg. 46). However, in the case of this paper, the instrumental variable z_j is actually $z_{-i,j}$ because it is constructed in a "leave-out" way using only a small subgroup whose FPS lies within a small window around the $x_0 = 11,734$ cut off. Define $left_{i,j} = \mathbf{1}\{FPS_{ij} \in [x_0 - \Delta, x_0]\}$ and $inside_{i,j} = \mathbf{1}\{FPS_{ij} \in [x_0 - \Delta, x_0 + \Delta]\}$, hence in this paper:

$$z_{-i,j} = \frac{\sum_k left_{kj} - left_{ij}}{\sum_k inside_{kj} - inside_{ij}} \tag{B.4}$$

Notice that this instrument takes zero or some positive value only within networks with at least one person with $FPS_{ij} \in [x_0 - \Delta, x_0 + \Delta]$, otherwise it takes missing values. Also notice that now the $z_{-i,j}$ is affected by the sum over the i subscripts in equation B.2, so the sums cannot be carried out through, and this implies that the coefficient $\hat{\beta} \neq 1$. As a result,

$$\hat{\beta} = \frac{\sum_j \sum_i z_{-i,j} y_{ij}}{\sum_j \sum_i S_j \bar{y}_{-i,j}} \neq 1 \tag{B.5}$$

B.4 Algorith Used

This section explains the algorithm used in comparing the fraction of people with SEJ under the randomized advertising versus the fraction under the targeted advertising campaigns.

1. Calculate the new 2SLS regression model and save the coefficients. This econometric model uses a sample of 89,440 persons.
2. Randomize the 47,047 persons who got SEJ among the admissible ones.
3. Calculate the new fraction of peers with SEJ inside each of the networks s_net and l_net.
4. If t = 0,
   a. Calculate the fraction of people with SEJ among the admissible people.
5. While 0<t<T,
   a. Use the regression model calculated in Step 1 to predict the new individual probability of applying to SEJ. This econometric model can be applied to only 129,563 persons.
   b. If the probability is greater than a threshold p* and she did not previously have SEJ before, then replace Y_{isl}=1.
   c. Recalculate the new fraction of peers with SEJ inside each of the networks s_net and l_net.
   d. Calculate the fraction of people with SEJ among the admissible people.

Repeat n times steps 3-6.