The London School of Economics and Political Science

Essays on the impact evaluation of social programs and public sector reforms

Tatiana Valeria Paredes-Torres

A thesis submitted to the Department of Social Policy of the London School of Economics for the degree of Doctor of Philosophy

London, January 2019

DECLARATION

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 62,856 words.

ABSTRACT

This thesis contains three essays on the impact evaluation of social programs and public sector reforms.

Chapter 2 investigates whether the effects of a cash transfer program persist or wear off in the long-run. I study the first two phases of Bono de Desarrollo Humano (BDH) in Ecuador, each of which lasted about five years. My identification strategy uses a regression discontinuity design and relies on the fact that at the threshold of eligibility, the second assignment to treatment (in 2008/9) was independent of the first assignment (in 2003). This allows me to disentangle the impact of a short exposure to the program (treatment during one phase) from a long exposure (treatment during two phases). Most of the gains in enrollment and schooling were achieved in the short-run among children that started treatment when they were about to start elementary school, eleventh grade or Baccalaureate. However, an extended exposure to BDH was not enough to keep raising children's education. Regarding labor market outcomes, BDH had a negative (not statistically significant) impact on the probability of working among young children but did not increase job opportunities among young adults in the long-run.

Chapter 3 evaluates the impact on in-hospital mortality of a reform that made all health professionals working part-time switch to full-time contracts at public hospitals in Ecuador. I take advantage of the staggered adoption of the reform and hospital panel data to implement an event study to evaluate the impact of the reform. The results for the sample of admissions to the emergency department show that mortality in public hospitals decreased by 0.1% on the adoption year and by 0.2% one year later. Results were robust to several alternative specifications and to the inclusion of pre-reform characteristics that could have been used by policymakers

to decide the order of implementation. More importantly, I show that the effects reported in this paper cannot be attributed to changes in other quality indicators at the hospital level like the length of stay or by changes in the patient mix.

Chapter 4 studies the impact of increased liability risk facing physicians on the use of c-sections and on indicators of maternal and infant health. I take advantage of a legal reform that led to the hardening of sentences for cases of professional malpractice in Ecuador. I use a difference-in-difference strategy that compares the outcomes of two neighboring countries, Ecuador and Colombia, before and after the introduction of the reform and perform several parallel trend tests on the outcomes of interest and test for the stability of the demographic composition of both countries to support my identification strategy. During the five quarters following the reform, Ecuadorian doctors reduced the c-section rate by 1.1% among women aged 15 to 24 years, and by 0.9% among women aged 25 to 34 years. The c-section rate remained unaffected for women aged 35 to 44 years, possibly because doctors have less discretion over riskier births. Interestingly, the observed reduction in the c-section rate did not affect the health outcomes of mothers or newborns.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Alistair McGuire and Stephen Jenkins for their support and continued guidance. Our conversations have taught me innumerable lessons about the different aspects of academic life that I am sure will accompany me throughout life.

Thanks also to Javier Brugués for all the love, encouragement and insightful comments to this thesis. To my dad for his devotion and patience in helping me request some of the data sets I needed. To my mum and sisters for their love and support.

I am grateful to Frank Lichtenberg for hosting me during my visit to Columbia University, for fruitful discussions and invaluable feedback on the second and third papers of this thesis. Special thanks to Miguel Urquiola, Eric Verhoogen, Rodrigo Soares, Steve Pischke, Bentley MacLeod, Michael Best, Almudena Sevilla and Yanos Zylberberg for their generosity and invaluable feedback.

I thank my friends and colleagues at the LSE and beyond who have provided helpful comments and suggestions at various points in time. In particular thanks to Wilson Guzmán, Nick Mathers, Diego Alburez, Cristian Crespo, Belén Sáenz, Julia Phillips, Celine Zipfel, Ellie Suh, Ewa Batyra, Milo Vandemoortele, Reinaldo Cervantes, Vladimir Proaño, Melody Serrano and Carlos Neptali López Castillo.

I could not have completed this thesis without the generous financial support of the Secretariat of Higher Education, Science, Technology and Innovation (SENESCYT). I also gratefully acknowledge the support provided by the Marshall Institute and the Academic Partnerships Office at LSE.

Finally, I would like to thank the Ministry of Social Development for access to Social Registry data and payments data on Bono de Desallorro Humano; and, the Secretariat of Higher Education, Science, Technology and Innovation for access to data on the ENES exams. I also would like to thank the Ministry of Finance for the data on hospital budgets and the Ministry of Health for access to several documents related to the reforms I study in this thesis.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

In developing countries, where social insurance covers a small fraction of the population, labor market regulations are thin and poorly enforced, and the incidence of poverty is high, the focus of social protection systems is on social assistance programs like non-contributory pensions and cash transfer programs (Barrientos, 2011). Latin American countries are at the forefront of developing countries' efforts to build stronger social protection systems (Naranjo, 2014). These countries not only are pioneers in the use of cash transfer programs, with the first conditional cash transfer programs being implemented in Mexico and Brazil by the end of the 90s. They have also been very active with the implementation of other initiatives including subsidized social security pensions for retired workers (where the government contributes with part of the cost and the other part is paid by the employee) and in the expansion of health protection (Naranjo, 2014).

Cash transfer programs are one of the most powerful tools in the fight against poverty and inequality in Latin America. By 2011 these programs had already spread across 18 countries in the region, covering close to 129 million beneficiaries (Stampini and Tornarolli, 2012). Many of the conditional cash transfer programs in the region have pursued an integrated approach to poverty reduction, balancing goals of social assistance and human capital formation. For this, they target women with children and condition the receipt of benefits to school enrollment and regular health checkups. Some examples include the conditional cash transfer programs in Mexico, Brazil, Colombia, El Salvador, Jamaica and Panama (Fiszbein et al., 2009).

Cash transfer programs vary widely in terms of the overall objectives they pursue and in terms of coverage. In Latin America, some programs put more emphasis on human capital formation while others focus more on income redistribution. There are also programs like Chile Solidario, were families commit to household-specific action plans as a precondition for receiving the benefits. In other regions of the globe these programs have been used to reduce gender disparities in education or to help alleviate the consequences of epidemics and natural dissasters. In general, the versatility of these programs explains why they have become so popular worldwide and why the interest in conditional cash transfers has spread from developing to developed countries — most recently to programs in New York City and Washington, DC. (Fiszbein et al., 2009).

In terms of coverage, most Latin American programs, like Mexico's Oportunidades and Chile Solidario, target only households living in extreme poverty, while other countries like Ecuador have a more inclusive approach, targeting a broader share of the population. More precisely, relative coverage ranges from approximately 40 percent of the population in Ecuador to about 20 percent in Brazil and Mexico and 5 percent in Chile (Fiszbein et al., 2009). The economic importance of these programs is such that in 13 countries of the region, by 2010, they represented over 20% of poor beneficiaries' incomes and studies suggest that the poverty headcount index in those countries would have been on average 13% higher, had conditional cash transfers not been implemented (Stampini and Tornarolli, 2012).

Cash transfers have become central to social protection systems by helping vulnerable households weather shocks while incentivizing human capital accumulation since the early stages of child development. A long literature reports that cash transfers have been effective in increasing enrollment and cognitive development among treated children; and in reducing child labor (Edmonds and Schady, 2012; Schady et al., 2008; Behrman, Parker, et al., 2011; Behrman, Sengupta, et al., 2005; Kugler and Rojas, 2018). The success of the basic model is prompting countries to address second and third rounds of challenges with complementary actions to ensure that conditional cash transfers have a greater impact. Some of them include expanding access to health care and education and exploring possible ways to make benefits conditional on performance and not conditional on mere service use (Fiszbein et al., 2009).

In the last two decades, cash transfers have become one of the main tools against poverty. However, inequality of income and opportunities remain high, in particular with regards to access to public services like health care and education (OECD, 2017).[1] Health systems in Latin America remain highly fragmented and unable to cover the most vulnerable population. There is also a high portion of the middle-class that does not have access to contributory social insurance programs (44% and 34% of those in the third and fourth quintile, respectively). It is precisely the low rates of social insurance coverage and the difficulty to reach out to informal workers, that led to the expansion of non-contributory programs (cash transfers) and the development of health-assistance programs for those excluded from the contributory systems (OECD, 2017).

Health systems in Latin America were inspired by the Beveridge-type model of a National Health Service based on universal access and funded from tax revenues, yet, it was never fully implemented (Giovanella et al., 2012). The 'debt crisis' during the 80s and 90s, intensified the problems. The structural adjustment policies imposed by international financial institutions prompted cuts in public spending and led some countries to set up partial health insurances with minimum services to cover only specific groups such as pregnant women, infants, and the extremely poor (Giovanella et al., 2012). As a result of these and some more recent reforms, the health systems in Latin America consist of a private sector block and two separate public sector blocks: a well resourced social security health system for salaried

[1] The average Gini coefficient went from 0.51 in the early 1990s to 0.42 in 2015 but relative poverty (i.e., households with incomes less than half of the median) decreased by only 4 percentage points between 2000 and 2014.

workers and their families and a Ministry of Health serving poor and vulnerable people with low quality standards (Cotlear et al., 2015).

During the 2000's or the so called period of economic prosperity in Latin America, several countries implemented health reforms focused on improving access and quality of health services in the public sector. For some countries of the region this also meant dealing with old problems related to the difficulties to attract health professionals, insufficient training, and a tendency to specialization among physcians at the expense of the large number of professionals required in the first level of care, all of which constitute major constraints to achieving the goal of universal health coverage (Maeda et al., 2014). Some of the reforms implemented in this regard include training programs for community doctors in Argentina and Venezuela, the creation of a specific secretariat responsible for the human resources management and training in Brazil (Giovanella et al., 2012), and a reform in Ecuador that made all health professionals working part-time switch to full-time contracts at public hospitals in Ecuador, which aimed to increase the available doctor-hours at emergency departments.

In this thesis, I study three different government interventions that took place in Ecuador in the era of economic recovery that followed the dollarization in January 2000. Ecuador is the eight largest economy in Latin America[2] with a developing economy that is highly dependent on commodities like petroleum and agricultural products. After the dollarization of the economy, in January 2000, Ecuador experienced inclusive growth fueled by a favorable external environment and by the increase in oil prices (Rubalcaba et al., 2017). Studying Ecuador during the era of economic recovery is interesting for several reasons. First, Ecuador is one of the ten countries in Latin America that between 2004 and 2009 underwent a profound transformation with left and center-left presidential candidates winning electoral

[2]Source: CIA World Factbook

contests (Pribble, 2013).[3] With some variations, these governments introduced several universalistic social policies aimed at guaranteeing coverage for a set of essential social services and ensuring a basic minimum income for all its citizens (Pribble, 2013).[4]

By 2011, Ecuador was the only country in the region that assigned more than 1% of its GDP to cash transfers, while other countries like Bolivia, Chile, Costa Rica, Panama, Paraguay, Peru and Uruguay only assigned between 0.09% and 0.35% of their GDP (Amarante and Brun, 2016). Stampini and Tornarolli (2012) compare the performance on poverty reduction of several cash transfers programs in various Latin American countries over the period 2004-2010. They found that Ecuador experienced the largest reduction in the poverty headcount index in absolute terms – a reduction of 3.3 percentage points –, followed by Brazil and Mexico with a reduction of 1.7 percentage points respectively, while in Chile, Costa Rica, Jamaica, Panama, and Paraguay, cash transfers reduced poverty by less than one percentage point. Amarante and Brun (2016) reached similar conclusions for the period between 2011 to 2013, when Ecuador exhibited the highest decrease in the incidence of poverty as a consequence of transfers – 2.4 percentage points –, followed by Uruguay with 1 percentage point.

As a result of a favorable external environment, Ecuador's gross national income-per-capita rose from 1,540 dollars in 2000 to 6,130 dollars in 2014;[5] and, the poverty rate decreased significantly between 2000 and 2014 from 64.4% to 21.7% in 2014.[6]. However, income inequality remained high. Even though the Gini index

[3]Argentina (2003, 2007, and 2011), Brazil (2002, 2006, and 2010), Bolivia (2005 and 2009), Chile (2000 and 2006), Ecuador (2006 and 2009), El Salvador (2009), Nicaragua (2006),Paraguay (2008), Uruguay (2004 and 2010), and Venezuela (1998, 2000, 2006, and 2012)

[4]Universalistic policies provide income transfers and public services of a similar size and quality for all citizens. Importantly, access to programs must be transparent, making sure that benefits are allocated on the basis of legal criteria (Pribble, 2013)

[5]Source: World Bank national accounts data, and OECD National Accounts data files. The figures are in current US dollars.

[6]Source: World Bank Indicators. Poverty is measured as the headcount ratio at the national

decreased from 56.4 in 2000 to 45 in 2014[7], by 2012, more than 50% of total income was concentrated by households in the richest quintile, while 26% was distributed among the 60% poorest households and 22% among households in the fourth income quintile (Naranjo, 2014). Inequality in access to public services is also a big concern in Ecuador. By 2004, almost a third of Ecuador's population lacked regular access to health services, while more than two-thirds had no health insurance and insufficient resources to pay for health care services (López-Cevallos and Chi, 2010). Whereas regarding access to education, children in the top income quintile achieve around 6 more years of education when compared to children of the same age in the lowest quintile (Cruces et al., 2014).

Economic prosperity paved the way for several reforms to the health sector and to the existing legislation seeking to expand access to public services. In 2008, Ecuador reformed its constitution and included access to health care as a right of all citizens and a responsibility of the state (Giovanella et al., 2012).[8] The new constitution also declared that public education should be free from the first year of basic education up to the undergraduate level. What followed was a marked increase in public spending during the 2000's. The budget for education grew significantly in the 2000-2012 period, from 3.4% of GDP in 2000 to 4.8% of GDP in 2012. Similarly, the health care budget increased from 0.8% of GDP in 2000 to 2.1% of GDP in 2012 (Naranjo, 2014). This growth in spending has been accompanied by an important improvement in some education and health indicators. For instance, the net enrollment rate for baccalaureate grew from 42.1% in 2003 to 65.1% in 2014, while the maternal mortality ratio decreased from 65.45 per one thousand inhabitants in 2000 to 49.16 per one thousand inhabitants in 2014; and, the infant

poverty line of 1.90 dollars

[7]Source: World Bank Indicators

[8]However, despite these reforms, according to specialists in the healthcare sector, Ecuador has never undergone a real reform process involving a deep and sustained change in the sector's fragmented structure (Naranjo, 2014).

mortality rate decreased from 15.46% in 2000 to 8.35% in 2014.[9] Yet, there is no rigorous evidence suggesting that the increase in spending generated commensurate improvements in health and education.

This thesis is composed of three independent papers (chapters 2, 3 and 4); each paper draws on methods in economics to evaluate three different interventions/reforms that happened during the period of economic prosperity in Ecuador. In each paper, I use quasi-experimental methods where institutions generate situations close to random assignment, to recover the causal effects of interest (Angrist and Krueger, 1999).

In Chapter 2, I investigate the short- and long-term effects of Bono de Desarrollo Humano (BDH) on young people's education and labor market outcomes. BDH is the main cash transfer program in Ecuador. It was initially advertised as a conditional cash transfer program requiring mothers to take children to health checkups and to enroll them at school. However, soon after its implementation, it became an unconditional transfer due to lack of administrative capacity to monitor the conditions. Several short-term evaluations show that BDH has had positive effects in enrollment and cognitive and socio-emotional development among treated children; as well as negative effects on child work (Edmonds and Schady, 2012; Schady et al., 2008). However, until recently there were no studies on the long-term effects of this program and much less evidence on the persistence of its short-term effects in the long-run.

By 2014, the program had two phases, each lasting 5 years. Hence, people who were eligible during the two phases of the program received the transfer for a maximum of 10 years. Unlike the previous literature that looks at how well the original treatment and control groups performed after several years of a program's implementation, the main question I address is whether children who were treated for a longer period of time continued achieving better outcomes compared to similar children who were

[9] Source: Compendio Estadístico del 2014, INEC.

treated for a shorter period of time. In other words, I study whether treatment continued being effective after several years of targeting the same population.

I build a unique data set that uses Ecuadorian Social Registry micro-data linked to BDH payments data that allows me to track the treatment status of children and their academic and labor market trajectories.[10] The main challenge was to construct a panel at the individual-level linking the three waves of the Ecuadorian Social Registry using probabilistic record linkage. I describe the main details of the process in the corresponding data section in Chapter 2, however, more details regarding attrition and sample balancing can be found in Appendix A.

My identification strategy relies on the fact that at the threshold of eligibility, the second assignment to treatment (in 2008/9) was independent of the first assignment (in 2003). This allows me to disentangle the impact of a long exposure to BDH (treatment during phases one and two) versus a short exposure (treatment during phase one only) by comparing the outcomes observed at the end of phase two of children who were treated during phase one and were also marginally treated (or not) during phase two. The results show that most of the gains in enrollment and schooling were achieved by children that were treated during one phase of the program and started treatment at juncture ages (5, 12 and 15 years). However, an extended exposure to BDH was not enough to keep raising children's education.

Apart from being one of the first studies to investigate the persistence of the effects of cash transfers in the long-run, the methodology proposed in this paper constitutes a tool that other countries could use to evaluate programs that have similar characteristics to BDH. In particular, programs that use Social Registry data and a secret targeting rule to select beneficiaries, which changes from time to time when

[10] Social Registries are information systems that support registration, and determination of potential eligibility for social programs (Leite et al., 2017). In their comparative study, Leite et al. (2017) analyze a sample of 32 countries that use social registries to illustrate their extensive use in social policy

governments update the information from the Social Registry or decide to change the targeting rule. This would help to create accountability and to avoid having ineffective and expensive cash transfer programs that are not reformed for political reasons.

Chapter 3 moves to the analysis of health care reforms in Ecuador. Several studies in developed countries suggest that there is a trend of declining working hours of doctors (Swami et al., 2017; Broadway et al., 2017).[11] In response to this phenomenon and a growing demand for healthcare services, several countries have implemented policies that operate in the intensive margin to increase the working hours of doctors and consequently expand health coverage (Dolton and Pathania, 2016; Swami et al., 2017). However, evidence from stated preference studies suggests that if doctors are given a choice, they would not choose to work more hours (Broadway et al., 2017). This paper studies for the first time a reform that replaced all part-time contracts in public health establishments and banned the creation of new part-time and occasional contracts, forcing doctors to work at least 8 hours a day or 40 hours per week. The reform – that started in 2009–, increased the number of doctor-hours at public hospitals. Hence the question I address is whether it had a positive effect on hospitals' quality as measured by in-hospital mortality.

I use information from hospital resources data, patient discharges and budget data to build a balanced panel of public hospitals from 2004 until 2014. I use the information on the number of health professionals with part-time contracts from hospital resources data and information from a series of resolutions issued by the Ministry of Labor that established particular adoption dates for each of the different groups of public hospitals, to identify the year of adoption for every hospital in the panel. For this, I identified the year of an unusually large reduction in part-time contracts

[11] Some of the reasons for this decline are a greater preference for working predictable and fewer hours and increased participation of women in the medical workforce

within the period of adoption signaled in the resolutions.[12] The staggered adoption of the reform across public hospitals allowed me to implement an event study to evaluate the impact of the reform on in-hospital mortality. Importantly, I show that once hospital and year fixed effects are controlled for in the model, the timing of adoption was exogenous to several measures of in-hospital mortality. Not only hospitals did not get to choose their adoption dates; but also, there is no evidence that the government decided the implementation schedule based on hospitals' observable characteristics. Proof of this is that the results of the econometric models remain robust to the inclusion of pre-reform characteristics that could be considered by the authorities that designed the implementation schedule. With this in mind, I argue that any change in hospital mortality observed after the adoption of the reform can be interpreted as the causal effect of the reform.

The labor reform analyzed in this chapter is of interest to other countries in the region that are trying to expand health coverage. Ecuador is to the best of my knowledge the only country in the region that has adopted such a drastic reform with the intention of homogenizing the working conditions of professionals in the public sector and expanding health coverage. I find that the reform was successful in reducing in-hospital mortality, especially mortality within the first 48 hours of admission. This was true for the complete sample of admissions as well as for admissions to the emergency department. Importantly, I show that the effects reported in this paper cannot be attributed to changes in other quality indicators at the hospital level like the length of stay or by changes in the patient mix.

Chapter 4 studies whether changes to malpractice pressure facing physicians have effects on the use of c-sections and on measures of infant health and maternal

[12]Instead of this data-driven approach, I plan to use the complete implementation schedule as soon as I get access to it, but I do not expect to see significant changes in the results, considering that the main adoption dates are established in the resolutions of the Ministry of Labor and that for the event study I do not require the exact date of adoption, just the year.

mortality. A large literature has found that the liability system is an important determinant of health care provision (Baicker et al., 2006; Frakes, 2013; Kerschbamer and Sutter, 2017; Kessler and McClellan, 1996). I evaluate the effects of a reform to the malpractice legislation in Ecuador that took place in August 2014 and led to the hardening of sentences for cases of professional malpractice. I focus on the field of obstetrics because this is an area where doctors are very prone to malpractice suits, as have been documented by the literature on this topic in developed countries (Dubay et al., 1999; Currie and MacLeod, 2008).

To investigate whether doctors in Ecuador reacted to the malpractice reform, I use a difference-in-difference strategy that compares the outcomes of interest in Ecuador and Colombia before and after the reform. The results of several parallel trend tests on the outcomes of interest and tests for the stability of the demographic composition of both countries support my identification strategy. Previous studies for Ecuador warn about the excessive use of unnecessary c-sections in the period prior to the malpractice reform. The reform provides a unique opportunity to asses whether these excess c-sections were socially wasteful or not. For this, I test whether the reduction in c-sections that was triggered by the malpractice reform was accompanied by an improvement in infant or maternal health indicators (or remained unaffected), in which case the extra c-sections are deemed socially wasteful (Kessler and McClellan, 2002; Dubay et al., 2001).

This paper contributes to the literature on over-provision of health services, one of the major drivers of health care costs and a source of concern throughout the world (Cohen et al., 2015). The case of c-sections deserves special attention in Latin America, especially because of the high c-section rates in some countries that are close to 50%, which has led many organizations to suggest mechanisms to reduce the c-section rates (Guzman et al., 2015).

For the different analysis, I use vital statistics and death certificates micro-data from Ecuador and Colombia. Even though I show that Colombia is the best control group for Ecuador, several problems related to data quality and difficulties in accessing data from other countries hindered the comparisons that could have been made with other countries of the region in a timely manner.[13] This highlights the heterogeneity in the level of development of the countries of the region in the management of public data and to some extent the lack of homogeneity in the handling of data (OECD, 2014). The results of this study show that malpractice laws can reduce the excessive use of unnecessary procedures in low-risk patients without affecting patients' health. This conclusion is of great interest not only for the countries of the region but for other countries facing increasing health care costs linked to the over-provision of health care services.

Chapter 5 presents the main conclusions of this thesis. Furthermore, it discusses how each identification strategy was chosen carefully to prevent the effects of a certain policy from contaminating the effects of another policy. One possible concern could be that the labor reform started in 2009, but some hospitals were assigned to adopt the reform as late as 2014, the year of the passing of the new Penal Code. I argue that because by 2013, 82% of public hospitals had already adopted the labor reform, it is unlikely that the effects of the labor reform could affect the estimates of the impact of the malpractice reform that happened in August 2014. However, to remove any doubt, I perform robustness checks which include setting aside hospitals that adopted the labor reform in 2013 or later, to make sure that the effects of the malpractice reform on the c-the section rate, maternal mortality, and infant health were not affected by the passing of the labor reform. It was reassuring to see that the results were robust to the exclusion of said hospitals.

[13] With the exception of Colombia, Mexico, and Uruguay, it was not possible to access the micro-data on vital statistics or death certificates of other countries of the region without having to make an official request to the statistical office of each country.

Chapter 5 also discusses the limitations and policy implications of each of the papers and suggest avenues for future research. More often than not, the limitations found in each paper have to do with data availability and data quality in Ecuador and other countries of the region. Hence, apart from the policy implications that are derived from the analysis in each paper. I suggest several ways how Ecuador and the countries of the region can improve the quality of administrative records and health statistics, to facilitate public scrutiny of governments and facilitate the creation of comparative studies across the countries of the region.

Finally, Appendix A contains an additional paper where I contrast the performance of probabilistic record linkage versus phonetic encoding when linking individuals across several waves of Social Registry data. This document was conceived as an independent paper but at this stage, it has been included in the Appendix. Its objective is to document the procedure used to generate the database that I use in Chapter 2. The contributions of this paper are twofold. First, is one of the very few papers that document the best practices for conducting probabilistic record linkage using big administrative datasets recorded in Spanish. Second, this paper addresses the issues related to attrition and differential link rates that arise when linking individuals across time and proposes some alternatives to limit those problems including the use of an algorithm to follow individuals that leave their original household and sample-balancing to correct for differential link rates.

All in all, this thesis addresses questions that have not been discussed enough in the literature on cash transfers, like the persistence of the short-term effects of cash transfers in the long-run. It also provides first-hand evidence on the effects of labor market regulations on hospital quality in a developing country context and adds to the literature that looks at possible solutions to mitigate the problem of rising health care costs in developing countries, by studying the role of malpractice reforms.

Chapter 2

THE SHORT AND LONG-TERM EFFECTS OF CASH TRANSFERS ON EDUCATION AND LABOR MARKET OUTCOMES

Abstract

This paper investigates whether the effects of a cash transfer program persist or wear off in the long-run. I study the first two phases of Bono de Desarrollo Humano (BDH) in Ecuador, each of which lasted about five years. My identification strategy uses a regression discontinuity design and relies on the fact that at the threshold of eligibility, the second assignment to treatment (in 2008/9) was independent of the first assignment (in 2003). This allows me to disentangle the impact of a long exposure to the program (treatment during two phases) from a short exposure (treatment during one phase). Most of the gains in enrollment and schooling were achieved in the short-run among children that started treatment when they were about to start elementary school, high school or baccalaureate. However, an extended exposure to BDH was not enough to keep raising children's education. Regarding labor market outcomes, BDH had a negative (not statistically significant) impact on the probability of working among young children but did not increase job opportunities among young adults in the long-run.

2.1 Introduction and literature review

It has been widely accepted in the literature that factors operating during early childhood play a more important role than tuition, school reforms, job training or family credit constraints in explaining gaps in socioeconomic attainment (Carneiro et al., 2002; Cunha, Heckman, and Schennach, 2010; Heckman, 2000). The reason is that human capital investment exhibits both self-productivity and complementarity. Self-productivity has to do with the fact that skill attainment at one stage of the life cycle raises skill attainment at later stages of the life cycle. On the other hand, complementarity reflects the fact that early investments are more productive if they are followed by later investments. This explains why late investments that attempt to compensate for the lack of adequate early investments are very costly (Cunha, Heckman, Lochner, et al., 2006).

Cash transfers targeted to the poor have become the most popular tool in developing countries to encourage investment in the health and education of young children. These programs range from pure unconditional cash transfers to fully monitored and enforced conditional cash transfers (CCTs) that link the receipt of benefits to conditions such as school enrollment and regular health checkups (Baird et al., 2014). This paper studies the effects on young people's education and labor market outcomes of a continuous exposure to Bono de Desarrollo Humano (BDH), the main unconditional cash transfer program in Ecuador. While the previous literature on the long-term effects of cash transfers study how well the original treatment and control groups perform on several dimensions of interest after several years of a program's implementation, the question I address in this paper is whether cash transfers continue to be effective after several years targeting the same population.

Janvry et al. (2006) describe a basic model of school enrollment and work decisions. In said model, a CCT reduces the net cost of school. When the transfer is lower than the opportunity cost of schooling, it brings to school children with school cost above the reservation cost by less than the transfer. However, if the transfer is larger than the opportunity cost of schooling, it becomes a net payment for going to school and it brings to school any child with a positive utility for school. On the other hand, the transfer induces an income effect that causes an overall decrease in work participation (Janvry et al., 2006). In this line, several empirical studies have shown that CCTs are effective in increasing enrollment rates among beneficiary children in the short-run (see Fiszbein et al., 2009 for a review and Saavedra and Garcia, 2012 for a meta-analysis). Large gains on school participation have also been reported as a result of unconditional cash transfer programs labeled as education support programs like in the paper by Benhassine et al. (2015).

In general, the impact of CCTs is overall much stronger on schooling than it is on child work. One possible explanation is that in most CCTs the conditionality applies to school and not to work, then a CCT is not expected to have a big effect in preventing parents from increasing child work in response to an adverse income shock, considering that school and work are not incompatible (Janvry et al., 2006). There is mixed evidence regarding the impacts of CCTs on child work, with some studies finding significant reductions on child work among children most vulnerable to transitioning from schooling to work (Edmonds and Schady, 2012; Schady et al., 2008; Skoufias and Parker, 2001; Baird et al., 2014) and others finding no effects (Janvry et al., 2006).

While there is sufficient evidence of the impact of cash transfers in the short-run, evidence about their impacts in the long-run is sparse. In a recent study, Aizer et al. (2016) analyzed the long-term impact of the Mothers' Pension program in the US on educational attainment, income in adulthood and other health-related outcomes. The study found that boys of mothers whose application to the program was accepted accumulated one third more years of schooling and had a higher income in adulthood. Also for the US, Hoynes et al. (2016) found that access to food stamps in utero and in early childhood leads to significant increases in educational attainment and earnings.

In developing countries, the few existing papers that examine long-term effects focus mainly on Latin America. Since most of these programs are relatively young (the oldest CCT program, PROGRESA/Oportunidades, started in 1997), the analysis has focused on the study of outcomes measured at the end of high school or during early adulthood. Most of the evidence comes from PROGRESA /Oportunidades (Behrman, Parker, et al., 2011; Behrman, Sengupta, et al., 2005; Kugler and Rojas, 2018), and Nicaragua's CCT program Red de Protección Social (Barham et al., 2013a; Barham et al., 2013b).[1] For both countries, there is consistent evidence of impacts on schooling as well as some evidence for Nicaragua of impacts on learning, off-farm employment, and income. Other studies have also used short-term estimates to extrapolate to long-run program impacts (Behrman, Sengupta, et al., 2005; Todd and Wolpin, 2006; Attanasio et al., 2012).[2]

In addition to these studies, a contemporaneous paper by Araujo et al. (2016) studies the impact of Ecuador's Bono de Desarrollo Humano (BDH) after 10 years of its implementation. The authors use data from a randomized evaluation in 2003 and survey data collected in 2014 to test the impacts of BDH on various cognitive tests. Furthermore, they use Social Registry data and a RD design to evaluate the effects of BDH on school attainment and employment status of young adults, aged 19-25 years old in 2013/14. The results suggest a modest impact on high school completion, between 1 and 2 percentage points among 19-25 year olds and no effects on cognitive tests.

Unlike the previous literature that studies how well the original treatment and control

[1] Also Baez and Camacho (2011) found that children who benefited from the CCT Familias en Acción in Colombia were on average between 4 and 8 percentage points more likely to graduate from high school but no significant effects were found on cognitive tests.

[2] There are some concerns about the results that stem from extrapolations; one is that short-term evaluations may reveal only temporary improvements in the outcomes of interest, which may vanish as time goes by. Another concern is that they may fail to detect any impact because the time span between the treatment and follow-up may be too short (King and Behrman, 2009)

groups perform after several years of a program's implementation, the question I address in this paper is whether children who are treated for a longer period of time continue achieving better outcomes compared to similar children who were treated for a shorter period of time. In other words, I study whether treatment continues being effective in the long-run. My identification strategy relies on the fact that there was a change to the targeting rule five years after the implementation of BDH and, at the threshold of eligibility, the second assignment to treatment (in 2008/9) was independent of the first assignment (in 2003). This allows me to estimate the short term effects of BDH and the differential effect of a short exposure to the program (treatment during phase one) versus a long exposure (treatment during phases one and two).

I build a unique dataset that follows individuals (not households) through the three waves of the Registro Social (RS), which allows me to control for transitions in and out of the program and account for family dynamics that may also introduce bias to the estimates. More importantly, I assigned the eligibility status to households according to the Selben score that was in effect in each phase of the program and assigned the 'treated' status to children living with women who actually claimed the transfer each year. This generates the variation in lengths of exposure to the program that I exploit in this paper.

The results show that treated children experienced important short-term gains in enrollment and high school graduation by the end of phase one. Most of the gains in enrollment and schooling were achieved among children that started treatment when they were about to start elementary school, high school or baccalaureate. However, an extended exposure to BDH was not enough to keep raising children's education. This explains why 18-year-olds that were treated during the two phases of the program were not more likely to finish high school when compared to similar children that were only treated during the first phase of the program. The only group that experienced positive differential effects in years of education after being exposed during the two phases of the program were children aged 0 when BDH began in 2003. Regarding labor market outcomes, BDH had a negative yet not statistically significant impact on the probability of working among young children and did not increase job opportunities among young adults in the long-run.

This paper contributes to fill the gap in the literature about the long-term effects of cash transfers with emphasis on long duration programs and investigates whether these programs lose their effectiveness after several years targeting the same population. It also contributes to the literature on the complementarities of human development investments, and highlights the use of individual panel data to produce reliable quasi-experimental evidence of the differential effects of a long exposure to social programs that use a proxy means test to target beneficiaries, an approach that is very common in social programs in developing countries.

In the next section, I explain the institutional background and operational aspects of BDH since its implementation. Section 2.3 discusses the three sources of administrative data used in this paper and the methodology used to build the panels. Section 2.4 presents the empirical strategy used to identify the different impacts of BDH. Section 2.5 presents evidence on the validity of the identification strategy. Section 2.6 presents the results from the RD design. Section 2.7 discusses some robustness checks using administrative data from a standardized exam taken at the end of high school and section 2.8 concludes.

2.2 Background

Ecuador is a middle-income country that has experienced significant progress in terms of poverty reduction in the last decade. In 2005, Ecuador's GNI per capita was $7,310 in PPP-adjusted current U.S. dollars.[3] At that time 84.3% of the rural

[3] Source: World Bank World Development Indicators.

population lived in poverty as well as 35.1% of the urban population.[4]. By 2013, Ecuador's GNI per capita was US$10,720 in PPP-adjusted current U.S. dollars.[5] However, poverty remained a major concern for policymakers with 57.8% of the rural population and 24.8% of the urban population living in poverty.[6]

The reduction in poverty has been accompanied by a marked improvement in the main educational indicators as well as in a more modest reduction of child labor. The Ecuadorian educational system has three phases: (i) initial education, for children ages 3 to 5 is not compulsory; (ii) basic general education, is compulsory and lasts 9 years for children ages 5 to 15 years, and (iii) the Baccalaureate lasts 3 years for children ages 15 to 18 years and is not compulsory. Since 2012, at the end of the Baccalaureate, students have to pass a standardized exam (ENES exam) to apply to university. Until October 2008, public education was free only up to the tenth year of Basic General Education for children aged 5 to 15 years[7]. In October 2008, the new constitution declared that public education should be free from the first year of basic education up to the undergraduate level from the 2008-2009 school year onwards.

Primary education is almost universal and most of the recent efforts of the government have focused on raising the secondary education enrollment rates. Gender differences in educational attainment are small when compared for all males and all females (García-Aracil and Winter, 2006). By 2003, the net enrollment rate for basic general education was 84.1% and increased up to 90.4% in 2008 and up to 95.7% in 2014. While the net enrollment rate for baccalaureate was 42.1% in 2003;

[4]Source: Sistema Integrado de Indicadores Sociales del Ecuador (SIISE) (http://www.siise.gob.ec)

[5]Source: World Bank World Development Indicators

[6]Source: SIISE (http://www.siise.gob.ec)

[7]Education was free at public schools, however, most schools charged so-called "voluntary contributions" at the beginning of the school year. Those contributions were eliminated after the reform in 2008.

53.6% in 2008 and 65.1% in 2014.[8]

Despite the progress made in terms of schooling since 2003, youth unemployment, among individuals between 15 and 24 years of age, remained the highest among all the age groups even though it decreased from 22.6% in 1998 to 9.7% in 2014 (García et al., 2016). On the other hand, the rate of child work for children ages 5 to 14 years has decreased significantly since 2007, when 8% of children in said age range reported having a job, compared to 2014, when only 3% of children ages 5 to 14 worked[9].

BDH was implemented in 2003, when the incumbent government merged two existing cash transfer programs: Bono Solidario (BS) and Beca Escolar (BE) and became the first program to use a proxy means test to target the poorest families in Ecuador.[10] BDH uses a proxy means test called "Selben index", which is computed every five years using the information contained in a Social Registry called "Registro Social" (RS). The RS is a census of poor households that contains individual and household-level information on all potential BDH beneficiaries including age, gender, education level, household composition, characteristics of the house, access to public services, among others. The index is computed using a principal components analysis that assigns numerical values to categorical variables. Around 30 variables that are highly correlated with consumption were considered to compute the index. Most of these variables are categorical or ordinal and describe the characteristics of the household head and of the house. With categorical variables the distance

[8]Source:http://www.siise.gob.ec/agenda/index.html?serial=13

[9]Source: Compendio Estadístico INEC, 2014. Using data from the labor force survey ENEMDU. http://www.ecuadorencifras.gob.ec/compendio-estadistico-2014/

[10]BS was an unconditional transfer to compensate poor families for the elimination of gas and electricity subsidies in 1998, and targeted mothers with earnings below 40 USD per month, people with disabilities, and senior citizens. By 2003, it had about 1 million beneficiary households (World Bank, 2005). The program had an open inscription process that relied on the identification of needy families by parish priests, who were considered to have reliable knowledge of whom among their parishioners was poor. BE was a CCT program that started in the late 1990s and consisted of monthly transfers of 5 USD per child (up to two children per household), conditional on children's enrollment in school and a 90% attendance rate (Carrillo and Ponce Jarrín, 2009).

between categories 1 and 2 may not be equal to the distance between categories 2 and 3, so they cannot be treated as quantitative variables. Consequently, the principal components analysis by alternating least squares algorithm (PRINCALS) (De Leeuw and Van Rijckevorsel, 1980) is used to estimate optimal scaling and principal components simultaneously. Finally, using the percentage contribution of each variable to the total variance of the principal component model it is possible to generate an index that ranges from 0 and 100 that reflects the socio-economic situation of the households (Fabara, 2009).

To select the cutoff of eligibility, given that the RS mostly covers poor households, the same process is reproduced using a nationally representative survey. Households are sorted considering their Selben score and the score corresponding to the households in the fourth decile is selected as the cutoff of eligibility (Fabara, 2009).[11] The Selben index correctly predicts that 95% of households in the poorest quintile are eligible for the benefits and erroneously excludes 5 percent of them (Fiszbein et al., 2009).

The amount of the monthly transfer was US$15 in 2003 for individuals with families in the lowest 20% of the Selben distribution, and US$11.50 for those located in the next 20% of the distribution (equivalent to 12% and 9% of the minimum wage in 2003 respectively). Since then, the amount has increased progressively. In 2007, the transfer increased to US$30 (18% of the minimum wage) for individuals in the bottom 40% of the Selben distribution. In 2009, the transfer was raised to US$35 (16% of the minimum wage) and in 2013 it was set at US$50 per month (16% of the minimum wage).

BDH was initially publicized as a CCT; eligible women with children aged 0 to 18 years were required to take their children for health checkups and to enroll them

[11] This is why more than 40% of the households in the RS fall below the chosen cutoff.

in school.[12] Compliance with the conditions was supposed to be monitored every two months; however, due to lack of administrative capacity, this did not happen. Nevertheless, the program used radio and television spots to explicitly link transfers with the conditions, and some BDH administrators stressed the importance of school enrollment at the implementation stage (Schady et al., 2008). Furthermore, in 2008, the Ministry of Social Inclusion (MIES) started a process of notifying some mothers who were not satisfying the conditions; however, no further action was taken to sanction non-compliance.[13]

By 2014, BDH had 10 years of operation and people who were eligible during the two phases of the program could have received the transfer for a maximum of 10 years. From 2003 until 2008, hereafter the first phase of the program, BDH targeted women with children aged 0-18 years scoring less than 50.65 points (Ponce and Bedi, 2010). In 2009 a new Selben score was computed (Selben II), using most of the variables used in the estimation of the previous score and the information from the second wave of the RS.[14] From August 2009 until March 2013, the new eligibility cutoff was set at 36.5 points (Buser, 2015).[15]

Figure 2.1 shows the chronology of BDH implementation and, specifically, the change in the targeting rule in 2009 that coincided with the collection of the second wave of the RS and marked the beginning of the second phase of the program. Figure 2.1 also shows the different lengths of treatment that exist and the children of interest in this study. Regarding the former, a short exposure means that individuals were eligible during one of the phases of the program either phase one or two; while

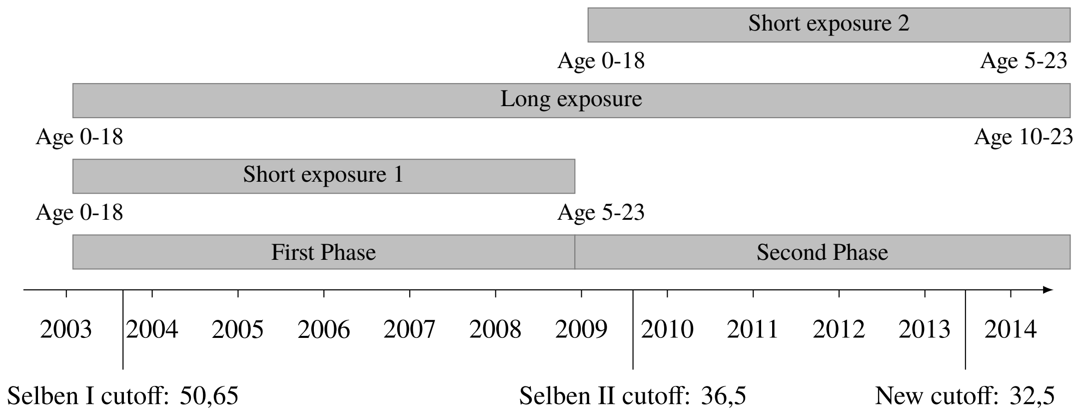
[12]The amount of the transfer does not depend on the age or number of other eligible children in the household, which may lead parents to choose the child in whom they want to invest the most.

[13]See Executive Decree No. 347-A of April 25 of 2003 published in the Official Registry No. 76 on May 7 of 2013.

[14]Cording to the Agreement MIES No. 00037

[15]Another change in the eligibility rule happened on March 2013 (according to the ministerial agreement No. 197 of 28 March 2013), when the beneficiaries whose score was between 32.5 and 36.5 points were excluded from the program.

Figure 2.1: Timeline of the implementation of BDH



long exposure means that individuals were eligible during the two phases of the program. Finally, individuals who were not eligible for the transfer may remain in this state during phase one, phase two or both.

The group of interest in this study is children aged 0-18 years when the program was launched in 2003, who grew up being exposed to BDH during phase one, phase two or both. At the end of phase one, these children were approximately 5 to 6 years older, and by the end of phase two, they were approximately 10 to 11 years older. Their exact age depends on the date they were surveyed in the first, second and third wave of RS. It is worth noting that for the analysis of the differential effects of BDH, the minimum age at the end of phase two is 10 years and the maximum age is 23 years because older children would not be eligible for treatment at the beginning of phase two.

2.3 Data and descriptive statistics

2.3.1 Description of the Data

I use social registry data called Registro Social (RS) and BDH payments data. The RS is considered a census of the poor because the first wave was conducted in 215 of the 223 registered counties in the 2010 census, and therefore covers most of the poor areas of the country. Furthermore, the second wave of the RS covers 2,393,377 of the 3,392,851 households that were registered in Ecuador in 2008 (Ponce and Falconí, 2011).[16] The RS records relevant information about BDH potential beneficiaries, including individual socio-economic information at the family and individual level, the ID number of the members (when available) and the Selben score assigned to each household. The geographic areas covered by the RS were selected using a poverty map that used information from the 2001 national census.

Three waves of the RS have been collected so far. The first wave was collected between 2001 and 2007. Of the 6,303,352 individuals that were surveyed 79% were surveyed before 2003 and 56.47% were surveyed in 2002 alone. The second wave was collected between 2007 and 2013. 70% of the 8,068,957 individuals in this wave were surveyed before 2009 and 68% were surveyed in 2008. Finally, the third wave contains the information of 6,930,701 individuals of which 36% was collected in 2013 and 64% in 2014.

The main challenge in linking the three waves of the RS is that many people do not report an ID (particularly children in the first wave of RS). For this reason, I use probabilistic record linkage to match individuals across waves using the two names and two last names recorded in the data and a common household ID as match keys. I also use an algorithm that takes into account the fact that some family members leave or join the household as time goes by (More details on the data linkage can be found in Appendix A.).[17]

The second source of administrative data is on BDH payments. These data are collected on a monthly basis and give an account of the amount and periodicity with

[16]Ecuador's population was 12,628,596 inhabitants by the year 2000, 14,447,600 inhabitants by 2008 and 15,661,312 inhabitants by 2013 according to the World Bank data (http://databank.worldbank.org/data/reports)

[17]400,000 individuals in the three-waves panel moved to (or from) other households

which beneficiaries collect their transfers from the different financial institutions. Compliance rates are high, on average, close to 90% of the individuals in the data set cash the transfers at least once every year. I merged the RS panel and payments data and assigned the actual treatment status to children who were living with women who claimed the transfer each year.

2.3.1.1 Descriptive statistics

I am able to follow 4,631,690 individuals across waves 1 and 2 of the RS and 5,439,749 individuals across waves 2 and 3. More importantly, I am able to follow 2,961,079 individuals across the three waves of the RS. The size of the three-waves panel is bounded by the number of matches between waves 1 and 2 because fewer people reported an ID in wave 1 (mainly children born in the 80's). There is also the fact that in the second wave there was an important influx of potential beneficiaries, which explains why wave 2 has 1.7 million more observations. Finally, in the last wave, at least 1.1 million individuals were not surveyed. The main reason was that during the data collection period, around 2,458 sectors were excluded from the RS leaving 735,479 individuals out of the survey.[18]

To tackle these issues that affect the matching rates, I adjusted the sampling weights on each panel-wave so that the totals of the adjusted weights on key characteristics match the corresponding population totals in each wave of RS. This process is known as raking or sample-balancing, and the key characteristics used in the process are known as auxiliary variables. Typical auxiliary variables include age, gender, years of education and geographic variables like province or county (see the Appendix A for more details).

The matching rate for households is high – 67% for waves 1 and 3 and 54% for wave

[18]This is less problematic than it seems because only a small fraction of the excluded households are close to threshold of eligibility based on their Selben Score. The data administrators also confirmed that the excluded sectors were left out because were more affluent on average.

2 –. On the other hand, the matching rate for individuals in the three-waves panel is around 37% to 47% in each wave of the RS (see Tables A.4 to A.6 in Appendix A).[19] However, once I focus on the groups of interest described in Figure 2.1 and on children who are close to the Selben cutoffs, the matching rates go up to 70% for some age groups as shown in Table 2.15.

In general the sample in the three-waves panel reproduces quite well the distribution of the Selben score in each wave, as shown in Figures 2.2, 2.3 and 2.4. Around 30 variables are involved in the estimation of the Selben index, among them, characteristics of the household head, features of the house, access to services, the number of household members who have migrated, assets, etc. The fact that the distribution in the panel-wave and the corresponding wave are very similar means that the panel represents households of all socio-economic backgrounds and that the households that I am not able to follow are not concentrated around the Selben cutoffs.

Furthermore, I formally test whether the eligibility status affects the probability to be in the three-waves panel among individuals who are close to the threshold of eligibility. I construct an indicator equal to one if the child is in the three-waves panel and zero otherwise and examine the association between this indicator and being eligible to receive the transfer. Table 2.15 shows that children in eligible households are not more likely to be in the three-waves panel compared to similar ineligible children. Only for children ages 18 and 19 years in wave 3, being eligible increases the likelihood of being in the panel in 1,3 percentage points.

The three-waves panel allows me to identify the individuals that were treated during the two phases of the program as well as those who left or joined the program at the beginning of phase two. It allows me to track the trajectory of the individuals in terms of their eligibility status, education and labor market outcomes. By following

[19]The matching rate for individuals goes up to 71% when we consider only households where at least one individual reported an ID. These households constitute the so called "likely to be linked" sample.

Figure 2.2: Kernel density graphs of the Selben score in the first wave of RS and and the corresponding panel wave

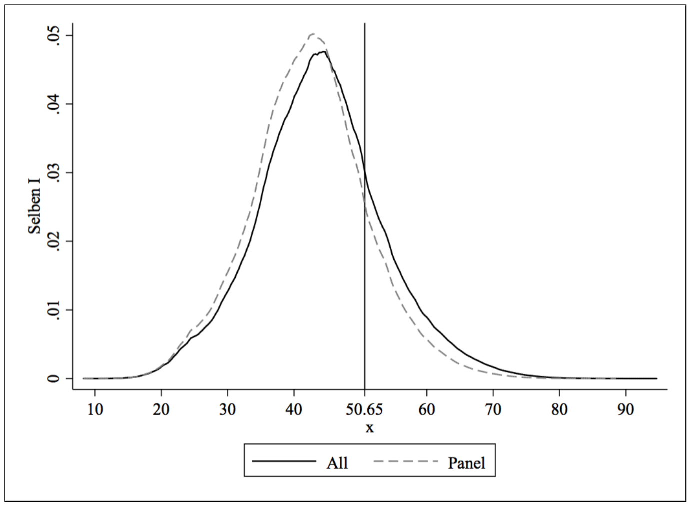


Figure 2.3: Kernel density graphs of the Selben score in the second wave of RS and and the corresponding panel wave

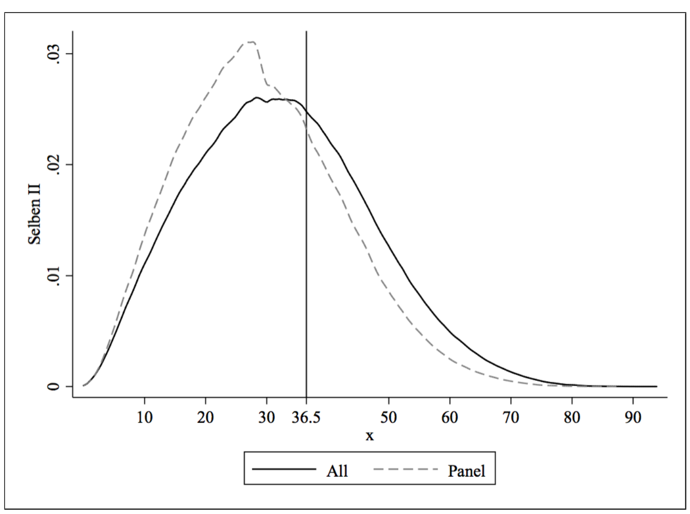
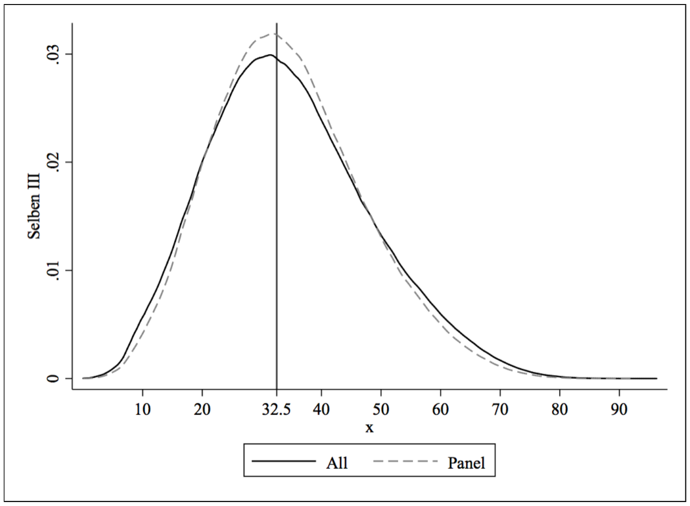


Figure 2.4: Kernel density graphs of the Selben score in the third wave of RS and and the corresponding panel wave



the children instead of the mothers, I make sure that I will not estimate the long-term effects of BDH on children that recently joined the household of a woman who was treated since 2003, but on children who lived with a treated woman since 2003 and may be currently living with her or not.

Table 2.1 shows the number of individuals who changed their eligibility status after the introduction of Selben II in 2009. There is substantial mobility in and out of the program. In fact, 20% of the people in the panel who were initially eligible to receive the transfer left the program by 2009 because they no longer met the selection criteria, and 28% of ineligible individuals became eligible. Looking specifically at the individuals around the Selben threshold established in 2003 (namely +/- 5 points from the 50.65 points cutoff), almost 35% of the individuals who were originally ineligible changed their status to eligible and 45% of those who were eligible became ineligible, which points to significant contamination of the original assignment groups.

Table 2.1: Eligibility status for individuals after the the introduction of Selben II in 2009

		Selben II		
		Ineligible	Eligible	Total
Selben I	Ineligible	148,417	58,274	206,691
		72%	28%	100%
	Ineligible +/- 5p	88,642	47,529	136,171
		65%	35%	100%
	Eligible	433,123	1,735,492	2,168,615
		20%	80%	100%
	Eligible +/- 5p	173,275	213,518	386,793
		45%	55%	100%
	Total	581,540	1,793,766	2,375,306
		24%	76%	100%
	Total +/- 5p	261,917	261,047	522,964
		50%	50%	100%

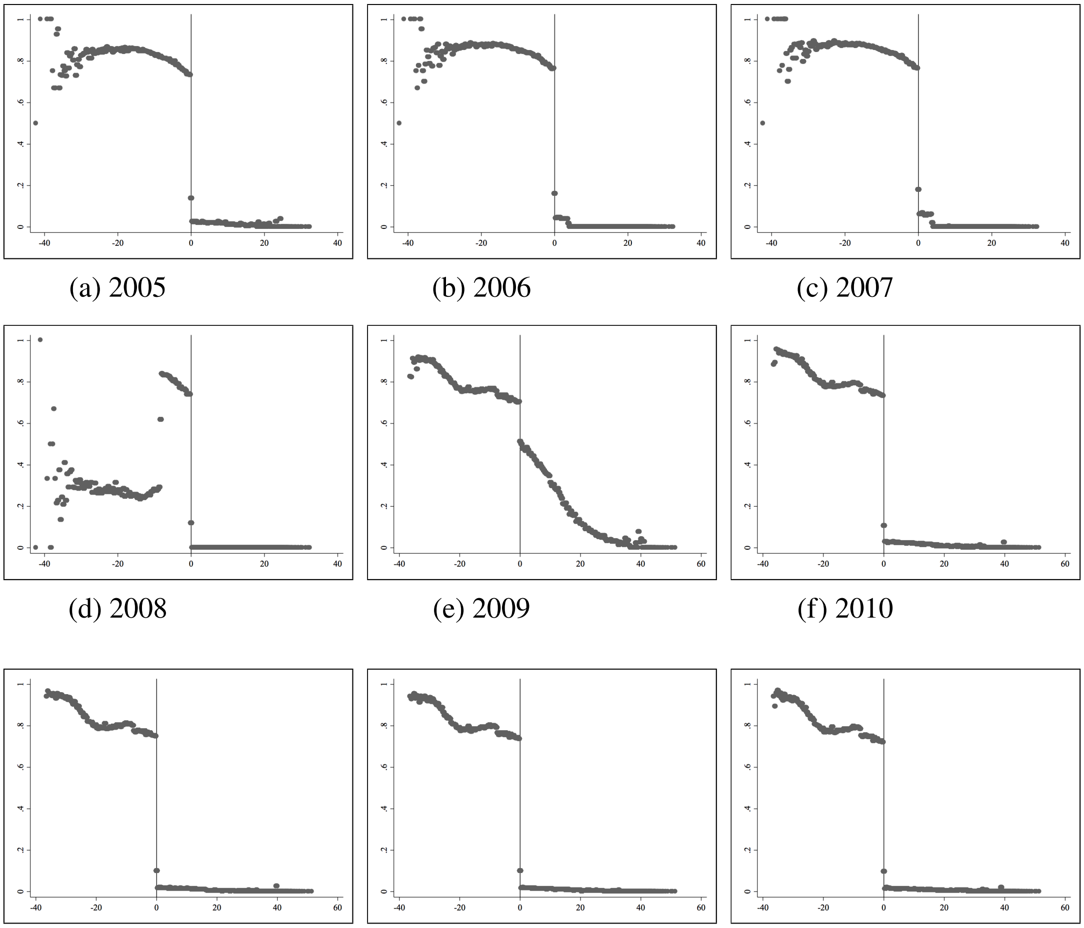
Notes: The table shows the transition matrix of the eligibility status to BDH before and after the introduction of Selben II. The total is 2,375,306 individuals instead of 2,961,043 because in the first wave not all the households had kids below age 18.

2.4 Empirical Strategy

2.4.1 Regression discontinuity design

The discontinuity in the assignment rule allows me to recover the local causal effects of exposure to BDH by comparing the outcomes of similar individuals who are just below the threshold (hence eligible for the transfer) and just above the threshold (not eligible for the transfer). Given that the Selben score predicts substantial but not perfect changes in the probability of treatment, meaning that the probability of treatment does not jump from 0 to 1 when the forcing variable crosses the threshold, I use a Fuzzy RD design. In the Fuzzy RD design, the treatment effect is obtained by dividing the jump in the outcome variable (Y) at the threshold to the jump in the treatment probability at the threshold as in an instrumental variable approach or the analogous Wald estimator. The different graphs in Figure 2.5 show the jump in the treatment probability at the Selben I threshold (50.65 points) and at the Selben II threshold (36.5 points) for each year starting in 2005, the first year for which

Figure 2.5: Proportion of treated households with respect to the threshold by year



(a) 2005 (b) 2006 (c) 2007

(d) 2008 (e) 2009 (f) 2010

(g) 2011 (h) 2012 (i) 2013

Notes: The graphs use RS data merged with administrative payments data. The sample is individuals that were surveyed before 2003 in the first wave and before 2009 in the second wave. The graphs show the proportion of households that received the treatment each year. The cutoff for the years 2005-2008 is 50.65 points and the cutoff for the years 2009-2013 is 36.5 points. The change in the assignment rule happened in August 2009 but the 2008 graph also exhibit some adjustment. In particular the 2008 payments dataset contains 676,068 individuals while the 2007 and 2009 datasets contain 1,127,909 and 1,280,367 individuals respectively.

payment data is available.

In the next section, I explain how I estimate the impact of BDH at the end of each phase of the program, as well as the differential impact of a long exposure (during phases one and two) versus a shorter exposure to BDH (during phase one).

2.4.2 Estimation

Several non-parametric methods have been proposed in the literature to estimate the local average treatment effect (LATE). One corresponds to the series estimation approach, which consists of the inclusion of polynomial functions of the forcing variable and provides estimates of the regression function over all the values of the forcing variable. The other non-parametric approach is kernel regressions. In the simplest case of the rectangular kernel, one computes the local average of the outcome (Y) in the closest bin to the left and right of the cutoff point and compare those means to get the RD estimate. However, Hahn, Todd, et al. (2001) argue that if the true model is upward sloping on both sides of the threshold, the RD estimate from kernel regression would be biased; moreover, any attempt to reduce the bias by reducing the bandwidth size would lead to very imprecise estimates in the absence of a large number of observations near the cutoff. To solve this problem Hahn, Todd, et al. (2001) suggest running local linear regressions at each side of the threshold instead of computing local averages within the closest bins. Hahn, Todd, et al. (2001) also proved that this approach reduces bias by one order of magnitude.

In the case of BDH, eligible individuals are located on the left-hand side of the cutoff, meaning that only people with a lower Selben score can benefit from the transfer. Following Lee and Lemieux (2010), the regression model on the left-hand side of the cutoff point ($S \leq c$) is:

$$Y = \alpha_l + f_l(S - c) + \epsilon \tag{2.1}$$

Y is the outcome variable, $f_l(.)$ and $f_r(.)$ are functional forms of the Selben score (S) that measures the distance to the cutoff c, α_l is the intercept. The regression model at the right hand side of the cutoff point ($S > c$) is:

$$Y = \alpha_r + f_r(S - c) + \epsilon \tag{2.2}$$

It is preferable to estimate the treatment effect with a pooled regression on both sides of the threshold. The advantage of this approach is that it directly yields estimates and standard errors of the treatment effect τ:

$$Y = \alpha_r + \tau T + f(S - c) + \epsilon \tag{2.3}$$

where $\tau = \alpha_l - \alpha_r$ and $f(S - c) = f_r(S - c) + T[f_l(S - c) - f_r(S - c)]$. The treatment status T is instrumented by D, which is a binary variable that takes a value of 1 when the Selben score is below the cutoff and 0 otherwise. It is important to let the regression function differ on both sides of the cutoff point by including interaction terms between T and S. In the linear case where $f_l(S - c) = \beta_l(S - c)$ and $f_r(S - c) = \beta_r(S - c)$, the pooled regression is:

$$Y = \alpha_r + \tau T + \beta_r(S - c) + (\beta_l - \beta_r)T(S - c) + \epsilon \tag{2.4}$$

The simplest fuzzy RD estimator uses only D as instrument without polynomial interactions of $f(S - c)$ with D. In this case, I allow for interaction terms in the first and second stage.

2.4.3 Short exposure

To estimate the effects of a short exposure to BDH at the end of phase one, I compare the outcomes (observed at the end of phase one) of children who were marginally eligible or not based on their proximity to the Selben score cutoff of 50.65 points set on 2003. Likewise, to estimate the effects of a short exposure to BDH, during phase two only, I restrict the sample to children who were not treated during phase one and compare the outcomes of children observed at the end of phase two with a Selben II score close to the cutoff for eligibility fixed at 36.5 points in 2009.

The analysis of the effects of each phase of the program is particularly important in this setting because BDH was publicized as a CCT. However, due to lack of administrative capacity the conditions were never enforced so it is likely that most of the impact was achieved in the first phase of the program when the transfer was believed to be conditional at least for a short period of time (De Brauw and Hoddinott, 2011; Baird et al., 2014; Benhassine et al., 2015). Furthermore, in October 2008, education became free in all public schools and universities in Ecuador, which may have caused the transfer to cease to have an effect on eligible children, since education became free for all eligible and ineligible children. It is important to contrast the results obtained in each of the phases separately to be able to identify the possible reasons why the effects of BDH may be different during phase one and two.

To estimate the effects of a short exposure to BDH, I estimate equation 2.4, where Y is the outcome variable observed at the end of phase one or at the end of phase two depending on the case. I instrument treatment with individual eligibility using the corresponding cutoff depending on whether I am evaluating the effects at the end of phase one or at the end of phase two. I try several bandwidths (+/-2.5, +/-5 and +/-7.5 points with respect to the cutoff) and test the robustness of the estimates to the inclusion of higher order polynomial terms. I estimate regressions for different age groups and include county fixed effects, gender, race and school year dummies because each wave of the RS was collected over more than one year.

Since each wave of the RS was collected over several years in the analyses I restrict the sample to individuals that were surveyed before 2003 in the first wave or who were surveyed before 2009 in the second wave. This is important for the balance tests of pre-treatment characteristics and also because in this way I avoid having cases where people respond to the first survey very late and shortly after respond to the following survey or cases where the information, on the contrary, is very spaced.

2.4.4 Differential effect of a long exposure vs a short exposure

A significant number of individuals changed their eligibility status after the introduction of Selben II in 2009. As a consequence, comparing the outcomes of children from treated households versus untreated households ten years later using the original assignment in 2003 would lead to underestimation of the treatment effect of the program, since both groups were exposed to the treatment at some point in time.

Another way to estimate long-term effects is to measure the differential impact of a long exposure to BDH (during phases one and two) versus a short exposure to BDH (during phase one). To do this, I compare the outcomes observed at the end of phase two of children who were marginally treated (or not) during phase two among children who were treated during phase one. Since most of these individuals were eligible during phase one, by comparing households that are very close in terms of their Selben II score, I would be comparing two groups of people that are very similar in terms of observable and unobservable characteristics.

To estimate the differential impact of a long exposure (during phases one and two) versus a short exposure to BDH (during phase one), I restrict the sample to individuals who were treated during the first phase of the program and estimate equation 2.4, where Y is the outcome variable observed at the end of phase two and $(S-c)$ is a function of the Selben II score (S) that measures the distance to the cutoff (36.5 points).

As with the short-term effects, I instrument treatment with individual eligibility and try several bandwidths (+/-2.5, +/-5 and +/-7.5 points with respect to the cutoff) as well as low polynomials of the distance to the cutoff to check the robustness of the RD estimates to different specifications. I estimate separate regressions for different age groups and include county fixed effects, a gender dummy and school year dummies in the regressions. I restrict the sample to individuals that I follow across the second and third wave of the RS that were surveyed before 2009 in the second wave.

2.5 Assessing the validity of the identification strategy

For the identification strategy to be valid, individuals should not be able to precisely manipulate their Selben score. This is unlikely in this setting because the Selben index is a complex "composite index" and its methodology has never been disclosed. People do not know the weights associated with their responses when they are surveyed, making it very difficult for them to determine which answers will make them end up on the left side of the cutoff. Furthermore, the change in the methodology in 2009, which involved the use of new variables to build the index, made cheating even more difficult. Figure 2.6 shows that there is no evidence of bunching or manipulation of the Selben I or Selben II scores. Moreover, the density tests shown in Table 2.2 fail to reject the hypothesis that the difference in densities on the two sides of the cutoff is zero.[20] These results rule out possible self-selection or non-random sorting of units into eligible or ineligible groups.

For identification, it is also important that there are no imbalances in baseline characteristics. To test for balance in the pre-treatment characteristics, I use a regression discontinuity approach and estimate local regressions of different polynomial orders using the observations within +/-2.5 points of distance to the cutoff. The variables were chosen from the list of variables used to estimate the Selben score.

I did this with the people who were surveyed before 2003 in the first wave of the RS, and with the people who became eligible for phase two, who were surveyed before 2009 in the second wave. I also tested whether among the former eligible children, those who were considered eligible or not for the second phase of the program had balanced characteristics. Results of the linear and quadratic specifications as well

[20]Cattaneo et al. (2016) propose a set of manipulation tests based on a novel local polynomial density estimator, which does not require pre-binning of the data as opposed to McCrary's test.

Figure 2.6: Frequency distribution of Selben I and Selben II score

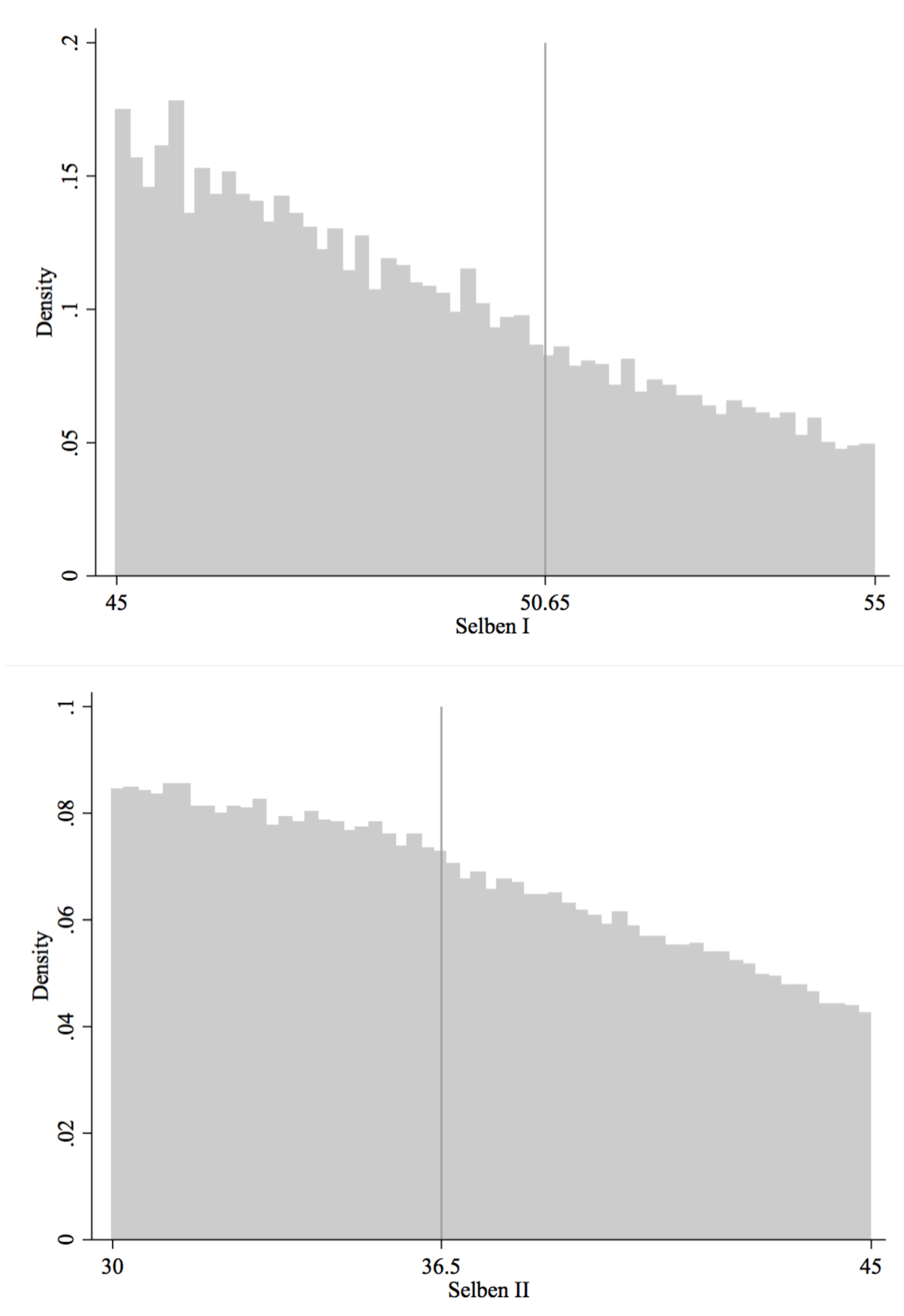


Table 2.2: Density tests for Selben I and Selben II

	Left of Selben I cuttof	Right of Selben I cuttof	Left of Selben II cuttof	Right of Selben II cuttof
Observations	651103	171121	555666	244332
Effective Observations	43763	36769	39043	40559
Bias corrected density	0.03	0.03	0.02	0.02
Standard error	0.00	0.00	0.00	0.00
Bandwidth values	1.54	1.55	2.00	2.27
Standard error test	0.00		0.00	
p-value	0.64		0.94	

Notes: Density tests based on Cattaneo et al. (2016). This local polynomial density estimator does not require pre-binning of the data as opposed to McCrary's test.

as the omnibus joint F-tests are reported in Tables 2.3, 2.4 and 2.5. In all the cases, at least for one of the specifications, the p-value on the omnibus F test was not statistically significant.

Table 2.3 shows that the pre-treatment characteristics of marginally eligible children in 2002 (around +/-2.5 points of the cutoff) present some minor imbalances in the linear specification that disappear in the quadratic specification. The quadratic specification performs better in terms of the omnibus joint F-test, but the linear specification was preferred in terms of the Akaike information criterion (AIC) and the goodness of fit test performed by jointly testing the significance of a set of bin dummies included as additional regressors in the model in order to select the optimal order of the polynomial.

Table 2.4 shows that among people who were not treated before 2009, the pre-treatment characteristics of the two comparison groups were balanced, including in terms of the Selben I score for both the linear and quadratic specifications. The omnibus joint F-test favors the linear specification, which is also preferred based on the AIC criterion and the goodness of fit tests.

Finally, I tested whether children who were treated in phase one, and were assigned

Table 2.3: Balance tests for pre-treatment characteristics (before 2003) for the analysis of a short exposure during phase one

	Linear specification		Quadratic specification	
Variables	Pt. Est	Std. Err	Pt. Est	Std. Err
Does the household own land	-0.0117**	(0.00546)	-0.0180**	(0.00818)
Electricity	0.00138*	(0.000731)	0.000621	(0.00110)
Does not have exclusive shower	0.00343	(0.00300)	0.00270	(0.00459)
Overcrowding	0.000363	(0.00630)	-0.000143	(0.00950)
Household members	0.00916	(0.0188)	0.00474	(0.0283)
Total earners including children	0.000235	(0.0106)	-0.0111	(0.0159)
Adult members who work	0.00771	(0.00701)	0.000952	(0.0105)
Members below 18 years	0.00678	(0.0132)	0.0286	(0.0199)
Children studying	0.0114	(0.0119)	0.0271	(0.0179)
Education level of the head	0.0190*	(0.0109)	0.0105	(0.0164)
Does the head have a job	0.00901**	(0.00451)	0.00267	(0.00677)
Does the head speaks native languages	-0.000442	(0.00223)	5.01e-05	(0.00334)
Is the head retired	-0.000450	(0.000859)	0.000284	(0.00122)
Years of education of the head	0.0471	(0.0463)	0.0434	(0.0694)
Joint F-test	19.50		10.32	
P-value	0.1467		0.7386	
Observations	94,965		94,965	

Notes: Robust standard errors shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Sample is households surveyed before 2003 in the first wave of RS that have an adult woman and children below age 18. Overcrowding is equal to 1 if the ratio of family members to bedrooms is above 3 and 0 otherwise.

to be treated or not in phase two based on their Selben II score, had balanced pre-treatment characteristics at the beginning of phase two. Results from the linear and quadratic specification show that those characteristics are well balanced including in terms of the Selben I score, meaning that in the neighborhood of the cutoff point, assignment in 2009 was independent of the first assignment in 2003. As shown in Table 2.5. Both specifications performed well in terms of the omnibus joint F-test, while the linear specification was slightly preferred based on the AIC criterion and the goodness of fit test.

Table 2.4: Balance tests for pre-treatment characteristics (before 2009) for the analysis of short exposure during the second phase of the program

	Linear specification		Quadratic specification	
Variables	Pt. Est	Std. Err	Pt. Est	Std. Err
Selben I score	-0.0977	(0.192)	0.122	(0.289)
Land	-0.00466	(0.00706)	-0.0116	(0.0107)
Electricity	0.000447	(0.00140)	0.00167	(0.00212)
Does not have exclusive shower	0.00153	(0.00847)	-0.00725	(0.0126)
Overcrowding	-0.00290	(0.00856)	-0.00510	(0.0128)
Household members	0.0220	(0.0382)	0.0683	(0.0575)
Total earners including children	-0.00413	(0.0159)	0.0280	(0.0234)
Adult members who work	-0.00189	(0.0140)	0.0120	(0.0211)
Members below 18 years	0.0238	(0.0287)	0.0357	(0.0431)
Children studying	0.0161	(0.0268)	0.00128	(0.0402)
Education level of the head	-0.0176	(0.0387)	0.0198	(0.0583)
Does the head have a job	-0.000451	(0.00970)	0.0203	(0.0147)
Does the head speaks native languages	-0.000207	(0.00365)	-0.000543	(0.00540)
Is the head retired	-0.00123	(0.00126)	-0.00281*	(0.00163)
Years of education of the head	0.0240	(0.0966)	0.133	(0.145)
Joint F-test	4.77		12.88	
P-value	0.9939		0.6113	
Observations	22,811		22,811	

Notes: Robust standard errors shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1.Sample is households surveyed before 2009 in the second wave of RS that have an adult woman and children below age 18. Overcrowding is equal to 1 if the ratio of family members to bedrooms is above 3 and 0 otherwise.

Table 2.5: Balance tests for pre-treatment characteristics (before 2009) for the analysis of long vs short exposure to BDH

	Linear specification		Quadratic specification	
Variables	Pt. Est	Std. Err	Pt. Est	Std. Err
Selben I score	-0.0108	(0.0809)	-0.0926	(0.121)
Land	-0.00157	(0.00417)	-0.00212	(0.00622)
Electricity	0.000776	(0.000881)	0.000221	(0.00138)
Does not have exclusive shower	-0.00264	(0.00495)	0.00254	(0.00737)
Overcrowding	-0.000476	(0.00548)	0.00724	(0.00818)
Household members	-0.0115	(0.0251)	-0.000694	(0.0375)
Total earners including children	-0.00601	(0.0105)	-0.00974	(0.0157)
Adult members who work	-0.00797	(0.00879)	0.00444	(0.0132)
Members below 18 years	-0.0155	(0.0193)	-0.0133	(0.0287)
Children studying	-0.00847	(0.0171)	-0.0134	(0.0254)
Education level of the head	-0.0236	(0.0225)	-0.0424	(0.0337)
Does the head have a job	-0.00344	(0.00606)	-0.000756	(0.00902)
Does the head speaks native languages	-0.000198	(0.00226)	0.000904	(0.00323)
Is the head retired	0.000573	(0.000592)	0.000565	(0.000881)
Years of education of the head	-0.0445	(0.0580)	-0.0849	(0.0864)
Joint F-test	7.23		5.71	
P-value	0.9508		0.9842	
Observations	56,872		56,872	

Notes: Robust standard errors shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Sample is households surveyed before 2009 in the second wave of RS that have an adult woman and children below age 18. Overcrowding is equal to 1 if the ratio of family members to bedrooms is above 3 and 0 otherwise.

2.6 Results

In this section, I present results of the 2SLS regressions in equation 2.4 that uses the treatment assignment as instrument for the actual treatment. Considering that there is some variation in treatment duration across households, I restrict the sample in the three-waves panel to households that were treated for at least three years in each of the phases of the program. The percentage of observations that meet this requirement is close to 99% of the sample, which is why the results for the complete sample are almost identical.[21]

2.6.1 Effects of a short exposure to BDH measured at the end of phase one

2.6.2 Education outcomes

Table 2.6 reports on the short-term effects of BDH on enrollment and years of education. Columns 1 and 3 show the results of the reduced form regressions (ITT), while columns 2 and 4 show the results of the 2SLS regression in equation 2.4. All the regressions include time and county fixed effects, individual-level control variables like gender and race and are estimated on the sample of individuals that are located within +/- 2.5 points of the Selben I cutoff.

I report the results for children aged 10 to 21 years by the end of phase one. The results for younger children can be found in Table 2.16 in the Appendix. These children were between 4 and 16 years old when the first phase began in 2003. By choosing these age groups, I am able to observe the effects of BDH on children who began treatment at all the juncture ages in their academic progress which are age 5, 12 and 15. These are juncture ages because at age 5 parents decide to enroll their children at elementary school, at age 12 they decide to enroll them in high school and at age 15 they decide whether or not to enroll them in Baccalaureate. Furthermore,

[21]It is worth noting that I replicated the analysis of the short term effects of BDH by the end of phase 1 using the two-waves panel that follows around 4,6 million individuals across waves 1 and 2 and found similar results to the ones I report in section 2.6.1 using the three-waves panel.

Table 2.6: Short-term effects of BDH on enrollment and years of education by the end of phase one

	Enrollment		Years of Education		
	(1)	(2)	(3)	(4)	
Age in 2008	ITT	2SLS	ITT	2SLS	N
Effect on 10 year olds	-0.00518	0.0130	0.00678	0.212**	5561
	(0.00422)	(0.00836)	(0.0468)	(0.0886)	
Pre- treatment mean	0.992	0.992	5.160	5.160	
Effect on 15 year olds	-0.0201*	0.0255	0.0180	0.147	6464
	(0.0121)	(0.0224)	(0.0662)	(0.117)	
Pre- treatment mean	0.886	0.886	8.863	8.863	
Effect on 16 year olds	0.0172	0.0127	0.0512	-0.0454	5861
	(0.0149)	(0.0264)	(0.0830)	(0.146)	
Pre- treatment mean	0.839	0.839	9.291	9.291	
Effect on 17 year olds	-0.00243	0.0872**	-0.107	0.170	5449
	(0.0183)	(0.0357)	(0.0911)	(0.174)	
Pre- treatment mean	0.723	0.723	9.791	9.791	
Effect on 18 year olds	0.000474	0.0836*	0.103	0.321	5343
	(0.0210)	(0.0437)	(0.105)	(0.214)	
Pre- treatment mean	0.539	0.539	10.42	10.42	
Effect on 19 year olds	-0.0220	0.0511	-0.110	0.296	5099
	(0.0215)	(0.0431)	(0.111)	(0.227)	
Pre- treatment mean	0.433	0.433	10.54	10.54	
Effect on 20 year olds	0.00882	0.111**	0.105	0.0170	4698
	(0.0219)	(0.0456)	(0.124)	(0.257)	
Pre- treatment mean	0.379	0.379	10.73	10.73	
Effect on 21 year olds	-0.0732***	0.0608	-0.110	0.441	4335
	(0.0222)	(0.0483)	(0.142)	(0.323)	
Pre- treatment mean	0.331	0.331	10.70	10.70	
County and time FE	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The table reports on the results of ITT and 2SLS regressions using a linear polynomial of the distance to the cutoff. Sample is children with a score within +/-2.5 points from the Selben I cutoff (50,65 points). All the regressions control for gender and race of the individuals.

I study children aged 4 to 16 years in 2003, because the analysis focuses on children who were treated for at least three years during phase one.

The results of the reduced form regressions in column 1 show negative effects of BDH on enrollment for some age groups, one reason is that the program did not reach all the households with a Selben score below the threshold of eligibility. Figure 2.5 shows that before 2009, take up was always below 80% in the area closer to the threshold. Once I correct for the endogeneity of take up by instrumenting the actual treatment with the assignment to treatment, the effects turn positive for all the age groups.

According to the 2SLS estimates reported in column 2, BDH was effective in raising the enrollment rates of 10 to 21 year olds. However, the effect was statistically significant only for 17, 18 and 20 year olds who were around 12, 13 and 15 years when the program started in 2003. Enrollment increased by 8.7 percentage points by the end of phase one among 17 year olds and by 8.3 and 11 percentage points for 18 and 20 year olds respectively. Schady et al. (2008) also investigates the short-term impacts of BDH on enrollment and other education and labor market outcomes. They study a randomized experiment carried out in 2003-2005 and find that school-aged children (6-17 years) from households who won the lottery were 3.2 to 4.0 percentage points more likely to be enrolled in school than children in households who did not win. Since Schady et al. (2008) report an average effect for school-aged children is hard to compare their estimates to the results discussed here. However, I cannot rule out that some age groups may have had similar effects to the ones I reported here that may have disappeared in the average.

Given that when the program started most people believed that one of the conditions to receive the transfer was to send children to school, I expected to see positive effects on enrollment, however, what is more important is whether these children completed more years of schooling. Columns 3 and 4 report on the short-term effects of BDH on years of education by the end of phase one. The reduced form specification reported in column 3 show some negative coefficients that turn positive after instrumenting the actual treatment with the assignment to treatment. Only 10 year olds, who were 5 years old when the first phase began experienced an increase in their years of education by 0.21 years, for children aged 15, 17, 18, 19, 20 and 21 years in 2008/2009, the effects were positive but not statistically significant.

The observed gains in enrollment are in line with a higher probability of graduating from high school for 18 year olds. For this group, being treated for at least three years during phase one increased the likelihood of graduating by 8.7 percentage points by the end of phase one (See column 2 of Table 2.7). This increase is equivalent to an 18% raise relative to a 48.7% pre-treatment high school graduation rate. Treated individuals aged 17 and 19 years by the end of phase one also experienced a higher probability of graduating from High School, although not statistically significant.

Furthermore, there was a positive effect of BDH on the likelihood of having some college education among young adults aged 19 year old in 2008/2009. These young people were 14 years old when the program started in 2003 and were old enough to benefit from free education at college level (established in 2008). The likelihood of having some college education increased by 7.4 percentage points for this group, relative to a pre-treatment rate of 17.3%, which implies a 43% raise. For 20 and 21 year olds, the effect of BDH was also positive but not statistically significant.

2.6.3 Labor market outcomes

Table 2.8 shows the results of the ITT and 2SLS regressions of the effect of BDH on the probability of having a job by the end of phase one. In line with previous literature (Skoufias and Parker, 2001; Janvry et al., 2006; Attanasio et al., 2012; Schady et al., 2008), the sign of the job coefficients is contrary to the sign of the enrollment coefficients for 16, 17, 18, 20 and 21 year olds. The 2SLS estimates reported in column 2 show that the likelihood of having a job was negative but not statistically significant for most age groups and negative and statistically significant for 21 year olds (who were 16 in 2003). By the end of phase one, BDH reduced the employment rate of this age group by 12.5 percentage points relative to a pre-treatment rate of 56%, which implies a 22% reduction for this group. The magnitude of this effect is comparable to the 9.9 percentage points decrease in paid employment

Table 2.7: Short-term effects of BDH on high school graduation and on the likelihood of having some college education by the end of phase one

	High School		Has some college		
	(1)	(2)	(3)	(4)	
Age in 2008	ITT	2SLS	ITT	2SLS	N
Effect on 17 year olds	-0.0478***	0.0249			5449
	(0.0166)	(0.0306)			
Pre- treatment mean	0.186	0.186			
Effect on 18 year olds	0.00984	0.0870**			5343
	(0.0211)	(0.0441)			
Pre- treatment mean	0.487	0.487			
Effect on 19 year olds	-0.0350*	0.0369	-0.0290	0.0737**	5099
	(0.0203)	(0.0419)	(0.0178)	(0.0355)	
Pre- treatment mean	0.573	0.573	0.173	0.173	
Effect on 20 year olds			-0.00882	0.0323	4698
			(0.0206)	(0.0417)	
Pre- treatment mean			0.228	0.228	
Effect on 21 year olds			-0.0375*	0.0482	4335
			(0.0213)	(0.0466)	
Pre- treatment mean			0.246	0.246	
County and time FE	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The table reports on the results of ITT and 2SLS regressions using a linear polynomial of the distance to the cutoff. Sample is children with a score within +/-2.5 points from the Selben I cutoff (50,65 points).All the regressions control for gender and race of the individuals.

reported by Edmonds and Schady (2012) who evaluate the impacts of BDH after 17 months of BDH's implementation on children aged 11–16 at baseline.

2.6.4 Effects of a short exposure to BDH measured at the end of phase two

2.6.5 Education outcomes

In terms of education outcomes, what makes the second phase of the program different from the first phase is that by the end of phase two, children attained more years of education but no significant effects were found in terms of enrollment. One possible explanation is that the size of the transfer was smaller in the first years of the program, furthermore, parents believed that BDH was conditional on enrollment.

Table 2.8: Short-term effects of BDH on the likelihood of having a job by the end of phase one

	Has a job		
	(1)	(2)	
Age in 2008	ITT	2SLS	N
Effect on 15 year olds	0.0136	0.0195	6465
	(0.00896)	(0.0166)	
Pre- treatment mean	0.0670	0.0670	
Effect on 16 year olds	-0.0142	-0.00178	5868
	(0.0117)	(0.0211)	
Pre- treatment mean	0.101	0.101	
Effect on 17 year olds	0.0177	-0.0287	5453
	(0.0163)	(0.0321)	
Pre- treatment mean	0.200	0.200	
Effect on 18 year olds	0.00423	-0.00627	5350
	(0.0198)	(0.0411)	
Pre- treatment mean	0.337	0.337	
Effect on 19 year olds	0.00693	0.0344	5105
	(0.0209)	(0.0418)	
Pre- treatment mean	0.449	0.449	
Effect on 20 year olds	-0.00196	-0.0600	4703
	(0.0217)	(0.0455)	
Pre- treatment mean	0.507	0.507	
Effect on 21 year olds	0.0288	-0.125***	4345
	(0.0214)	(0.0474)	
Pre- treatment mean	0.562	0.562	
County and time FE	Yes	Yes	
Controls	Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The table reports on the results of ITT and 2SLS regressions using a linear polynomial of the distance to the cutoff. Sample is children with a score within +/-2.5 points from the Selben I cutoff (50,65 points). All the regressions control for gender and race of the individuals.

However, since the condition was on enrollment and not on grade progression, there were no effects on years of education.

By contrast, in phase two, not only parents received a bigger transfer, but also, education became free after tenth grade with the reform of 2008. This may have reinforced the effect of BDH especially at the juncture ages of 12 and 15 years, when children decide whether or not to enroll at high school and at the first year of Baccalaureate, respectively.

Columns 1 and 2 in Table 2.9 report on the short-term effects of BDH on enrollment by the end of phase two. My preferred specification is in column 2 and shows that BDH did not have a significant positive effect on enrollment for any of the age groups. If anything, there was a significant reduction on the enrollment rate of 19 year olds by 22.4 percentage points.

Columns 3 and 4 show the impacts of BDH on years of education by the end of phase two. Results from the 2SLS regressions show a positive and statistically significant effect on years of schooling for 17 and 20 year olds. The total years of education for 17 year olds increased by 1.2 years, relative to a pre-treatment mean of 10.11 years for this age group. While, for 20 year olds, schooling increased by 1.9 years, relative to a pre-treatment mean of 11.12 years of education.

Table 2.10 shows the short-term effects of BDH on high school graduation rates and the likelihood of having some college education by the end of phase two. The 2SLS estimates reported in column 2 show a negative and statistically significant effect on the high school graduation rate among 18 year olds. Children who are close to finishing elementary school or tenth grade are less likely to drop out of school at least until they reach that goal. 18 year olds were around 13 years when the second phase of the program began and they were two years away from finishing tenth grade; so, receiving the transfer may not have been enough incentive to prevent

Table 2.9: Short-term effects of BDH on enrollment and years of education by the end of phase two

	Enrollment		Years of Education		
	(1)	(2)	(3)	(4)	
Age in 2014	ITT	2SLS	ITT	2SLS	N
Effect on 10 year olds	0.00597	-0.00765	-0.128	-0.899**	468
	(0.0146)	(0.0293)	(0.151)	(0.384)	
Pre- treatment mean	0.988	0.988	4.960	4.960	
Effect on 15 year olds	0.0347	0.0949	0.0247	0.445	1289
	(0.0251)	(0.0643)	(0.155)	(0.423)	
Pre- treatment mean	0.919	0.919	9.103	9.103	
Effect on 16 year olds	-0.0650**	-0.0694	-0.346**	0.398	1381
	(0.0285)	(0.0704)	(0.166)	(0.399)	
Pre- treatment mean	0.869	0.869	9.458	9.458	
Effect on 17 year olds	-0.0105	0.0480	-0.0143	1.204**	1482
	(0.0346)	(0.0952)	(0.181)	(0.512)	
Pre- treatment mean	0.728	0.728	10.11	10.11	
Effect on 18 year olds	-0.0376	-0.0708	-0.254	-0.141	1483
	(0.0386)	(0.110)	(0.190)	(0.555)	
Pre- treatment mean	0.447	0.447	10.70	10.70	
Effect on 19 year olds	0.0314	-0.224**	0.293	-0.267	1450
	(0.0375)	(0.106)	(0.197)	(0.521)	
Pre- treatment mean	0.344	0.344	10.97	10.97	
Effect on 20 year olds	-0.0205	0.193	-0.347	1.882**	1334
	(0.0371)	(0.129)	(0.223)	(0.825)	
Pre- treatment mean	0.279	0.279	11.12	11.12	
Effect on 21 year olds	-0.0421	0.0790	-0.0365	-0.158	1261
	(0.0366)	(0.175)	(0.255)	(1.204)	
Pre- treatment mean	0.256	0.256	11.20	11.20	
County and time FE	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The table reports on the results of ITT and 2SLS regressions using a linear polynomial of the distance to the cutoff. Sample is children with a score within +/-2.5 points from the Selben II cutoff (36.5 points). All the regressions control for gender and race of the individuals.

Table 2.10: Short-term effects of BDH on high school graduation and on the likelihood of having some college education by the end of phase two

	High School		Has some college		
	(1)	(2)	(3)	(4)	
Age in 2014	ITT	2SLS	ITT	2SLS	N
Effect on 17 year olds	0.00363	0.0240			1482
	(0.0292)	(0.0764)			
Pre- treatment mean	0.168	0.168			
Effect on 18 year olds	-0.0492	-0.224*			1483
	(0.0384)	(0.115)			
Pre- treatment mean	0.505	0.505			
Effect on 19 year olds	0.0178	0.0446	0.0477	-0.0516	1450
	(0.0380)	(0.102)	(0.0311)	(0.0870)	
Pre- treatment mean	0.624	0.624	0.167	0.167	
Effect on 20 year olds			-0.00727	0.381***	1334
			(0.0360)	(0.130)	
Pre- treatment mean			0.212	0.212	
Effect on 21 year olds			-0.0549	-0.115	1261
			(0.0369)	(0.177)	
Pre- treatment mean			0.248	0.248	
County and time FE	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The table reports on the results of ITT and 2SLS regressions using a linear polynomial of the distance to the cutoff. Sample is children with a score within +/-2.5 points from the Selben II cutoff (36.5 points). All the regressions control for gender and race of the individuals.

some of them from leaving school. I also present results for 17 and 19 year olds to allow for the possibility of early and late high school graduations. The effects were positive but not statistically significant for these groups.

On the other hand, column 4 shows that 20 year olds in 2014, who were 15 when the second phase began in 2008/2009 were in fact 38.1 percentage points more likely to have some college education. This is equivalent to a 180% increase relative to the pre-treatment rate.

2.6.6 Labor market outcomes

Table 2.11 reports on the short-term effects of BDH on the employment rate of 15 to 21 year olds by the end of phase two. Column 2 shows that BDH had a negative and statistically significant effect on the employment rate of 17 year olds. The probability of working drops by 12.5 percentage points for this group, relative to a pre-treatment rate of 14.3% which is equivalent to an 87% reduction. This is consistent with the results reported in Table 2.9 that showed a statistically significant increase of 1.2 years of education for this age group accompanied by a positive effect, although not statistically significant, in enrollment.

2.6.7 Differential effects of a long versus a short exposure to BDH

2.6.8 Education outcomes

In this section, I discuss the differential effects of a short exposure to BDH (treatment during phase one) versus a long exposure to BDH (treatment during phases one and two). I focus on children aged 10 to 23 years at the end of phase two, who were treated during phase one and were 5 to 18 years old when they were assigned to be treated (or not) at the beginning of phase two. Since I restrict the sample to children that received treatment for at least three years in each of the phases of the program, the final sample contains children aged 10 to 21 years by the end of phase two.

Table 2.12, shows the differential effects of BDH on enrollment and years of education by the end of phase two. Column 2 shows the results of the 2SLS regressions in equation 2.4. It shows positive differential effects on enrollment for children aged 10, 14 and 16 years in 2013/2014, who were 0, 4, and 6 years in 2003. However, none of the age groups were significantly more likely to be enrolled in an educational establishment after being treated during the two phases of the program when compared to similar children that were treated only during the first phase.

Column 4 shows the results for years of education. The general results are not

Table 2.11: Short-term effects of BDH on the employment rate by the end of phase two

	Has a job		
	(1)	(2)	
Age in 2014	ITT	2SLS	N
Effect on 15 year olds	-0.0141	0.0528	1289
	(0.0176)	(0.0461)	
Pre- treatment mean	0.0480	0.0480	
Effect on 16 year olds	0.0163	-0.0117	1381
	(0.0207)	(0.0531)	
Pre- treatment mean	0.0770	0.0770	
Effect on 17 year olds	0.0163	-0.125*	1482
	(0.0269)	(0.0726)	
Pre- treatment mean	0.143	0.143	
Effect on 18 year olds	0.00301	-0.0260	1483
	(0.0371)	(0.107)	
Pre- treatment mean	0.311	0.311	
Effect on 19 year olds	0.0173	0.0928	1450
	(0.0396)	(0.107)	
Pre- treatment mean	0.440	0.440	
Effect on 20 year olds	0.0453	-0.165	1334
	(0.0421)	(0.150)	
Pre- treatment mean	0.523	0.523	
Effect on 21 year olds	-0.0575	0.240	1261
	(0.0427)	(0.218)	
Pre- treatment mean	0.587	0.587	
County and time FE	Yes	Yes	
Controls	Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The table reports on the results of ITT and 2SLS regressions using a linear polynomial of the distance to the cutoff. Sample is children with a score within +/-2.5 points from the Selben II cutoff (36.5 points). All the regressions control for gender and race of the individuals.

Table 2.12: Differential effects of BDH on enrollment and years of education

	Enrollment		Years of Education		
	(1)	(2)	(3)	(4)	
Age in 2014	ITT	2SLS	ITT	2SLS	N
Effect on 10 year olds	-0.00379	0.0166	0.00307	0.280**	1787
	(0.00650)	(0.0110)	(0.0609)	(0.114)	
Pre- treatment mean	0.988	0.988	4.960	4.960	
Effect on 11 year olds	0.00869*	-0.00686	0.0399	-0.159*	3401
	(0.00447)	(0.00788)	(0.0479)	(0.0927)	
Pre- treatment mean	0.987	0.987	5.917	5.917	
Effect on 12 year olds	0.00381	-0.00729	-0.0550	-0.0670	7002
	(0.00447)	(0.00811)	(0.0406)	(0.0759)	
Pre- treatment mean	0.980	0.980	6.817	6.817	
Effect on 13 year olds	0.00397	-0.0147	0.000286	0.0401	7470
	(0.00557)	(0.0103)	(0.0484)	(0.0890)	
Pre- treatment mean	0.971	0.971	7.677	7.677	
Effect on 14 year olds	-0.00218	0.0119	0.0316	0.0613	7629
	(0.00731)	(0.0142)	(0.0528)	(0.100)	
Pre- treatment mean	0.949	0.949	8.481	8.481	
Effect on 15 year olds	-0.0130	-0.00875	8.24e-05	-0.0803	6767
	(0.00923)	(0.0183)	(0.0631)	(0.117)	
Pre- treatment mean	0.919	0.919	9.103	9.103	
Effect on 16 year olds	-0.0257**	0.0353	-0.165**	-0.0312	7084
	(0.0112)	(0.0229)	(0.0679)	(0.140)	
Pre- treatment mean	0.869	0.869	9.458	9.458	
Effect on 17 year olds	-0.0108	-0.0239	-0.0563	-0.0606	6678
	(0.0156)	(0.0320)	(0.0752)	(0.151)	
Pre- treatment mean	0.728	0.728	10.11	10.11	
Effect on 18 year olds	0.0139	-0.0135	-0.106	-0.217	5734
	(0.0187)	(0.0386)	(0.0878)	(0.182)	
Pre- treatment mean	0.447	0.447	10.70	10.70	
Effect on 19 year olds	-0.0189	-0.0289	0.0275	-0.0322	5409
	(0.0185)	(0.0426)	(0.0933)	(0.211)	
Pre- treatment mean	0.344	0.344	10.97	10.97	
Effect on 20 year olds	0.0301*	-0.0613	0.0137	-0.291	5127
	(0.0176)	(0.0411)	(0.103)	(0.234)	
Pre- treatment mean	0.279	0.279	11.12	11.12	
Effect on 21 year olds	0.0209	-0.0501	-0.0429	-0.182	4605
	(0.0183)	(0.0434)	(0.118)	(0.281)	
Pre- treatment mean	0.256	0.256	11.20	11.20	
County and time FE	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The table reports on the results of ITT and 2SLS regressions using a linear polynomial of the distance to the cutoff. Sample is children with a score within +/-2.5 points from the Selben II cutoff (36.5 points). All the regressions control for gender and race of the individuals.

very encouraging either. The only group that experienced an increase in years of schooling were 10 year olds. These children had not reached the first year of age when the program started in 2003. By the end of phase two, they increased their years of education in 0.28 years, relative to a pre-treatment mean of 4.96 years for that age group, which is equivalent to a 6% increase in schooling. These children were exposed to BDH for 10 years so it is likely that the benefits received during their first 5 years of life were spent in food and health checkups, while during the remaining five years they were used to cover education-related expenses. These results are in line with the evidence on the complementarity of early investments that argue that early investments facilitate the productivity of later investments (Cunha, Heckman, Lochner, et al., 2006).

The differential effects of BDH on high school graduation and on the likelihood of having some college education are reported in Table 2.13. There were no statistically significant gains in terms of the likelihood of graduating from high school for 17 to 19 year olds. The reason could be that these children were between 12 and 14 years when the second phase began. One possible explanation is that because education is compulsory up to age 15 they were less likely to drop out from school and after education became free, the outcomes of those who were not treated in phase two were at least as good as those of children treated during the two phases. Likewise, there were no significant effects on the probability of having some college. This probability was even negative for 21 year olds, probably because households treated during the two phases are the most vulnerable ones.

2.6.9 Labor market outcomes

The last wave of the RS collects information on the occupation sector in which people work. This allows me to classify the different sectors into two general groups, agricultural work and white-collar jobs. In this way, it is possible to investigate the

Table 2.13: Differential effects of BDH on high school graduation and on the likelihood of having some college education

	High School		Has some college		
	(1)	(2)	(3)	(4)	
Age in 2014	ITT	2SLS	ITT	2SLS	N
Effect on 17 year olds	0.0101	-0.0201			6678
	(0.0122)	(0.0251)			
Pre- treatment mean	0.168	0.168			
Effect on 18 year olds	-0.0151	0.00686			5734
	(0.0184)	(0.0379)			
Pre- treatment mean	0.505	0.505			
Effect on 19 year olds	-0.00511	0.0155	-0.00358	0.0206	5409
	(0.0187)	(0.0430)	(0.0142)	(0.0331)	
Pre- treatment mean	0.624	0.624	0.167	0.167	
Effect on 20 year olds			0.0152	-0.0444	5127
			(0.0155)	(0.0362)	
Pre- treatment mean			0.212	0.212	
Effect on 21 year olds			-0.00815	-0.0845**	4605
			(0.0183)	(0.0426)	
Pre- treatment mean			0.248	0.248	
County and time FE	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The table reports on the results of ITT and 2SLS regressions using a linear polynomial of the distance to the cutoff. Sample is children with a score within +/-2.5 points from the Selben II cutoff (36.5 points). All the regressions control for gender and race of the individuals.

effects of BDH not only on the probability that young adults work but also on the type of work they have.

In general, being treated during the two phases of the program reduced the likelihood of having a job among the younger age groups. However, this effect was not statistically significant. Similarly, there were no significant effects on the probability of having a white collar job or a job in the agricultural sector as shown in Table 2.14. These results are consistent with a recent study by Hahn, Islam, et al. (2018) that examines the long-term effects of a stipend program that made secondary education free for rural girls in Bangladesh and find no apparent effect of the stipend on the

Table 2.14: Differential effects of BDH on the employment rate

	Has a job		Agriculture job		White collar job		
	(1)	(2)	(3)	(4)	(5)	(6)	
Age in 2014	ITT	2SLS	ITT	2SLS	ITT	2SLS	N
Effect on 15 year olds	0.00515	-0.00410	0.0192	-0.183	-0.0188	-0.0783	6767
	(0.00719)	(0.0139)	(0.0823)	(0.126)	(0.0900)	(0.148)	
Pre- treatment mean	0.0480	0.0480	0.293	0.293	0.233	0.233	
Effect on 16 year olds	0.0141	-0.0109	0.0150	0.105	-0.143**	-0.0461	7084
	(0.00861)	(0.0175)	(0.0493)	(0.0883)	(0.0707)	(0.107)	
Pre- treatment mean	0.0770	0.0770	0.239	0.239	0.256	0.256	
Effect on 17 year olds	-0.00332	0.00744	0.00185	0.0278	-0.000628	0.0194	6678
	(0.0124)	(0.0251)	(0.0361)	(0.0730)	(0.0450)	(0.0913)	
Pre- treatment mean	0.143	0.143	0.194	0.194	0.270	0.270	
Effect on 18 year olds	-0.00152	0.0275	0.0231	-0.0398	0.00414	0.0290	5734
	(0.0170)	(0.0353)	(0.0244)	(0.0452)	(0.0340)	(0.0614)	
Pre- treatment mean	0.311	0.311	0.151	0.151	0.317	0.317	
Effect on 19 year olds	0.0128	-0.00477	0.00411	0.0377	0.0313	-0.0539	5409
	(0.0183)	(0.0425)	(0.0189)	(0.0425)	(0.0287)	(0.0666)	
Pre- treatment mean	0.440	0.440	0.144	0.144	0.337	0.337	
Effect on 20 year olds	0.00931	0.0221	0.00679	-0.0381	0.0432	0.0108	5127
	(0.0183)	(0.0422)	(0.0187)	(0.0408)	(0.0280)	(0.0624)	
Pre- treatment mean	0.523	0.523	0.129	0.129	0.362	0.362	
Effect on 21 year olds	-0.00188	0.0333	0.0402**	0.00300	-0.0612**	0.0633	4605
	(0.0188)	(0.0439)	(0.0168)	(0.0402)	(0.0275)	(0.0626)	
Pre- treatment mean	0.587	0.587	0.127	0.127	0.377	0.377	
County and time FE	Yes	Yes	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The table reports on the results of ITT and 2SLS regressions using a linear polynomial of the distance to the cutoff. Sample is children with a score within +/-2.5 points from the Selben II cutoff (36.5 points). All the regressions control for gender and race of the individuals.

likelihood of women working in adulthood.

2.7 Robustness checks

2.7.1 Results using ENES exam data

Considering that an important number of counties were left out of the third wave of the RS, in this section, I use administrative data on the ENES exam (Examen Nacional para la Educación Superior) to test the robustness of the results regarding high school graduation. ENES is a standardized exam that students in the last year of high school have to pass in order to go to university. It is compulsory for all students at private and public schools; hence, taking the test is a good predictor for high school graduation.

The exam was first administered in 2012, so it is possible to merge the 2013 and 2014 ENES databases to the last wave of my three-wave panel using the students' IDs. Of the 336,791 students in the ENES dataset, 168,481 were also in the RS. As in the previous section, I run similar 2SLS regressions to the ones I estimated before using as the outcome of interest a binary variable that takes the value of one if the child was on the ENES dataset, which implies that she was in the final year of high school, and zero if not.

As shown in Table 2.10, being treated during the second phase of the program significantly reduced the probability of graduating from high school among 18-year-olds by 22.4 percentage points, whereas when using the ENES exam data, the result was a reduction of 13.4 percentage points and was not statistically significant, as shown in column 2 Table 2.17 in the Appendix. On the other hand, Appendix Table 2.18 reports on the results of the differential effects of a long versus a short exposure to BDH on the probability of high school graduation. The results were in line to those reported in Table 2.13 that showed no statistically significant effects for 17, 18 and 19 year olds.

2.7.2 Effects on fertility and migration

There are several factors that may affect children outcomes apart from the transfer. One concern is that CTs could incentivize other behaviours like an increase in fertility or migration. Poor women may choose to have children to become eligible to receive the transfer or may choose to migrate to places with public offices where cashing the transfers is easier. Changes in household composition may affect the schooling decisions of poor children since a new family member implies more expenses, forcing some children out of school. On the other hand, remittances from migrants may also affect schooling decisions but in the opposite direction. I explore whether BDH had effects on fertility and migration in the short and long-run. As a proxy for fertility, I focus on the total number of children below 12 years living in the household. To study the impacts on migration, I focus on the number of household members who migrated since 2000 (a question that is asked in waves 2 and 3). Results shown in Table 2.19 suggest that there is little evidence of endogenous fertility induced by the transfer in the short or long-run for any of the age groups except for children ages 8 and 9 years for whom I found a statistically significant effect in the short-run. Finally, there is no evidence of impacts on migration in the short or long-run.

2.8 Conclusions

This paper studies whether cash transfers continue to be effective after several years targeting the same population. The availability of individual-level social registry data allowed me to identify children that were treated during one or the two phases of the program and track their performance in terms of education and labor market outcomes. With this information and knowing that at the threshold of eligibility, the second assignment to treatment (in 2008/9) was independent of the first assignment (in 2003), I was able to disentangle the impact of a short exposure to BDH (treatment during phase one) versus a long exposure (treatment during phases one and two).

The short-term effects of BDH – measured after five years of its implementation – were positive and statistically significant on enrollment among children that began treatment at ages 8, 9, 12, 13, and 15. Children were also more likely to graduate from high school and to have some undergraduate studies. The results by the end of phase two (among children who were treated only during phase two) also showed positive effects on years of education among children who began treatment at ages 12 and 15. The lack of short-term effects in enrollment by the end of phase two may be explained by the fact that soon after implementation, people realized that the authorities were not monitoring the conditions.

The lack of differential effects on education and labor market outcomes among children that were treated during the two phases versus those treated during just the first phase, is not explained by an attenuation of BDH effects for all the age groups in phase two. In fact, the analysis of the short-term effects at the end of phase two showed positive and important effects of BDH at juncture ages. This contradicts the hypotheses that educational gratuity or lack of monitoring of the conditions attenuated the effects of BDH during phase two. A more plausible reason for the lack of differential effects is that once children reach the education level they have planned to achieve, or alternatively, once they reach a certain age when the opportunity cost of schooling becomes sufficiently high, an unconditional transfer does not provide enough incentive to keep them at school.

The only group that experienced positive differential effects in years of education after being exposed during the two phases of the program were children aged 0 when BDH began in 2003. Children who began treatment at an older age did not experience similar gains despite being treated with similar intensity possibly because once cognitive gaps appear, the process cannot be reversed with CTs. It is likely that older children were already lagging behind perhaps with low grades and attendance problems, which explains why BDH did not have positive effects on their education. These results are in line with the notion that human capital investment exhibits both self-productivity and complementarity and explains why late investments that attempt to compensate for the lack of adequate early investments can be very costly and ineffective (Cunha, Heckman, Lochner, et al., 2006).

Regarding labor market outcomes in the short-run, results were not conclusive about whether the negative effect on the likelihood of having a job for 17 year olds in 2013/2014 was caused by a concurrent raise in enrollment for this age group. Moreover, being exposed to BDH for two phases versus just one did not give treated children an advantage in the labor market by the end of phase two. The reason could be that children exposed during the two phases of the program did not achieve more years of education after all. It is possible that people treated during the two phases could be more vulnerable, and that other things that are not captured by the Selben score like social networks or lack thereof, important to access the job market, could be attenuating the effects of BDH on labor market outcomes.

The lack of monitoring of the conditions is one of the factors that may explain why BDH did not achieve its goal of improving educational attainment consistently in the long-run. Had the transfer been conditioned on school registration and on grade progression it is likely that better results would have been found in terms of years of education. In a study for Ecuador, Schady et al. (2008) found that the short-term gains from BDH were significantly larger among households who believed that there was a school enrollment requirement attached to transfers. Evidence from micro-simulation models for Mexico and Brazil also conclude that conditions attached to transfers explain the bulk of the effect of CCT programs on school enrollment (Ferreira and Leite, 2002; Todd and Wolpin, 2006). Another reason for the lack of lasting effects is that the transfer was not big enough to compensate for the wages that older children could get in the labor market. Finally, when there are other children at home, there is no way to prevent parents from spending the transfer on the older children instead of those below 18 years. This would also cause an attenuation of the program's impact.

Looking strictly at the effects of BDH on education and labor market outcomes, the results from this paper stress the need for a redesign of BDH. Attanasio et al. (2012) argue that for the case of PROGRESA, a revenue neutral change that increases the grants for secondary school children while eliminating it for primary school children would have positive effects on enrollment on the latter and minor effects on the former. I would expect similar results for the case of BDH considering that in Ecuador education is compulsory until age 15. Transfers should also take into account the number of children in the household and should increase with age in order to reduce the opportunity cost of work for children aged 15 to 18 years. Furthermore, given the results found in this paper, the government should also increase the transfers to poor mothers with children below 5 years.

Finally, it is worth noting that even when BDH became inefficient in the long-run in improving the educational outcomes of children, it does not mean necessarily that this applies also to health outcomes. Future research should look at the long-term effects of BDH on other outcomes to inform any reform to the program.

References

Aizer, By Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney (2016). “The Long-Run Impact of Cash Transfers to Poor Families”. In: *American Economic Review* 106.4, pp. 935–971.

Araujo, M. Caridad, Mariano Bosch, and Norbert Schady (2016). “Can Cash Transfers Help Households Escape an Inter-Generational Poverty Trap?” In: *NBER Working Paper Series*. NBER Working Paper Series, pp. 1–30.

Attanasio, Orazio P., Costas Meghir, and Ana Santiago (2012). “Education choices in Mexico: Using a structural model and a randomized experiment to evaluate PROGRESA”. In: *Review of Economic Studies* 79.1, pp. 37–66.

Baez, Javier E and Adriana Camacho (2011). “Assessing the Long-term Effects of Conditional Cash Transfers on Human Capital Evidence from Colombia”.

Baird, Sarah, Francisco H G Ferreira, Berk Özler, and Michael Woolcock (2014). “Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes”. In: *Journal of Development Effectiveness* 6.1, pp. 1–43.

Barham, By Tania, Karen Macours, and John A. Maluccio (2013a). “Boys’ Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages”. In: *American Economic Review* 103.3, pp. 467–471.

Barham, Tania, Karen Macours, and John a Maluccio (2013b). “More Schooling and More Learning? Effects of a Three-Year Conditional Cash Transfer Program after 10 Years”.

Behrman, Jere R., Susan W. Parker, and Petra E. Todd (2011). “Do Conditional Cash Transfers for Schooling Generate Lasting Benefits?: A Five-Year Followup of PROGRESA/Oportunidades”. In: *Journal of Human Resources* 46.1, pp. 203–236.

Behrman, Jere R., Piyali Sengupta, and Petra Todd (2005). “Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico”. In: *Economic Development and Cultural Change* 54.1, pp. 237–275.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen (2015). “Turning a shove into a nudge? A "labeled cash transfer" for education”. In: *American Economic Journal: Economic Policy* 7.3, pp. 1–48.

Buser, Thomas (2015). “The Effect of Income on Religiousness”. In: *American Economic Journal: Applied Economics* 7.3, pp. 178–195.

Carneiro, Pedro, James J Heckman, David Bravo, Partha Dasgupta, Steve Levitt, Lance Lochner, Costas Meghir, Kathleen Mullen, and Casey Mulligan (2002). “The Evidence on Credit Constraints in Post-Secondary Schooling”. In: *The Economic Journal* 112, pp. 705–734.

Carrillo, Paul E. and Juan Ponce Jarrín (2009). “Efficient delivery of subsidies to the poor: Improving the design of a cash transfer program in Ecuador”. In: *Journal of Development Economics* 90.2, pp. 276–284.

Cattaneo, Matias D, Michael Jansson, and Xinwei Ma (2016). “rddensity : Manipulation Testing based on Density Discontinuity”. In: *The Stata Journal* ii, pp. 1–18.

Cunha, Flavio, James J Heckman, and Susanne M Schennach (2010). “Estimating the Technology of Cognitive and Noncognitive Skill Formation”. In: *Econometrica* 78.3, pp. 883–931.

Cunha, Flavio, James Heckman, Lance Lochner, and Dimitriy Masterov (2006). "Chapter 12. Interpreting the Evidence on Life Cycle Skill Formation". In: *Handbook of the Economics of Education*. Vol. 1, pp. 697–812.

De Brauw, Alan and John Hoddinott (2011). "Must conditional cash transfer programs be conditioned to be effective? The impact of conditioning transfers on school enrollment in Mexico". In: *Journal of Development Economics* 96.2, pp. 359–370.

De Leeuw, J and J Van Rijckevorsel (1980). "HOMALS and PRINCALS: Some generalizations of principal components analysis." In: *In E. Diday et al., eds. Data analysis and informatics, pp. 231-241. Amsterdam, Netherlands, North-Holland.*

Edmonds, Eric V. and Norbert Schady (2012). "Poverty alleviation and child labor". In: *American Economic Journal: Economic Policy* 4.4, pp. 100–124.

Fabara, Cristina (2009). *Reformulacion del indice de clasificacion socioeconomica del Registro Social*. Tech. rep. Quito-Ecuador: Ministerio Coordinador de Desarrollo Social, pp. 1–11.

Ferreira, Francisco and Phillippe Leite (2002). "Ex-Ante Evaluation of Conditional Cash Transfer Programs: The Case of Bolsa Escola". Washington, DC.

Fiszbein, Ariel, Norbert Schady, Francisco Ferreira, Margaret Grosh, Nial Kelleher, Pedro Olinto, and Emmanuel Skoufias (2009). "Conditional Cash Transfers. Reducing present and future poverty". In: *The World Bank*, pp. 1–361.

García, María Isabel, Natalia Garzón, Carolina Patiño, Drichelmo Tamayo, Ana María Grijalva, and Juan Carlos Palacios (2016). *Panorama Laboral y Empresarial del Ecuador 2016*. Tech. rep. Quito - Ecuador: INEC, p. 55.

García-Aracil, Adela and Carolyn Winter (2006). “Gender and ethnicity differentials in school attainment and labor market earnings in Ecuador”. In: *World Development* 34.2, pp. 289–307.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw (2001). “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design”. In: *Econometrica* 69.1, pp. 201–209.

Hahn, Youjin, Asadul Islam, Kanti Nuzhat, Russell Smyth, and Hee-Seung Yang (2018). “Education, Marriage, and Fertility: Long-Term Evidence from a Female Stipend Program in Bangladesh”. In: *Economic Development and Cultural Change* 8.

Heckman, James J (2000). “Policies to foster human capital”. In: *Research in Economics* 54.December 1998, pp. 3–56.

Hoynes, Hilary W, Diane Whitmore Schanzenbach, and Douglas Almond (2016). “Long Run Impacts of Childhood Access to the Safety Net”. In: *American Economic Review* 106.4, pp. 903–934.

Janvry, Alain de, Frederico Finan, Elisabeth Sadoulet, and Renos Vakis (2006). “Can conditional cash transfer programs serve as safety nets in keeping children at school and from working when exposed to shocks?” In: *Journal of Development Economics* 79.2, pp. 349–373.

King, Elizabeth M. and Jere R. Behrman (2009). “Timing and duration of exposure in evaluations of social programs”. In: *World Bank Research Observer* 24.1, pp. 55–82.

Kugler, Adriana and Ingrid Rojas (2018). “Do CCTs improve employment and earnings in the very long-term? Evidence from Mexico”. Cambridge, MA.

Lee, David S and Thomas Lemieux (2010). “Regression Discontinuity Designs in Economics”. In: *Journal of Economic Literature* 48.June, pp. 281–355.

Ponce, J and F Falconí (2011). *El trabajo infantil en Ecuador: marco institucional, evolución histórica y análisis costo beneficio de su erradicación*. Ed. by Gabriela Malo. Quito - Ecuador, pp. 1–87.

Ponce, Juan and Arjun S. Bedi (2010). “The impact of a cash transfer program on cognitive achievement: The Bono de Desarrollo Humano of Ecuador”. In: *Economics of Education Review* 29.1, pp. 116–125.

Saavedra, Juan Esteban and Sandra Garcia (2012). “Impacts of Conditional Cash Transfer Programs on Educational Outcomes in Developing Countries: A Meta-Analysis”.

Schady, Norbert, Maria Araujo, Ximena Peña, and Luis López-Calva (2008). “Cash Transfers, Conditions, and School Enrollment in Ecuador”. In: *Economía* 8.2, pp. 43–77.

Skoufias, Emmanuel and Susan Wendy Parker (2001). “Conditional Cash Transfers and Their Impact on Child Work and Schooling: Evidence from the PROGRESA Program in Mexico”. In: *Economía* 2.1, pp. 45–86.

Todd, Petra E and Kenneth I Wolpin (2006). “Assessing the Impact of a School Subsidy Program in Mexico : Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility”. In: *The American Economic Review* 96.5, pp. 1384–1417.

World Bank (2005). *Project information document (PID)*. Tech. rep. Washington DC: (AB1915). The World Bank.

2.9 Appendix

Table 2.15: Selective matching and BDH eligibility

Ages	First Wave	Second Wave	Third Wave
0 and 1 year olds	0.0106		
	(0.00798)		
N	29886		
Mean	0.388		
2 and 3 year olds	-0.00136		
	(0.00800)		
N	29218		
Mean	0.383		
4 and 5 year olds	0.00143		
	(0.00826)		
N	27783		
Mean	0.385		
6 and 7 year olds	0.00342	-0.00181	
	(0.00707)	(0.00664)	
N	37439	39476	
Mean	0.385	0.377	
8 and 9 year olds	0.00499	0.00357	
	(0.00716)	(0.00677)	
N	36551	39518	
Mean	0.388	0.459	
10 and 11 year olds	0.00821	0.00794	0.00184
	(0.00728)	(0.00711)	(0.00527)
N	35711	35999	51686
Mean	0.389	0.456	0.253
12 and 13 year olds	0.00937	0.00465	0.00630
	(0.00712)	(0.00704)	(0.00560)
N	37580	35826	55086
Mean	0.392	0.421	0.648
14 and 15 year olds	0.00829	0.0106	-0.00219
	(0.00720)	(0.00723)	(0.00556)
N	36146	33618	50046
Mean	0.385	0.388	0.710
16 and 17 year olds	0.00732	-0.00418	-0.00381
	(0.00694)	(0.00747)	(0.00583)
N	38343	30019	45604
Mean	0.386	0.356	0.711
18 and 19 year olds	-0.00492	-0.00561	0.0135**
	(0.00741)	(0.00794)	(0.00635)
N	33554	26732	39267
Mean	0.392	0.345	0.694
20 and 21 year olds	-0.00185	0.0106	0.00203
	(0.00757)	(0.00806)	(0.00700)
N	32022	24884	34839
Mean	0.389	0.310	0.647

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample includes children in each wave of RS data within +/- 2.5 points of the eligibility thresholds in each wave. The dependent variable is a dummy equals to 1 if a child is in the 3-waves panel and 0 otherwise. Models include county, year of survey FE. Control covariates include child's gender, age. Each coefficient comes from a separate regression by age group.

Table 2.16: Short-term effects of BDH on enrollment and years of education by the end of phase one

	Enrollment		Years of Education		
	(1)	(2)	(3)	(4)	
Age in 2008	ITT	2SLS	ITT	2SLS	N
Effect on 6 year olds	-0.00452	0.00560	-0.0636	0.103	2925
	(0.00683)	(0.0133)	(0.0493)	(0.0836)	
Pre- treatment mean	0.984	0.984	1.449	1.449	
Effect on 7 year olds	0.00731	0.00610	-0.00226	0.0771	5660
	(0.00489)	(0.00875)	(0.0346)	(0.0655)	
Pre- treatment mean	0.990	0.990	2.344	2.344	
Effect on 8 year olds	-0.00179	-0.00564	0.0653*	0.0843	6326
	(0.00442)	(0.00739)	(0.0353)	(0.0648)	
Pre- treatment mean	0.991	0.991	3.264	3.264	
Effect on 9 year olds	0.00800**	-0.00115	0.0198	0.0421	6038
	(0.00354)	(0.00536)	(0.0394)	(0.0699)	
Pre- treatment mean	0.993	0.993	4.221	4.221	
Effect on 10 year olds	-0.00518	0.0130	0.00678	0.212**	5561
	(0.00422)	(0.00836)	(0.0468)	(0.0886)	
Pre- treatment mean	0.992	0.992	5.160	5.160	
Effect on 11 year olds	-0.00207	-0.00215	0.0485	-0.00706	6411
	(0.00507)	(0.00929)	(0.0497)	(0.0846)	
Pre- treatment mean	0.980	0.980	5.992	5.992	
Effect on 12 year olds	-0.00904	0.0161	0.0140	0.0577	7009
	(0.00601)	(0.0116)	(0.0510)	(0.0905)	
Pre- treatment mean	0.963	0.963	6.796	6.796	
Effect on 13 year olds	0.000741	0.0297**	-0.0253	0.0744	7135
	(0.00849)	(0.0151)	(0.0583)	(0.103)	
Pre- treatment mean	0.942	0.942	7.520	7.520	
Effect on 14 year olds	0.00231	0.0383**	0.00406	0.0674	6649
	(0.00950)	(0.0179)	(0.0670)	(0.121)	
Pre- treatment mean	0.923	0.923	8.233	8.233	
County and time FE	Yes	Yes	Yes	Yes	
Controls	Yes	Yes	Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The table reports on the results of ITT and 2SLS regressions using a linear polynomial of the distance to the cutoff. Sample is children with a score within +/-2.5 points from the Selben I cutoff (50,65 points). Models include county and year of survey FE. Control covariates include child's gender, age.

Table 2.17: Short-term effects of BDH by the end of phase two using ENES data

	Enes exam		
	(1)	(2)	
Age in 2014	ITT	2SLS	N
Effect on 17 year olds	-0.0313	0.0559	1482
	(0.0376)	(0.100)	
Effect on 18 year olds	-0.0251	-0.134	1483
	(0.0393)	(0.112)	
Effect on 19 year olds	-2.21e-05	0.0478	1450
	(0.0365)	(0.0986)	
County and time FE	Yes	Yes	
Controls	Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Table reports the coefficients of ITT and 2SLS regressions. Sample is children who took the ENES exam and has a Selben score within +/-2.5 points from the Selben II cutoff (36.5 points) who were untreated on phase 1. All the regressions control for gender and race of the individuals.

Table 2.18: Differential effects of a long versus a short exposure measured by the end of phase two using ENES data

	Enes exam		
	(1)	(2)	
Age in 2014	ITT	2SLS	N
Effect on 17 year olds	-0.0352**	-0.0239	6678
	(0.0172)	(0.0352)	
Effect on 18 year olds	0.0198	-0.0353	5734
	(0.0188)	(0.0387)	
Effect on 19 year olds	0.00128	-0.0666	5409
	(0.0180)	(0.0421)	
County and time FE	Yes	Yes	
Controls	Yes	Yes	

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. P-values from goodness of fit test after standard errors. Table reports the coefficients of ITT and 2SLS regressions. Sample is children who took the ENES exam and has a Selben score within +/-2.5 points from the Selben II cutoff (36.5 points) who were treated on phase 1. All the regressions control for gender and race of the individuals.

Table 2.19: Short-term effects and differential effects of BDH on fertility and migration

	Short-term effects			Differential effects		
	(1)	(2)		(3)	(4)	
Ages	Children <=12 years	Migrants	N	Children <=12 years	Migrants	N
Effect on 6 and 7 year olds	0.0289	-0.00776	8587			
	(0.0666)	(0.0177)				
Pre- treatment mean	2.325	0.0420				
Effect on 8 and 9 year olds	-0.158***	0.0123	12364			
	(0.0543)	(0.0155)				
Pre- treatment mean	2.331	0.0510				
Effect on 10 and 11 year olds	-0.0610	-0.00864	11973	0.0506	-0.0833	5188
	(0.0520)	(0.0150)		(0.0887)	(0.0574)	
Pre- treatment mean	2.166	0.0460		2.207	0.0310	
Effect on 12 and 13 year olds	-0.0643	-0.00764	14147	0.0862	-0.0191	14473
	(0.0500)	(0.0128)		(0.0539)	(0.0140)	
Pre- treatment mean	1.453	0.0530		1.680	0.0310	
Effect on 14 and 15 year olds	0.00519	-0.000401	13116	-0.00412	0.0137	14397
	(0.0488)	(0.0126)		(0.0517)	(0.0156)	
Pre- treatment mean	0.961	0.0530		1.173	0.0350	
Effect on 16 and 17 year olds	-0.0809	-0.00661	11321	0.00904	-0.0424	13762
	(0.0546)	(0.0162)		(0.0534)	(0.0285)	
Pre- treatment mean	0.921	0.0590		1.034	0.0380	
Effect on 18 and 19 year olds	-0.0706	-0.0138	10463	0.00849	-0.00367	11143
	(0.0644)	(0.0185)		(0.0638)	(0.0185)	
Pre- treatment mean	0.911	0.0650		0.971	0.0340	
Effect on 20 and 21 year olds	0.0692	-0.0227	9071	0.0578	-0.00140	9742
	(0.0743)	(0.0230)		(0.0707)	(0.0157)	
Pre- treatment mean	0.953	0.0620		0.969	0.0320	
County and time FE	Yes	Yes		Yes	Yes	
Controls	Yes	Yes		Yes	Yes	

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The table reports on the results of 2SLS regressions using a linear polynomial of the distance to the cutoff. Sample is children with a score within +/-2.5 points from the Selben I cutoff (50,65 points) for the short-term effects and within +/-2.5 points from the Selben II cutoff (36,5 points) for the differential effects. Models include county, year of survey FE. Control covariates include child's gender, age. For columns 1 and 2 the age is reported by the end of phase1 and for columns 3 and 4 age is reported by the end of phase 2.

Chapter 3

LABOR MARKET REGULATION AND HOSPITAL QUALITY

Abstract

I evaluate the impact on in-hospital mortality of a reform that made all health professionals working part-time switch to full-time contracts at public hospitals in Ecuador. I take advantage of the staggered adoption of the reform and hospital panel data to implement an event study to evaluate the impact of the reform. The results for the sample of admissions to the emergency department show that mortality in public hospitals decreased by 0.1% on the adoption year and by 0.2% one year later. Results were robust to several alternative specifications and to the inclusion of pre-reform hospital characteristics that could have been used by policymakers to decide the order of implementation. More importantly, I show that the effects reported in this paper cannot be attributed to changes in other quality indicators at the hospital level like the length of stay or by changes in the patient mix.

3.1 Introduction and literature review

Recent studies in developed countries suggest that there is a trend of declining working hours of doctors (Swami et al., 2017; Broadway et al., 2017).[1] In response to this phenomenon and a growing demand for healthcare services, several countries have implemented policies that operate in the intensive margin to increase working hours and improve health care coverage. In the United Kingdom, for instance, the government launched a 7-day opening of General Practitioner (GP) practices to improve primary care access for patients (Dolton and Pathania, 2016), and in Australia, the government provides financial incentives for GPs to offer additional after-hours coverage to patients (Swami et al., 2017).

Hiring on-call staff remains the more common response to sudden surges in demand. The problem with on-call staff is that when these professionals are brought into the hospital on very short notice; they are likely to be unfamiliar with the procedures, practices, equipment and their colleagues, and are therefore expected to have worse results than regular staff (Bartel et al., 2014). Furthermore, evidence from stated preference studies suggests that GPs would pay a large proportion of their annual income just to avoid on-call work, which suggests that policies that operate on the intensive margin would have little impact on increasing the total hours worked by GPs (Broadway et al., 2017).

This paper studies the effects on hospital mortality of a labor reform that increased the number of hours worked by doctors at public health facilities in Ecuador. The reform replaced progressively all existing part-time contracts with full-time contracts and affected all healthcare professionals (excluding those in administrative jobs). There are several reasons to believe that this reform would reduce in-hospital mortality. First, it is reasonable to expect that higher staff-patient ratios would lead to a

[1] Some of the reasons for this decline are a shift in preference for working predictable and fewer hours and increased participation of women in the medical workforce

reduction in hospital mortality, in fact, much of the early literature on the relationship between staffing levels and hospital quality finds a negative correlation between staffing levels and patient mortality, measured as risk-adjusted 30-day mortality and failure to rescue rates (Aiken et al., 2017; Needleman, Buerhaus, Matteke, et al., 2002; Needleman, Buerhaus, Pankratz, et al., 2011). Second, the reduction in the number of part-time and on-call contracts is expected to reduce disruptions to team functioning which in turn would reduce mortality (Bartel et al., 2014).

However, other studies have found evidence of diminishing returns to nurse staffing improvements. One example is the study by Sochalski et al. (2008) that uses within-hospital variation over time to assess whether hospitals that increased nurse staffing levels saw improvements in mortality among acute myocardial infarction patients. The results show diminishing returns to nurse staffing improvements and lack of significant evidence that staffing reduces mortality. This "flat of the curve" argument that emphasizes inefficiency of medical spending has also been found by Fisher, Wennberg, et al. (2003) who argue that a 30 percent reduction in Medicare spending in high-use regions would not have negative effects on patient outcomes or satisfaction. Hence, it is not clear whether increases in staffing would necessarily translate into better outcomes for patients, a fact that is corroborated by several empirical studies that have found mixed results, which I discuss briefly below.

The main challenges to uncover a causal relationship between changes in staffing levels and in-hospital mortality reside on omitted variable bias and selection of sicker patients to the best hospitals. Even after controlling for severity of the conditions, if the unobserved component of the severity of illness affects a patient's choice of hospital, then the selection problem remains and this endogenous sorting would lead to underestimation of the beneficial effect of higher staffing levels (Cook et al., 2012). To limit these concerns, other studies have used hospital-level panel data that control for time-invariant unobserved hospital characteristics, minimizing the omitted variable problem in cross-sectional studies.

A natural solution to the selection problem is to randomly assign patients to hospitals with different staffing levels and then compare the outcomes of patients. Since this is not possible, researchers have to use natural experiments that create exogenous shifts to hospital staffing and use that variation to study the impact on patients' outcomes. Cook et al. (2012) for example, evaluate the impact of California Assembly Bill 394, which mandated maximum levels of patients per nurse in the hospital setting. They instrument the observed patient-nurse ratios by the difference between the required ratio and the ratio observed before the reform. They find that the reform decreased patient/nurse ratios but did not improve patients' outcomes, as opposed to other studies that use the same IV strategy to analyze the effects of similar reforms in the US, but register significant effects on quality of care (Lin, 2014; Tong, 2011).[2]

Matsudaira (2014) also evaluates the case for minimum nurse-to-patient ratios at nursing homes. He uses a difference-in-difference strategy that compares changes in quality of care for firms with varying degrees of exposure to the minimum staffing law and finds no impact of the regulation on overall facility quality. Finally, Evans and Kim (2006) exploits natural variation in patient loads registered at particular days in the week and find that patients admitted when patient loads are high tend to have higher mortality, however, the effects they report are small and not statistically significant for some specifications. Moreover, the variation they isolate is temporary as suggested by Matsudaira (2014) given that is caused by temporary shocks, and therefore may not identify the effect of a permanent increase in nurse staffing contrary to the reform I study in this paper.

This paper contributes to at least two strands of the literature. The first one studies the effects of staffing regulation and alternative work arrangements on hospital quality

[2] Is worth noting that a potential problem with this IV strategy is a possible correlation between policy-induced changes in nurse staffing and patient acuity (Tong, 2011).

and the second more broadly relates to the effects of labor market regulation on firm outcomes. Most of the literature on the effects of labor market regulation on firm outcomes focuses on employment protection legislation (EPL) reforms (Cappellari et al., 2012; Cingano et al., 2016; Griffith and Macartney, 2014). EPL affects firm outcomes through the increase in dismissal costs, which increases job security for existing workers. Efficiency wage arguments suggest that higher job security increases the value of employment for the worker and their (unobservable) effort, improving firm outcomes (Griffith and Macartney, 2014). This paper is related to the EPL literature because, like these regulations, the reform of Ecuador increased the labor stability of workers in the healthcare sector. More importantly, to the best of my knowledge, there are no other studies that evaluate the effects of a reform that drastically eliminates the possibility of hiring part-time staff in the future.

The labor reform that I study in this paper affected all public hospitals at different points in time according to an official implementation schedule. Hence, I use an event study to evaluate the effects of the reform on in-hospital mortality. The key identifying assumption in the event study is that the timing of adoption of the reform has to be random so that after controlling for time fixed effects and group fixed effects, there are no pre-existing differences in the outcome of interest between hospitals that were treated before and later. In other words, hospitals treated at different points in time follow parallel trends in the outcome of interest before the reform.

I build a dataset that merges information on hospital resources, hospital admissions, and budgets and to deal with the selection of patients to hospitals, I present results for the sample of admissions to the emergency department and contrast with the results obtained for the complete sample. The results for the sample of admissions to the emergency department show that mortality in public hospitals decreased by 0.1% on the adoption year and by 0.2% one year later. Results were robust to several alternative specifications and to the inclusion of pre-reform characteristics that could have been used by policymakers to decide the order of implementation.

The main contributions of this paper are: First, to the best of my knowledge, it is the first paper that studies the effects of a reform that replaced all part-time contracts with full-time contracts in the healthcare sector. Second, unlike the previous literature that has mostly focused on nurse staffing regulations, this paper studies a reform that affected all health professionals. Finally, it is one of the few papers looking at these topics in a developing country context. The rest of the paper is organized as follows: Section 3.2 explains the structure of the health system in Ecuador and describes the implementation of the labor reform at public sector hospitals. Section 3.3 describes the three sources of administrative data that are used to build the balanced panel of hospitals that will be used in all the analysis. Section 3.4 discusses the empirical strategy. Section 3.5 presents evidence on the validity of the identification strategy. Section 3.6 discusses the results of the event studies and some alternative specifications. Section 3.7 presents some additional robustness checks; and, Section 3.8 concludes.

3.2 Background

In this section, I describe the Ecuadorian health system and the process of implementation of the labor reform that started in 2009 and ended in 2014 with almost every part-time contract in the public sector being replaced by a full-time contract.

3.2.1 The health care system

Ecuador has a mixed health system made up of the public and private sectors. The public sector comprises the Ministry of Public Health (MSP), with its institutional providers which cater to individuals who lack health insurance and the National System of Social Security which caters formal workers. The National System of Social Security is made up of the Ecuadorian Institute of Social Security (IESS), which includes Rural Social Insurance, the Armed Forces Social Security Institute, and the National Police Social Security Institute (Giovanella et al., 2012). The private sector is made up of for-profit institutions (hospitals, clinics, community clinics, physicians' offices, pharmacies) and private non-profit organizations such as NGOs, popular medical service organizations, social service associations, etc. (Giovanella et al., 2012).

Although the largest number of facilities with hospital capacity belong to the private sector, the public sector concentrates the highest percentage of hospital admissions. In 2009, the Ministry of Public Health was in charge of 125 public hospitals, the Ecuadorian Institute of Social Security had 19 and the remaining 44 belonged to the other arms of the National System of Social Security. The private sector, on the other hand, had 534 facilities with hospital capacity. In terms of hospital discharges, the public sector covers nearly 62% of the total, while private hospitals cover only 38% of the total.[3]. Unlike the public sector, in the private sector, there is a great turnover of hospitals. Only during the period between 2014 and 2015, 19 health facilities entered the market; among them, 17 were private and 2 were public, while 90 left the market, 88 were private and 2 were public[4].

The public sector also hires the biggest share of healthcare professionals. Among those who registered their degrees between 2014 and 2018, 54.64% of the total worked in the public sector and 30.13% worked in the private sector, while the remaining 15.22% remains unclassified.[5] There are no statistics available about the qualifications held by health professionals in the public and private sector but in the aggregate, 25% hold a graduate degree, 68% hold an undergraduate degree and 7% hold a technical degree.[6] There is also very little information about the situation of

[3]Source: Hospital Discharges Yearbook INEC 2014.

[4]Source: Hospital Discharges Yearbook INEC 2014.

[5]Ministry of Health - ACESS statistics: http://www.calidadsalud.gob.ec/acess-app-servicio-ciudadano/public/estadistica/personalSalud.jsf

[6]Ministry of Health - ACESS statistics: http://www.calidadsalud.gob.ec/acess-app-servicio-

the health workforce in Ecuador. Most of the available information is published by the National Statistics Office and comes from hospital resources administrative data at the establishment level because of the confidential nature of micro-data.[7]

For many years, the professionals working at the Ministry of Health received the lowest remunerations of the public sector. As an example, the remuneration of a specialist physician was almost half of the market remuneration (Giovanella et al., 2012). As part of a wider reform which sought to standardize the salaries of all public officials, the government decided to move from a system that allowed hiring workers for 4 and 6 hours a day[8] to an 8-hour-a-day system in all the establishments and institutions of the public sector. In the next section, I explain in more detail the implementation of this labor reform in the public health system.

3.2.2 The labor reform

Employment contracts in the public sector can be indefinite or fixed-term. Indefinite contracts are called 'nombramientos' and can be provisional or definitive. Fixed term contracts have a maximum duration of twelve months renewable for one more year. By 2003, only 20% of the health care professionals working at public hospitals had an indefinite contract (nombramiento), while the other 80% had a fixed term contract (Merino, 2006). This situation remained almost unchanged until 2012.[9]

Regarding workloads, contracts can be full-time, part-time or occasional/on-call. Full-time contracts involve 8 hours of work per day, part-time contracts involve less than 8 hours of work per day and on-call contracts are those whose object is the attention of emergent or extraordinary needs, not linked to the habitual activity of

ciudadano/public/estadistica/personalSalud.jsf

[7] The Resolution 001-INEC-DIJU-NT-2015 establishes a lengthy process for researchers to access administrative microdata and when access is possible, the requirements for data processing are very demanding.

[8] These types of contracts were common in the healthcare and education sectors among others.

[9] Source:https://www.salud.gob.ec/ministerio-de-salud-entrega-500-nombramientos-a-profesionales-del-ramo/

the employer, hence their duration shall not exceed thirty days in a year. Before 2008, more than 50% of public hospitals had more than 60% of their payroll working less than 8 hours per day. This situation changed drastically with the enactment of a series of resolutions by the Ministry of Labor that banned the creation of new part-time contracts and replaced the existing part-time contracts with full-time contracts at all public institutions. The result was a sharp decline in the share of health professionals with part-time and occasional/on-call contracts and a marked increase in the share of full-time contracts.

Table 3.1 shows how the share of full-time, part-time and occasional/on-call contracts changed at public hospitals from the period between 2004 and 2007 to the period between 2012 and 2014. Hospitals are classified in quartiles of the share of part-time staff (SPTS) hired in 2004-2007.[10] Hospitals in the last quartile, those with the higher share of part-time staff in 2004-2007, experienced the biggest increase in the share of full-time staff in 2012-2014 (an increase of 77 percentage points), as well as the biggest reduction in the share of part-time staff, (a reduction of 75 percentage points).

The reform was implemented in a staggered way, following the resolutions issued by the Ministry of Labor for the effect. The first resolution (SENRES 2009-90) was enacted in May 2009 and affected only health care professionals working at facilities of the Ministry of Public Health. The resolution affected professionals with fixed-term contracts and applied at the moment of contract renewal. Adoption within this group of hospitals was progressive according to the implementation schedule that the Ministry of Labor and the Ministry of Health created for that purpose.[11] A second resolution (MRL-2011-33) was issued in February 2011 and included

[10]The share of part-time staff is the ratio of part-time workers to total staff hired in a year

[11]Several documents state that there was an official schedule that would determine the hospitals or groups of hospitals that would implement the reform first, one of them is the Agreement 916 of October 2011.

Table 3.1: Change in the share of full-time, part-time, and occasional/on-call contracts in public hospitals

	Mean 2004-2007	Mean 2012-2014	Difference	N
Share full-time doctors (%)				
$25 > SPTS \geq 0$	73	94	21	41
$50 > SPTS \geq 25$	50	91	41	50
$75 > SPTS \geq 50$	32	93	62	46
$100 > SPTS \geq 75$	12	89	77	47
Share part-time doctors (%)				
$25 > SPTS \geq 0$	13	3	-10	41
$50 > SPTS \geq 25$	38	7	-31	50
$75 > SPTS \geq 50$	63	6	-57	46
$100 > SPTS \geq 75$	86	11	-75	47
Share occasional/on-call doctors (%)				
$25 > SPTS \geq 0$	13	2	-11	41
$50 > SPTS \geq 25$	12	2	-10	50
$75 > SPTS \geq 50$	6	1	-5	46
$100 > SPTS \geq 75$	2	0	-2	47

Notes: The sample corresponds to an unbalanced panel of 184 public hospitals from the Hospital resources dataset compiled by the National Statistics Institute (INEC).

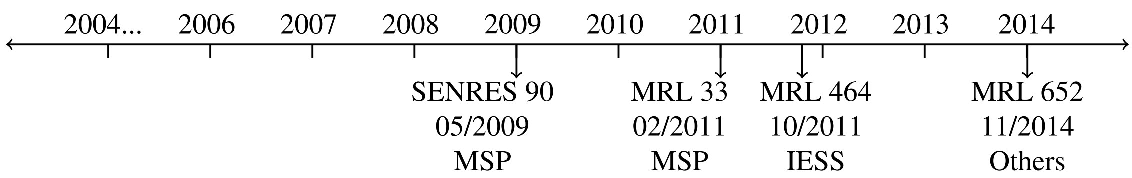
employees with indefinite appointments working at establishments of the Ministry of Health.

Given that by 2011, circa 80% of health professionals working at the Ministry of Health had a fixed-term contract[12], the hospitals that adopted the reform in 2009 and 2010 had already undergone the biggest changes by the time the second resolution was issued. While hospitals that had not adopted the reform yet continued to do so progressively according to the official schedule. A third resolution (MRL-2011-464), was issued in October 2011, with the intention of including all the facilities of the Ecuadorian Institute of Social Security. Finally, a fourth resolution (MRL-DM-2014-652) was published in November 2014 and included the remaining facilities of the Public Health System. Figure 3.1 shows the timetable of the reform.

The reform affected all health care professionals including physicians, dentists,

[12]https://www.salud.gob.ec/ministerio-de-salud-entrega-500-nombramientos-a-profesionales-del-ramo/

Figure 3.1: Timeline of the reform



obstetricians, psychologists, nurses, and medical technologists and excluded only the administrative positions because those professionals were already working 8 hours per day. All health care professionals were offered a proportional raise to compensate them for the hours that were added to their contracts.

The reform affected a group of highly educated professionals with sector-specific skills. Hence, mobility to other sectors of the economy is unlikely in particular for specialist physicians. However, horizontal movements to the private health care sector cannot be discarded as well as migration. Unfortunately, the analysis of spillover effects to the private sector is beyond the scope of this paper and there is no available information to quantify the extent of migration in the health sector.

3.3 Data and sample

In this section, I explain the three sources of administrative data that I use in the analysis, namely, hospital resources data, hospital discharges data (patient-level data) and hospitals' budget data. Then, I explain the process to build the balanced panel of hospitals that constitute the main sample in this study and discuss the outcomes of interest to this study.

3.3.1 Hospital resources data

I use administrative data on hospital resources (Recursos de Salud) compiled by the National Statistics Institute (INEC) from 2004 to 2014.[13] This dataset includes

[13] By the time of this study, individual-level data – that would allow me to identify the health professionals that were affected by the reform and their response in terms of working hours or job

all public and private hospitals in the country. Hospital resources data contain information on the number of health care professionals in any of the three types of contract: full-time, part-time and occasional/on-call.[14] In the dataset, full-time contracts are defined as those involving 8 hours of work per day, part-time contracts involve 4 to 6 hours of work per day and occasional/on-call require two hours or less of work per day.[15] Other available variables are medical equipment and number of consultations, exams, and tests reported in every hospital on a calendar year basis.

To identify the year when hospitals adopted the reform I requested the official implementation schedule to the Ministry of Health, however, the information I have only allows identifying the period when the different groups of public hospitals adopted the reform, not the specific dates when each hospital adopted it. Hence, to identify the adoption dates for each hospital, I use a data-driven approach based on three conditions. First, I calculated the growth rate of total part-time workers at hospital level with respect to the average for 2006-2007 and assumed that treatment started on the first year I observed a large reduction in the growth rate of part-time contracts. Since there are only 144 hospitals it is possible to test several thresholds (including -15%, -25%, -35%, -45%, -55%, and -65%) and determine which threshold better identifies the most likely adoption year based on visual inspection of the leaps in the series. In addition to this, I calculated the average reduction in part-time contracts for each group of public hospitals and found that during the adoption periods established by the Ministry of Labor, the different groups of hospitals experienced reductions in the number of part-time in the order

transitions – was not available. Recently, I got an authorization to access these type of data, which I plan to use in the next stage of this project.

[14]The specialties considered are general practitioners, surgeons, pediatricians, gynecologists, obstetricians, traumatologists, anesthesiologists, cardiologists, neurologists, psychiatrists, ophthalmologists, otolaryngologists, neonatology, intensivists, hematologists, pulmonologists, plastic surgeons, nephrologists, gastroenterologists, gerontologists, oncologists, urologists, residents, nurses, nutritionists, psychologists, health educators, sanitary engineers, environmental engineers, physiotherapy technician, anesthesiology, food technologists.

[15]The staffing hours reported in my data do not differentiate between patient-care hours and hours devoted to administrative work, teaching,etc.

of 35% or above, so I selected 35% as the threshold to identify the adoption year at the hospital level.

Furthermore, from the replication of the analyses using other thresholds (not shown) is possible to conclude that lower rates (below 25%) are problematic and cause the parallel trend tests in the event studies to fail, as expected, since with less stringent criteria for the treatment variable, I erroneously identify as the adoption year any year with a small change, and, comparisons are done between less similar hospitals. On the other hand, rates above 35% rendered similar results in the event study regressions to the ones obtained with the 35% threshold.

Since the reform is expected to increase the number of doctor-hours at public hospitals, the second condition to identify the adoption year is that there has to be an increase in the number of full-time equivalent doctors in the hospital. Full-time equivalent doctors per hospital are obtained by adding the number of doctors on each type of contract times the number of hours worked per day and dividing this by 8. Finally, the adoption year has to be within the adoption period determined by the resolutions of the Ministry of Labor for the different groups of hospitals. For hospitals of the Ministry of Health, the adoption date had to occur after the enactment of Resolution SENRES-2009-90. For the hospitals of the Ecuadorian Institute of Social Security, the adoption date had to occur at any point after the enactment of Resolution MRL-2011-464 and for the rest of health facilities, the adoption year was 2014 since this is the last year of the data and corresponds to the year when Resolution MRL-2014-652 was enacted. Consistent with this, out of the 144 public hospitals in my final sample, 42 hospitals in the panel adopted the reform in 2009, 17 in 2010, 23 in 2011, 36 in 2012, 3 in 2013 and 23 in 2014.[16]

[16] I plan to use the complete implementation schedule as soon as I get access to it, but I do not expect to see significant changes in the results, considering that the resolutions of the Ministry of Labor establish relatively short adoption periods and that for my identification strategy I do not require the exact date of adoption, just the year.

Drawing from the time-series literature on structural breaks, other studies have used data-driven methods to recover missing information, for instance, to determine the breakpoint for a regression-discontinuity design when the true breakpoint is not known (Chay et al., 2005; Carneiro et al., 2018). In my case, to identify structural breaks in the number of part-time contracts – that would signal the adoption of the reform –, I used a longer panel of hospital resources data covering the period 2000-2014. Unfortunately, with 15 observations per hospital, I was not able to identify the precise year of adoption within the period indicated in the resolutions of the Ministry of Labor.[17]

3.3.2 Hospital discharge data

Hospital-specific mortality rates are the most commonly used outcome-based measure of hospital quality (Gowrisankaran and Town, 1999). To study the effects of the reform on hospital mortality, I use hospital discharge data from 2004 to 2014. These data are compiled by the National Statistics Institute (INEC) and provide information on each patient discharged from public and private hospitals, including the patient's condition at the moment of discharge (alive or not) and other patient characteristics like age, sex, province and county of residence, precise date of hospitalization and discharge, reason of hospitalization using the corresponding ICD-10 code, and length of stay.

A limitation of hospital discharge data is that is not possible to construct several risk-adjusted mortality rates and 30 days mortality indicators since patients' survival is not observed after hospital discharge. However, it is still possible to study in-hospital mortality and this is the indicator that I will focus on throughout this paper. Using the ICD-10 code attached to every admission and the condition at the moment of discharge (alive or not), I obtain average mortality rates at hospital level for the

[17] I used the Stata commands 'sbsingle' and 'sbknown' to identify structural breaks.

complete sample of admissions and for admissions to the emergency department. I also compute hospital level weights defined as the ratio of hospital discharges to total discharges in the public sector in a year.

3.3.3 Budget data

Budget data is available for 101 public hospitals for the period between 2004 to 2014 and was accessed via a special petition to the Ministry of Finance. It contains very disaggregated information about all types of expenses and for some years have been classified into seven more general groups: (i) wages; (ii) medication; (iii) long-term goods; (iv) maintenance and repairs; (v) infrastructure; (vi) hospital and complementary medical services; and, (vii) other expenses. I use this information to control for hospital budget in the regressions. I distinguish between personnel expenditures and all other expenditures because the former are strongly correlated with the labor reform, while the latter are expected to be less affected, so this is the variable that I use as a control in the analyses.

3.3.4 Analytical sample

For the main sample in this study, I merged hospital discharge data and hospital resources data and constructed a balanced panel of hospitals covering the period between 2004 and 2014.[18] I used a balanced panel to reduce the likelihood of confounding effects due to changes in the supply of services at the local level, which is expected to affect mortality.

Before obtaining average mortality rates at hospital level, I restricted the sample to patients who were discharged within one year of admission to avoid having unusually large hospital stays in the sample; and, as is standard in the literature, I ignored patients aged 75 years and older and newborns and neonates because they

[18] I am currently restricted to this period of analysis until I get access to the hospital IDs for 2005, 2015 and 2016.

can have high mortality rates for reasons having no relation to the quality of care received that may bias the estimates of the impact of the reform (Evans and Kim, 2006).

The main sample contains 144 public hospitals with information for all the years, which represent 86.95% of all the discharges in the public sector in the period of analysis. The sample drops to 101 hospitals in the regressions that control for hospital budget. Table 3.2 provides a brief description of the public hospitals in the sample of 144 public hospitals and Table 3.9 in the appendix shows a description of the sample of 101 public hospitals.

Table 3.2: Characteristics of hospitals in the sample. Balanced panel of 144 public hospitals

	2009	2010	2011	2012	2013	2014
Admissions	12487	13005	13564	13191	13019	12225
Mortality	0.01	0.01	0.01	0.01	0.01	0.01
Mortality within 48 hours	0.00	0.00	0.00	0.00	0.00	0.00
Mortality after 48 hours	0.01	0.01	0.01	0.01	0.01	0.01
Mortality (ED)	0.02	0.03	0.03	0.02	0.02	0.03
Mortality within 48 h. (ED)	0.01	0.01	0.01	0.01	0.01	0.00
Mortality after 48 h. (ED)	0.02	0.02	0.02	0.01	0.02	0.03
Full-time staff	61.39	79.20	183.59	248.49	251.25	285.23
Part-time staff	160.70	157.57	71.75	32.37	18.96	28.97
Occasional staff	0.72	0.68	0.51	0.63	0.33	0.06
Full-time equiv. doctors	168.74	182.83	229.00	268.49	262.97	303.31
Patients - staff ratio	79.30	78.23	67.56	56.74	60.22	52.21
Length of stay (days)	4.41	4.40	4.36	4.31	4.41	4.59
Case mix index	1.42	1.42	1.43	1.42	1.43	1.43
Imaging equipment	7.09	7.83	8.50	8.41	9.45	9.61
Diagnostic equipment	8.27	8.90	9.38	10.40	10.79	14.15
Treatment equipment	14.53	13.11	16.88	13.52	16.21	17.18
Operating rooms	9.70	9.53	9.49	7.13	8.37	7.27

Notes: The sample corresponds to a balanced panel of 144 public hospitals.

3.3.5 Outcomes

I study the effects of the labor reform on in-hospital mortality, an outcome that is frequently used in the literature (Evans and Kim, 2006; Bell and Redelmeier, 2001;

Pronovost et al., 1999). In-hospital mortality is defined as the ratio of deaths to total discharges in a hospital. The problem with in-hospital mortality is that not all deaths are recorded during the hospital stay. However, part of the deaths that occur after discharge are unrelated to the care received, like when patients do not follow the recommendations for an adequate recovery.

In Ecuador, the leading causes of death are ischemic heart diseases, diabetes, cerebrovascular and hypertensive diseases, pneumonia and car accidents.[19] In this paper, I distinguish between mortality within the first 48 hours of hospital admission and after 48 hours because the first 48 hours are crucial in determining patients' survival, which has given rise to the term "golden hours", in particular with regards to myocardial infarction, stroke, birth-related bleeding, among other conditions. (Fisher and Gardner, 2012; Herlitz et al., 2010; Karoshi and Keith, 2009; Latense, 2009; Sircar et al., 2007)

Furthermore, several studies on the management of traumatic injuries have shown that most deaths occur in the first 24 to 48 hours (Jayawardena et al., 2007; Demetriades et al., 2004; Valdez et al., 2016). The trimodal distribution of traumatic death was first described by Trunkey in 1983, who demonstrated that most deaths occur in the first 24 hours. In fact, most patients with traumatic injuries die after the first 6 hours of admission (Demetriades et al., 2004). However, it has been documented that fewer patients die after arrival to a surgical center, which highlights the importance of the early presence of an experienced trauma surgeon, timely surgical intervention, and improvement in bleeding control (Valdez et al., 2016). All this suggest that a labor reform like the one I study in this paper, that increased the number of doctor hours and consequently raised the likelihood of having more specialists available at all times would have a direct effect on mortality within 48 hours. I also expect that these effects would be bigger than the ones stemming from

[19]Source: http://www.ecuadorencifras.gob.ec/vdatos/

nurse staffing reforms alone where the literature has reported null effects.

3.4 Empirical Strategy

3.4.1 Identification strategy

To credibly assess the impact of the labor reform on in-hospital mortality, ideally, hospitals should be randomly assigned to adopt the reform or not. Since this is not possible, I adopt a quasi-experimental approach that takes advantage of its staggered implementation. To claim causality, the reform itself should not respond to deficiencies in the public health system. I argue that the reform was exogenous to the health system because it was part of wider public sector reform to homogenize wages at all public institutions. Moreover, there was an official implementation schedule that established the timing of adoption for every hospital and as I will show in a moment, it was exogenous to hospital performance, meaning that the government did not assign hospitals with higher or lower mortality rates to be treated first.

Not only hospitals did not get to choose their adoption dates but also there is no evidence that the government decided the implementation schedule based on hospitals' observable characteristics. Proof of this are several conversations with former authorities of the Medical Association of Pichincha and Guayaquil who confirmed that there were no clear guidelines – based on observable characteristics of the hospitals – which would allow them to anticipate the approximate date on which they should initiate the process. That is why hospital directors should call the Ministry of Health to consult on the dates in which they should initiate the process.

A final piece of evidence to support the argument that the government did not decide the implementation schedule based on hospitals' observable characteristics is that the results of the econometric models – that I will discuss shortly – remained robust to the inclusion of pre-reform characteristics that could be considered by the authorities that designed the implementation schedule. All these elements lend themselves to the implementation of an event study.

3.4.2 Event study

In an event study all the groups in the panel are treated at different points in time. The key identifying assumption is that the timing of the treatment is random so that after controlling for time fixed effects and group fixed effects, there are no pre-existing differences in the outcome of interest between the groups that were treated before and the ones that were treated later, which is equivalent to a test for parallel trends in a difference-in-difference model. In this way, any difference that is observed after the adoption of a reform can be attributed to the policy. Following Dobkin et al. (2018) the basic specification of a non-parametric event study is the following:

$$Y_{jt} = \alpha_j + \lambda_t + \sum_{k=a}^{b} D_{jt}^{k}\theta_k + X_{jt}\beta + C_{jt}\gamma + \epsilon_{jt} \tag{3.1}$$

where Y_{jt} is the mortality rate at hospital j in year t, α_j are hospital fixed effects to control for any systematic differences across hospitals, λ_t are year fixed effects to control for shocks to mortality rates that are common to all hospitals in a year. The inclusion of hospital fixed effects is a natural approach to addressing potential bias related to attrition; with this, the impact of the labor reform is estimated entirely off of within-hospital changes and therefore should not be contaminated by any differential attrition correlated with the level of hospital mortality (Dobkin et al., 2018). However, if there is heterogeneity in treatment effects across hospitals, the event study coefficients could still be affected by compositional changes in the group of hospitals used to identify a given relative year coefficient θ_k, for that reason I use a balanced panel of hospitals (Dobkin et al., 2018).

Let $\tilde{t}$ be the adoption year, so $D_{jt}^{k} \equiv 1\{t = \tilde{t} + k\}$ are indicators for time relative to adoption and θ_k are coefficients on said indicators. In all the specifications, I set

$a = -3$ and $b = 3$.[20] The coefficients θ_k when $k > 0$ reflect the effect of the reform one, two and three years after the adoption year relative to the omitted period θ_{-1} which is the period immediately before adoption and is different across hospitals since it depends on the adoption year. ϵ_{jt} is the error term. Standard errors are clustered at county level.

X_{jt} is a vector of time-variant controls at the hospital level that include patient demographics at the hospital level like the share of patients in different age groups, the share of female patients and a proxy for hospital budget. C_{jt} is a vector of county-level characteristics and contains the number of public and private health facilities in the county. To account for zeros in the data, the log transformation log(variable +1) is used for the main outcomes and for the control variables. If the implementation schedule was determined taking into account pre-treatment hospital characteristics or county-level characteristics that are fixed in time, these would be controlled for by the hospital fixed effects dummies. But, if the implementation schedule was determined based on current hospital or county-level characteristics like the supply of health services at the local level or current hospitals' budget, one should control for those variables in the regressions.

Unfortunately, any labor reform will necessarily affect the current budget allocated to hospitals through changes in the wage bill. Yet, failure to control for budget in a regression that explains in-hospital mortality would introduce bias in the estimates. In order to control for this, I subtracted all personnel-related expenditures from hospitals' budgets and created a variable that contains only "other expenditures".

If we were suspicious that the assumption of independence between the timing of adoption and the outcome was violated (Dobkin et al., 2018) we would estimate a

[20] I chose the period of analysis to be within 3 years of the adoption year because hospital panel data is only available until 2014 and because analyzing pre-trends before 2006 means moving to far from the period of the Correa administration, to a period when health care policies were very different.

parametric event study. The baseline specification allows for a linear pre-trend in event time r. The basic specification is the following:

$$Y_{jt} = \alpha_j + \lambda_t + \delta r + \sum_{k=0}^{3} D_{jt}^{k}\theta_k + X_{jt}\beta + C_{jt}\gamma + \epsilon_{jt} \tag{3.2}$$

Since equation 3.2 includes a linear pre-trend in event time (k), the coefficients of interest θ_k, reflect the change in the outcome relative to any preexisting time trend δ (Dobkin et al., 2018). Note that apart from δr all other elements in the equation remain the same as in equation 3.1, except that I set $a = 0$ and $b = 3$. Researchers normally implement first the non-parametric event study to test for pre-trends and if this fails then they implement the parametric event study that controls for any existing pre-trend.

The two assumptions for the difference-in-difference estimator to provide accurate estimates of treatment effects also need to be met by the event study. First, time effects must be common across the treatment and control group. Second, the composition of the comparison groups must remain stable before and after the policy change (Blundell and Macurdy, 1999). In other words, hospitals that adopted the reform earlier than others should have been exposed to similar shocks than those that adopted it later. Also, hospital composition should follow similar trends in early and late adopters and should remain unaffected after the reform so that changes in mortality can be attributed to the reform and not to changes in case mix that could be caused by other reasons.

3.4.3 The selection problem

In Ecuador, public and private hospitals are required by law to treat all emergency patients no matter whether they have insurance or not. Only after the emergency has been overcome and the patient is stable can payment be requested for the care received. In case the patient does not have insurance, the State covers the cost of care.[21] Hence, to deal with the problem of non-random sorting of sicker patients to better hospitals, I focus on admissions to the emergency department, where patients have less discretion about the hospitals where they are treated.[22] Other studies have used the exogenous variation in ambulance assignment to hospitals (Doyle et al., 2015) or have focused on admissions for "nondeferrable" conditions to circumvent the selection problem (Doyle et al., 2015; Card et al., 2009).

3.5 Assessing the validity of the identification strategy

I argue that the reform was exogenous to hospitals' performance because hospitals did not decide when to adopt the reform. Instead, they followed an official schedule. To test that the reform was exogenous to hospitals' performance, I implement the event study described in equation 3.1 and check for pre-trends in mortality (Autor, 2003; Dobkin et al., 2018; Frakes, 2013). The test for the exogeneity of the reform is equivalent to a test for pre-trends in the outcome, meaning that each θ_k for $k < 0$ should not be statistically different from zero.

In this test, after controlling for hospital and year fixed effects, the coefficient θ_{-3}, for example, reflects the difference in the mortality reported in t-3 by hospitals that were treated in year t (for which $D_{jt}^{-3} = 1$), versus the mortality reported by those same hospitals over time (from t-2 until t+3, for which $D_{jt}^{-3} = 0$) and also relative to the mortality reported by hospitals treated in other years (for which $D_{jt}^{-3} = 0$). This excludes other hospitals treated in the same year t because for them $D_{jt}^{-3} = 1$. If the timing of adoption is truly random, mortality should not be significantly different across time for the same hospital and neither across hospitals treated at different years, once hospital and year fixed effects are controlled for in the model. If this is

[21] See Art.365 of the Constitution and Art.7 of the Ecuadorian Health Law.

[22] Since patient discharge data do not include a variable to distinguish between emergency admissions and other admissions, I have requested a list of the most common Emergency Department conditions to the Ministry of Public Health. For the moment, I use the Emergency Department ICD-10-AM Principal Diagnosis Short List published by the Independent Hospital Pricing Authority of Australia, since is the most comprehensive list that I have found so far.

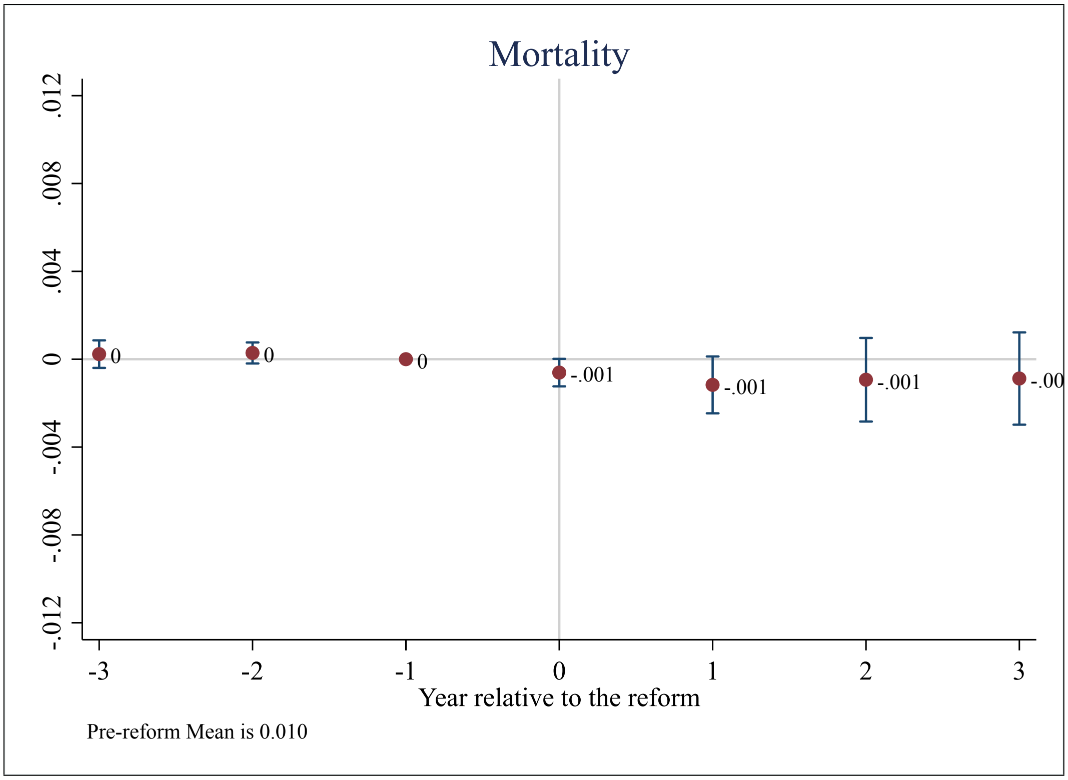
true, then any change in mortality observed after the adoption of the reform can be interpreted as the causal effect of the reform.

Figure 3.2 shows the evolution of the coefficients of the indicator variables relative to the adoption year (θ_k) when the outcome is mortality rate among all admissions. It includes time and hospital fixed effects, additional controls, including the share of patients in different age groups, the share of female patients and the number of public and private health facilities in the county. All analyses include hospital level weights defined as a share of total discharges.

Figure 3.2 shows that during the three years prior to adoption, hospitals that implemented the reform in year t had a similar mortality rate to hospitals that adopted it at other years. The difference only becomes statistically significant on the implementation year and later, showing a reduction in mortality that will be discussed in the next section, where I present in more detail these results and those of several other specifications. Figure 3.3 also shows that there is no evidence of pre-existing trends in mortality when looking at the sample of admissions to the emergency department.

The absence of pre-trends is taken as evidence in favor of the strict exogeneity of the policy change but is not sufficient to remove all concerns about policy endogeneity. It may be that pre-trends were not detected due to limited statistical power (Freyaldenhoven et al., 2018). The inclusion of hospital fixed effects to control for permanent differences across hospitals in the timming of adoption and outcomes would solve the endogeneity of the policy only if the systematic determinants of adoption are additive, time-invariant hospital characteristics; however, if the source of endogeneity are time variant characteristic, the problem remains (Besley and Case, 2000). One solution is to find an instrument for policy adoption across hospitals. In cross-state fixed effect estimation and difference-in-difference estimation, possible instruments are variables that account for changes in political represen-

Figure 3.2: θ_k estimates from the event study when the outcome is mortality

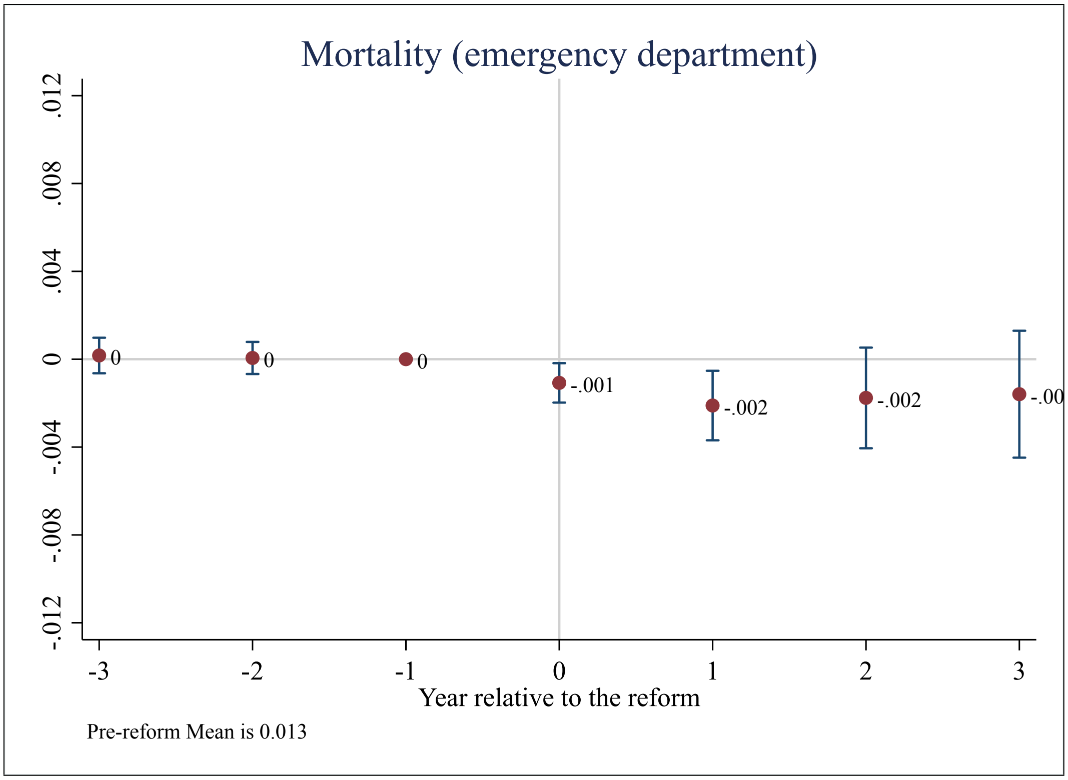


Notes: The sample is the balanced panel of 144 public hospitals. The figure plots the estimates from my preferred specification of the event study in equation 3.1. The range plot depicts the 95 percent confidence intervals. All estimates are weighted by the share of hospital admissions among all admissions in a year.

tation that determine policy adoption at the state level. However, in my setting, the reform was decided in a centralized way and all hospitals had to abide by it. Another approach for handling possible policy endogeneity is to explicitly control for the variables that may have been considered by the policymakers when choosing which hospitals would be treated first (Duflo, 2001). One example is total staff hired in 2007 since is possible that hospitals in need of more staff may choose to adopt the reform earlier. Since these variables are fixed in time, they must be interacted with a time-varying variable (the year in this case), otherwise, they would be absorbed by the hospital fixed effects.

In the next section, I present the results of the event study controlling for pre-reform

Figure 3.3: θ_k estimates from the event study when the outcome is mortality



Notes: The sample is the balanced panel of 144 public hospitals. The figure plots the estimates from my preferred specification of the event study in equation 3.1. The range plot depicts the 95 percent confidence intervals. All estimates are weighted by the share of hospital admissions among all admissions in a year.

characteristics and contrast them with the results of the basic specification in 3.1. It is reassuring that the estimates of both specifications are very similar. However, due to lack of statistical power, most of the coefficients of the specification that controls for pre-reform characteristics are not statistically significant. To tackle this problem, I estimate a slightly different model, where I impose the restriction that there are no pre-trends in the outcomes of interest so that instead of including the complete set of leads and lags relative to the adoption date, I assume that the coefficients of all the leads are zero. Before discussing these results, I present the results of the non-parametric event study from equation 3.1 and show that in fact there are no pre-trends in any of the outcomes of interest. I estimate the following specification:

$$Y_{jt} = \alpha_j + \lambda_t + \sum_{k=0}^{3} D_{jt}^k \theta_k + C_{j,2007} * \delta_{jt} + X_{jt}\beta + C_{jt}\gamma + \epsilon_{jt} \tag{3.3}$$

where Y_{jt} is the mortality rate at hospital j in year t, α_j are hospital fixed effects, λ_t are year fixed effects. $D_{jt}^k \equiv 1\{t = \tilde{t} + k\}$ are indicators for time relative to adoption (lags), δ_{jt} are dummy variables that indicate the year and $C_{j,2007}$ is the logarithm of total staff hired in 2007.

This increases efficiency and leads to more precise estimates of the effects of the program since now the reference category is not just the period immediately before adoption but all the periods before adoption. The increase in precision of the estimates is evidenced by an increase in the number of statistically significant coefficients.

3.6 Results

In this section, I present the results of the non-parametric event study described in equation 3.1, and as robustness checks, I also present the results of the specification that controls for pre-treatment characteristics (equation 3.3) and of the parametric event study specified in equation 3.2. The main outcome of interest is overall in-hospital mortality; however, I also investigate the effects on mortality within the first 48 hours of admission and after 48 hours of admission. To deal with the problem of non-random sorting of sicker patients to hospitals, I present results for the full sample of admissions and for the sample of admissions to the emergency department. All analyses include hospital level weights defined as the share of total discharges in a year.

3.6.1 Overall mortality

Columns 1 to 3 in Table 3.3 report the results of different specifications of equation 3.1. The specification in column 1 does not include hospital fixed effects while

Table 3.3: Estimates from the event study regressions. Mortality among all admissions and among admissions to the emergency department

Specification	(1) Panel	(2) Panel FE	(3) Controls for budget	(4) Pre reform controls	(5) Parametric ES
Panel A: Mortality among all admissions					
Efect in t	0.000122	-0.000610*	-0.000245	-0.000912*	-0.000550**
	(0.849)	(0.056)	(0.438)	(0.065)	(0.050)
Efect in t+1	0.000952	-0.00117*	-0.000994	-0.00155	-0.000994
	(0.307)	(0.076)	(0.105)	(0.204)	(0.300)
Efect in t+2	0.00240*	-0.000935	-0.000854	-0.00172	-0.000626
	(0.088)	(0.332)	(0.254)	(0.337)	(0.656)
Efect in t+3	0.00228	-0.000879	-0.00132	-0.00217	-0.000472
	(0.145)	(0.408)	(0.183)	(0.330)	(0.769)
Pre-reform means	0.0100	0.00900	0.00900	0.00900	0.00900
P-value (F-test of leads)	0.00166	0.453	0.513		
N	900	900	622	900	900
R2	0.672	0.932	0.950	0.937	0.932
Panel B: Mortality among admissions to the emergency department					
Efect in t	-0.000425	-0.00108**	-0.000644	-0.00123	-0.000984*
	(0.584)	(0.019)	(0.298)	(0.125)	(0.097)
Efect in t+1	-0.000133	-0.00211***	-0.00179**	-0.00241*	-0.00193*
	(0.913)	(0.009)	(0.012)	(0.098)	(0.095)
Efect in t+2	0.00138	-0.00176	-0.00198	-0.00280	-0.00150
	(0.444)	(0.130)	(0.129)	(0.200)	(0.381)
Efect in t+3	0.00126	-0.00159	-0.00234	-0.00310	-0.00125
	(0.560)	(0.276)	(0.144)	(0.267)	(0.555)
Pre-reform means	0.0140	0.0130	0.0130	0.0130	0.0130
P-value (F-test of leads)	0.0726	0.908	0.663		
Panel	Yes	Yes	Yes	Yes	Yes
Hospital FE	No	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Hospital Linear trends					Yes
N	900	900	622	900	900
R2	0.643	0.924	0.916	0.930	0.924

Notes: Robust standard errors clustered at county level shown in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Column 1 reports on the results of equation 3.1 without including hospital fixed effects while the specification in column 2 includes hospital fixed effects. The specification in column 3 is estimated on the sample of 101 public hospitals for which budget data is available and controls for a proxy for the annual budget. Column 4 reports on the results of the event study in equation 3.3 and controls for total staff hired in 2007 interacted with the year. Column 5 shows the results of the parametric event study in equation 3.2.

the specification in column 2 includes hospital fixed effects. Both specifications are estimated on the sample of 144 public hospitals in the balanced panel. The specification in column 3 on the other hand, is estimated on the sample of 101 public hospitals for which budget data is available and controls for a proxy of the yearly budget allocated to public hospitals. Column 4 reports on the results of the event study in equation 3.3 and controls for total staff hired in 2007 interacted with the year to deal with possible endogeneity in adoption. Finally, column 5 shows the results of the parametric event study in equation 3.2 that controls for a linear trend in event time.

In all the specifications, I report on the effects of the reform up to three years after adoption and present separate results for the complete sample of admissions and for the sample of admissions to the emergency department. Panel A in Table 3.3 shows that the estimated effects of the reform vary very little across specifications. My preferred specification is in column 2, it includes hospital and time fixed effects and a set of controls including the share of patients in different age groups, the share of female patients and the number of public and private health facilities in the county. The specifications in columns 3 to 5 serve as robustness checks for the results in my preferred specification.

Results in column 2, panel A, show that for the complete sample of admissions, the reform reduced mortality by 0.06% on the adoption year relative to a pre-reform mean of 0.9%, meaning that mortality decreased in 0.7 per 1000 individuals on the adoption year relative to the year immediately prior adoption ($k = -1$). The effect was statistically significant at 10% level. Similarly, one year after adoption, there was a statistically significant reduction in mortality of 0.1%, equivalent to 1 per 1000 individuals. The results reported in column 2 are very similar to those in columns 3 to 5. Results in column 3 are the least similar though, possibly because they are estimated on a selected sample since hospitals for which budget data are available seem to be slightly different to the rest of hospitals as shown in Table 3.9 in the Appendix.[23]

Panel B shows the results for the sample of admissions to the emergency department. In general, results are bigger in magnitude and remain statistically significant even after including additional controls like a proxy for hospital budget and pre-reform characteristics. Column 2 shows that on the adoption year, mortality decreased by 0.1% and one year later it decreased by 0.2% relative to a mean mortality rate of 1.3% (in $k = -1$), which means that one year after adoption, mortality among admissions to the emergency department dropped in 0.8 and 1.5 per 1000 individuals respectively. These results were significant at 5% and 1% level, respectively.

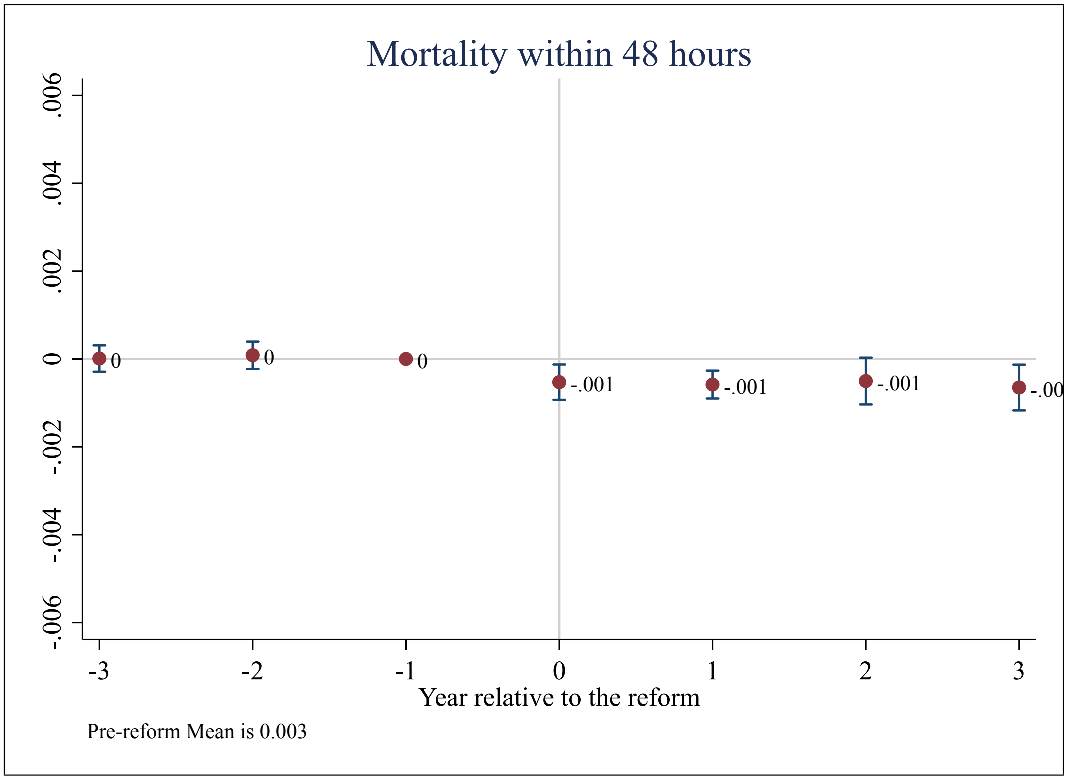
3.6.2 Mortality within 48 hours of admission

Figures 3.4 and 3.5 show that three years prior to adoption, there is no evidence of pre-trends in mortality within 48 hours. This is true for the complete sample of admissions and for the sample of admissions to the emergency department. Given that there are no pre-trends before adoption, my preferred specification is once again the non-parametric event study described in equation 3.1 and reported in column 2 of Table 3.4.

Mortality within 48 hours of admission decreased progressively after adoption. This was true for the complete sample and for the subsample of admissions to the emergency department. However, the effects on the sample of admissions to the emergency department were larger than for the complete sample. Column 2 in panel A shows that for the complete sample, mortality within 48 hours decreased by around 0.06% in the adoption year and the years that followed, relative to a pre-reform rate of 0.3% which is equivalent to a reduction of 2 persons per 1000 individuals in each case.

[23] I have requested information on the budgets of more public hospitals and I plan to use these data in the future.

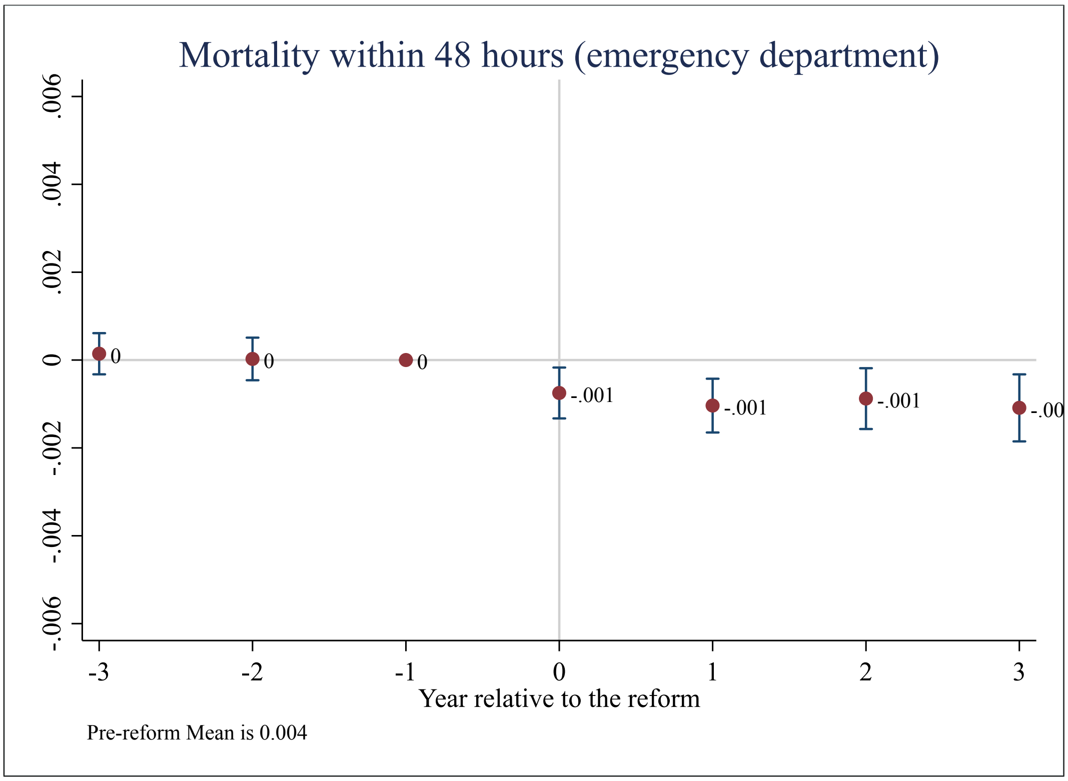
Figure 3.4: θ_k estimates from the event study when the outcome is mortality within 48 hours



Notes: The sample is the balanced panel of 144 public hospitals. The points in each figure represent the estimated effects from my preferred specification of the event study in equation 3.1. The range plot with capped spikes present the 95 percent confidence intervals. All estimates are weighted by the share of hospital discharges among all discharges in a year.

For the sample of admissions to the emergency department, Panel B shows that mortality decreased between 0.08% and 0.1% after the adoption of the labor reform relative to a pre-reform rate of 0.4%. This effect is equivalent to a mortality reduction of 2.3 per 1000 individuals.

Figure 3.5: θ_k estimates from the event study when the outcome is mortality within 48 hours



Notes: The sample is the balanced panel of 144 public hospitals. The figure plots the estimates from my preferred specification of the event study in equation 3.1. The range plot depicts the 95 percent confidence intervals. All estimates are weighted by the share of hospital admissions among all admissions in a year.

Table 3.4: Estimates from the event study regressions. Mortality within 48 hours, among all admissions and among admissions to the emergency department

Specification	(1) Panel	(2) Panel FE	(3) Controls for budget	(4) Pre reform controls	(5) Parametric ES
Panel A: Mortality within 48 hours among all admissions					
Efect in t	-0.000568**	-0.000528**	-0.000228	-0.000607***	-0.000549***
	(0.025)	(0.011)	(0.381)	(0.001)	(0.006)
Efect in t+1	-0.000494**	-0.000582***	-0.000400	-0.000711***	-0.000598**
	(0.038)	(0.000)	(0.177)	(0.000)	(0.014)
Efect in t+2	-0.000231	-0.000502*	-0.000485	-0.000681*	-0.000504
	(0.637)	(0.064)	(0.192)	(0.060)	(0.260)
Efect in t+3	-0.000593	-0.000650**	-0.000730**	-0.000916**	-0.000655
	(0.299)	(0.015)	(0.034)	(0.036)	(0.182)
Pre-reform means	0.00300	0.00300	0.00300	0.00300	0.00300
P-value (F-test of leads)	0.790	0.774	0.332		
N	900	900	622	900	900
R2	0.568	0.826	0.840	0.832	0.826
Panel B: Mortality within 48 hours among admissions to the emergency department					
Efect in t	-0.000741**	-0.000750**	-0.000476	-0.000688**	-0.000663**
	(0.014)	(0.012)	(0.240)	(0.037)	(0.043)
Efect in t+1	-0.000839**	-0.00104***	-0.000860	-0.000906**	-0.000878**
	(0.014)	(0.001)	(0.111)	(0.039)	(0.032)
Efect in t+2	-0.000432	-0.000878**	-0.00103	-0.000777	-0.000652
	(0.474)	(0.013)	(0.123)	(0.183)	(0.296)
Efect in t+3	-0.000808	-0.00109***	-0.00145**	-0.00101	-0.000786
	(0.228)	(0.006)	(0.012)	(0.147)	(0.260)
Pre-reform means	0.00400	0.00400	0.00400	0.00400	0.00400
P-value (F-test of leads)	0.931	0.800	0.925		
Panel	Yes	Yes	Yes	Yes	Yes
Hospital FE	No	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Hospital Linear trends					Yes
N	900	900	622	900	900
R2	0.539	0.792	0.758	0.798	0.792

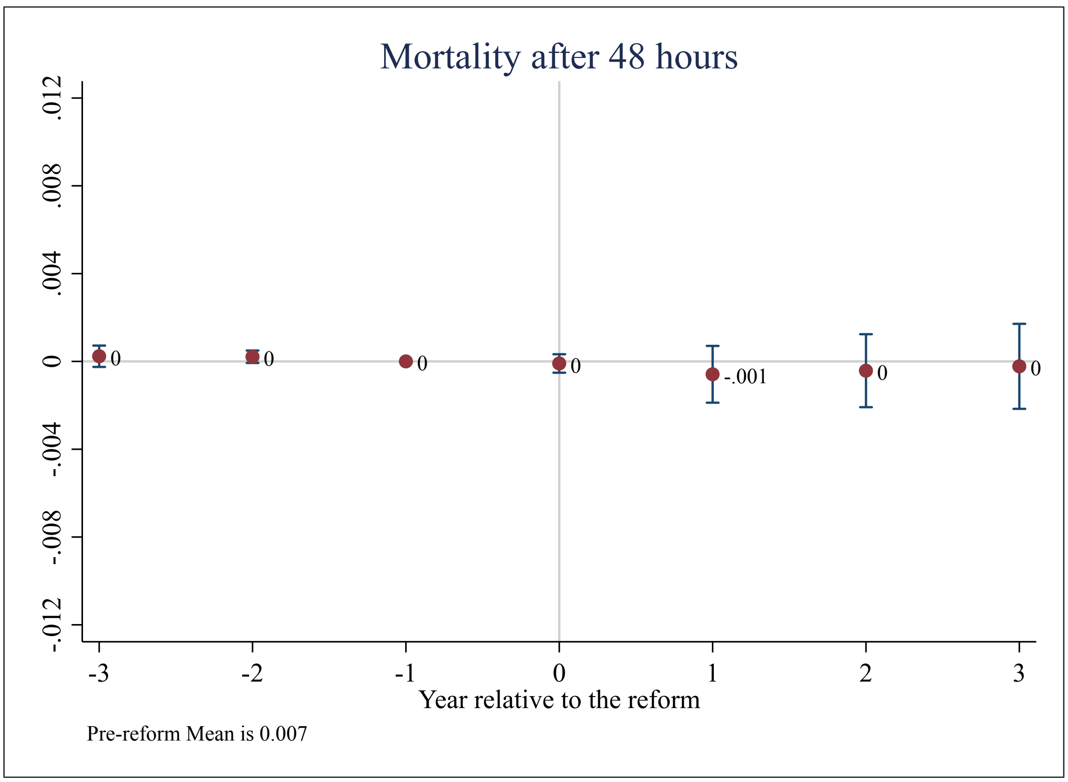
Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column 1 reports on the results of equation 3.1 without including hospital fixed effects while the specification in column 2 includes hospital fixed effects. The specification in column 3 is estimated on the sample of 101 public hospitals for which budget data is available and controls for a proxy for the annual budget. Column 4 reports on the results of the event study in equation 3.3 and controls for total staff hired in 2007 interacted with the year. Column 5 shows the results of the parametric event study in equation 3.2.

3.6.3 Mortality after 48 hours of admission

Figures 3.6 and 3.7 add further evidence about the exogeneity of the reform since they show that there are no pre-trends in mortality rates after 48 hours of admission for the complete sample and for the sample of admissions to the emergency department.

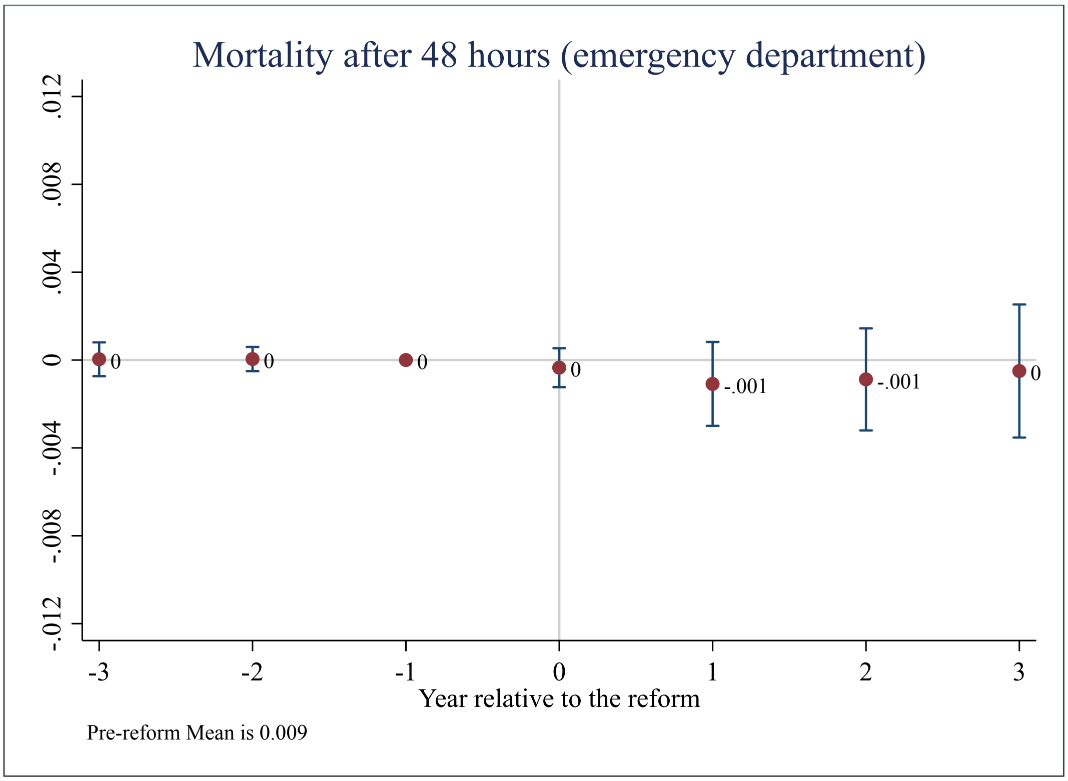
Column 4 in Table 3.5 shows that for the complete sample and for the subsample of admissions to the emergency department, the reform had a negative but not statistically significant effect on mortality after 48 hours.

Figure 3.6: θ_k estimates from the event study when the outcome is mortality after 48 hours



Notes: The sample is the balanced panel of 144 public hospitals. The figure plots the estimates from my preferred specification of the event study in equation 3.1. The range plot depicts the 95 percent confidence intervals. All estimates are weighted by the share of hospital admissions among all admissions in a year.

Figure 3.7: θ_k estimates from the event study when the outcome is mortality after 48 hours



Notes: The sample is the balanced panel of 144 public hospitals. The figure plots the estimates from my preferred specification of the event study in equation 3.1. The range plot depicts the 95 percent confidence intervals. All estimates are weighted by the share of hospital admissions among all admissions in a year.

Table 3.5: Estimates from the event study regressions. Mortality after 48 hours among all admissions and among admissions to the emergency department

Specification	(1) Panel	(2) Panel FE	(3) Controls for budget	(4) Pre reform controls	(5) Parametric ES
Panel A: Mortality after 48 hours among all admissions					
Efect in t	0.000677	-9.35e-05	-1.30e-05	-0.000311	-6.97e-06
	(0.116)	(0.660)	(0.956)	(0.547)	(0.979)
Efect in t+1	0.00145*	-0.000587	-0.000593	-0.000827	-0.000383
	(0.054)	(0.370)	(0.336)	(0.484)	(0.673)
Efect in t+2	0.00266***	-0.000423	-0.000363	-0.00101	-9.05e-05
	(0.006)	(0.615)	(0.533)	(0.531)	(0.940)
Efect in t+3	0.00287***	-0.000223	-0.000591	-0.00123	0.000217
	(0.007)	(0.819)	(0.522)	(0.554)	(0.880)
Pre-reform means	0.00800	0.00700	0.00700	0.00700	0.00700
P-value (F-test of leads)	0.00121	0.261	0.529		
N	900	900	622	900	900
R2	0.685	0.944	0.957	0.949	0.944
Panel B: Mortality after 48 hours among admissions to the emergency department					
Efect in t	0.000292	-0.000348	-0.000160	-0.000555	-0.000338
	(0.612)	(0.438)	(0.737)	(0.527)	(0.593)
Efect in t+1	0.000698	-0.00109	-0.000943	-0.00151	-0.00106
	(0.484)	(0.261)	(0.197)	(0.356)	(0.419)
Efect in t+2	0.00183	-0.000881	-0.000955	-0.00201	-0.000832
	(0.158)	(0.454)	(0.336)	(0.361)	(0.625)
Efect in t+3	0.00206	-0.000499	-0.000904	-0.00207	-0.000435
	(0.197)	(0.744)	(0.533)	(0.478)	(0.843)
Pre-reform means	0.00900	0.00800	0.00800	0.00800	0.00800
P-value (F-test of leads)	0.0721	0.986	0.686		
Panel	Yes	Yes	Yes	Yes	Yes
Hospital FE	No	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Hospital Linear trends					Yes
N	900	900	622	900	900
R2	0.656	0.934	0.926	0.941	0.934

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column 1 reports on the results of equation 3.1 without including hospital fixed effects while the specification in column 2 includes hospital fixed effects. The specification in column 3 is estimated on the sample of 101 public hospitals for which budget data is available and controls for a proxy for the annual budget. Column 4 reports on the results of the event study in equation 3.3 and controls for total staff hired in 2007 interacted with the year. Column 5 shows the results of the parametric event study in equation 3.2.

3.7 Robustness checks

To make sure that the observed effects were caused by the labor reform, hospital composition should follow similar trends in early and late adopters, and should not be affected by the reform so that the reduction in mortality can be attributed to the increase in the available doctor-hours and not to changes in the patient mix, for instance, which can be explained by other reasons. Neither should I observe changes in other indicators of hospital quality that would explain a reduction in mortality after the reform. To test this, I estimate the event study in equation 3.1 and show evidence of the stability in hospitals' patient mix, and length of stay.

3.7.1 Case mix index

In all the analyses, I use a balanced panel of hospitals so that the event study coefficients are not affected by compositional changes in the group of hospitals used to identify a given relative year coefficient (Dobkin et al., 2018). However, a concern that remains is that hospital patient mix itself could have changed across time. To test this, I generate two alternative measures of the patient mix at the hospital level. The first one is the case mix index (CMI), which is calculated by summing the diagnosis-related group (DRG) weights for all the discharges and dividing by the number of discharges. Ecuador does not use DRG groups, and there is no official definition for a CMI for Ecuador, so to construct the CMI, I matched the ICD-10 code reported as the main reason for admission into hospital with the corresponding DRG code used by The Centers for Medicare & Medicaid Services (CMS), which also provide the weights associated to each DRG.[24]

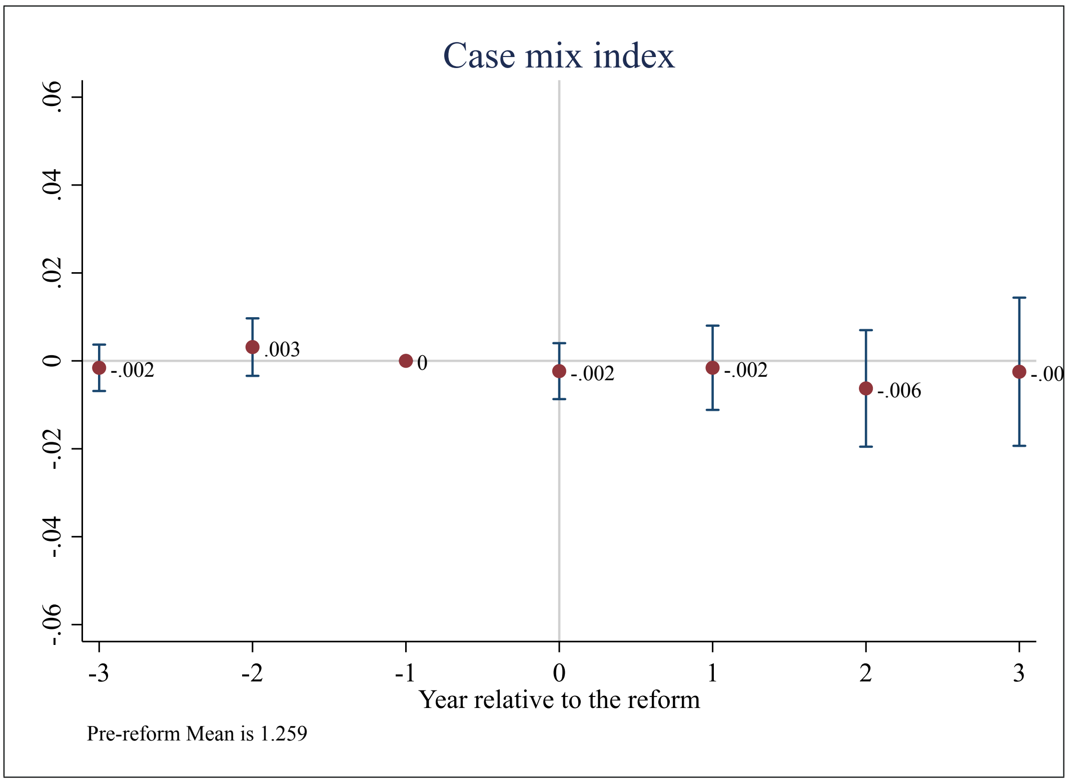
For the second measure of patient mix, I obtained the weight that each condition represented within the total admissions to a given hospital in 2007. I assigned these

[24] Since health care costs and reimbursement systems are different in the US, I just use the fact that the DRG weights are fixed in time and provide an easier way to distinguish more complicated cases from simpler ones.

fixed weights to each admission in the remaining years using the ICD-10 codes. Finally, I added all the weights corresponding to the different conditions observed in a hospital in each year to get an alternative measure of the patient mix at the hospital and year level.

Figure 3.8 shows graphical evidence on the parallel trends assumption for the logarithm of the CMI.

Figure 3.8: θ_k estimates from the event study when the outcome is the case mix index



Notes: The sample is the balanced panel of 144 public hospitals. The figure plots the estimates from my preferred specification of the event study in equation 3.1. The range plot depicts the 95 percent confidence intervals. All estimates are weighted by the share of hospital admissions among all admissions in a year.

Using equation 3.1 I test the parallel trends assumption by evaluating the coefficients θ_k for $k < 0$, and checking that are not statistically different from zero. Figure 3.8 shows that three years before adoption hospitals treated at different points in time

Table 3.6: Estimates from the event study regressions. Case Mix Index

Specification	(1) Panel	(2) Panel FE	(3) Controls for budget	(4) Pre reform controls	(5) Parametric ES
Efect in t	0.00481	-0.00234	-0.000433	-0.00545	-0.00441
	(0.385)	(0.468)	(0.937)	(0.175)	(0.246)
Efect in t+1	0.0101*	-0.00157	0.00660	-0.00523	-0.00443
	(0.095)	(0.746)	(0.359)	(0.498)	(0.495)
Efect in t+2	0.0180*	-0.00627	0.00482	-0.0115	-0.00947
	(0.074)	(0.350)	(0.641)	(0.250)	(0.285)
Efect in t+3	0.0293*	-0.00248	0.00769	-0.0110	-0.00686
	(0.062)	(0.771)	(0.476)	(0.324)	(0.506)
Pre-reform means	1.270	1.250	1.250	1.250	1.250
P-value (F-test of leads)	0.209	0.470	0.0494		
Panel	Yes	Yes	Yes	Yes	Yes
Hospital FE	No	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Hospital Linear trends					Yes
N	900	900	622	900	900
R2	0.864	0.945	0.940	0.946	0.945

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column 1 reports on the results of equation 3.1 without including hospital fixed effects while the specification in column 2 includes hospital fixed effects. The specification in column 3 is estimated on the sample of 101 public hospitals for which budget data is available and controls for a proxy for the annual budget. Column 4 reports on the results of the event study in equation 3.3 and controls for total staff hired in 2007 interacted with the year. Column 5 shows the results of the parametric event study in equation 3.2.

are not significantly different in terms of their CMI. The observed parallel trend before adoption in the CMI provides further evidence that endogenous sorting of more severe cases to certain hospitals is not correlated to the timing of adoption meaning that any sorting of patients to hospitals acts in the same way on average for the group of hospitals that adopted the reform earlier and those that adopted it later.

Table 3.6 shows the results for different specifications and samples. Like the results reported in section 3.6, my preferred specification is in column 2 and shows that the reform did not have a significant effect on CMI in none of the years that followed the adoption. All in all, the results of the event study on CMI lend credibility to the results found in section 3.6. Hospitals that adopted the reform at different points in time followed similar paths in terms of the CMI three years before and after the adoption of the reform. Table 3.7 shows similar results when the outcome of interest is the logarithm of the proposed measure of patient mix.

Table 3.7: Estimates from the event study regressions. Patient Mix

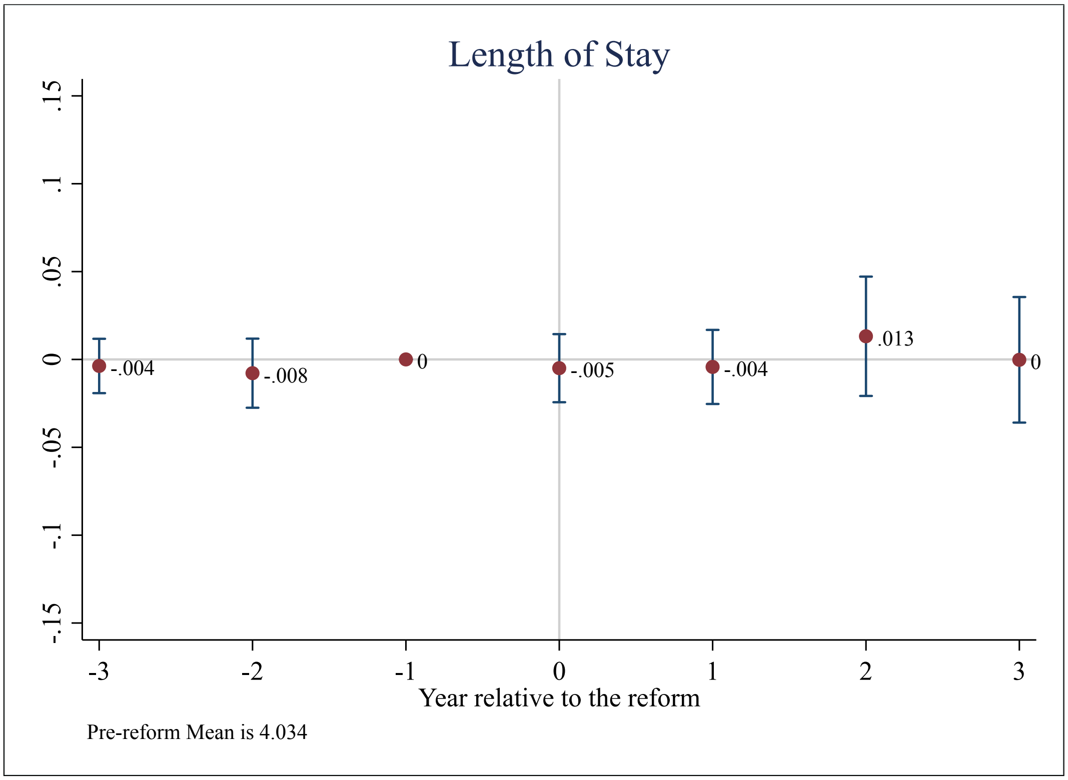
	(1)	(2)	(3)	(4)	(5)
Specification	Panel	Panel FE	Controls for budget	Pre reform controls	Parametric ES
Efect in t	0.0366	-0.0100	-0.0244	-0.0401	-0.0271
	(0.349)	(0.657)	(0.418)	(0.206)	(0.356)
Efect in t+1	0.174***	0.0576	-0.00782	0.0104	0.0368
	(0.000)	(0.163)	(0.857)	(0.836)	(0.464)
Efect in t+2	0.245***	0.0692	-0.0176	0.0403	0.0491
	(0.001)	(0.249)	(0.758)	(0.625)	(0.564)
Efect in t+3	0.282**	0.0811	0.00133	0.0340	0.0531
	(0.014)	(0.430)	(0.991)	(0.800)	(0.699)
Pre-reform means	414	460	460	460	460
P-value (F-test of leads)	0.00518	0.112	0.602		
Panel	Yes	Yes	Yes	Yes	Yes
Hospital FE	No	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Hospital Linear trends					Yes
N	900	900	622	900	900
R2	0.856	0.972	0.975	0.973	0.972

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column 1 reports on the results of equation 3.1 without including hospital fixed effects while the specification in column 2 includes hospital fixed effects. The specification in column 3 is estimated on the sample of 101 public hospitals for which budget data is available and controls for a proxy for the annual budget. Column 4 reports on the results of the event study in equation 3.3 and controls for total staff hired in 2007 interacted with the year. Column 5 shows the results of the parametric event study in equation 3.2.

3.7.2 Length of stay

To attribute the observed reduction in mortality to the labor reform, an additional check looks at changes in other indicators of hospital quality that would explain the observed effects. While some quality indicators like patient satisfaction are directly linked to the care provided by hospital staff and the number of doctors and nurses available at different times, indicators such as length of stay will be less affected by staffing changes.

Figure 3.9: θ_k estimates from the event study when the outcome is the length of stay



Notes: The sample is the balanced panel of 144 public hospitals. The figure plots the estimates from my preferred specification of the event study in equation 3.1. The range plot depicts the 95 percent confidence intervals. All estimates are weighted by the share of hospital admissions among all admissions in a year.

Length of stay depends on the availability of beds in a hospital, but it may also be affected by staff availability if, for example, when there is not enough staff, doctors decide to discharge some patients before they have fully recovered. Other reasons for sudden changes in length of stay may respond to budgetary shocks, sudden surges in demand or changes in the severity of patients (caused by an epidemic or natural disaster). Some studies suggest that sufficient nurse-to-patient ratios are associated with lower hospital mortality and reduced length of stay (Pham et al., 2011). It has also been found in the literature that hospitals utilizing more licensed nurses and fewer unlicensed aides have a shorter residual length of stay because the former are more experienced and have acquired more human capital (Bartel et al., 2014).

To check for pre-trends in the average length of stay at hospitals treated at different points in time, I estimate the event study regression in equation 3.1. The outcome of interest is the logarithm of the length of stay. Figure 3.9 and the results presented in column 2 of Table 3.8 show that length of stay remained comparable across hospitals treated at different points in time during the three years before and after adoption.

Table 3.8: Estimates from the event study regressions. Length of stay

	(1)	(2)	(3)	(4)	(5)
Specification	Panel	Panel FE	Controls for budget	Pre reform controls	Parametric ES
Efect in t	-0.0256	-0.00497	0.00962	-0.00928	-0.00487
	(0.453)	(0.612)	(0.402)	(0.471)	(0.644)
Efect in t+1	-0.0261	-0.00427	0.0226	-0.0107	-0.00601
	(0.649)	(0.689)	(0.316)	(0.434)	(0.681)
Efect in t+2	-0.0258	0.0132	0.0441	-1.83e-05	0.00890
	(0.769)	(0.443)	(0.389)	(0.999)	(0.692)
Efect in t+3	-0.0772	-0.000216	0.0321	-0.0218	-0.00575
	(0.477)	(0.990)	(0.221)	(0.485)	(0.787)
Pre-reform means	4.080	4.030	4.030	4.030	4.030
P-value (F-test of leads)	0.631	0.708	0.536		
Panel	Yes	Yes	Yes	Yes	Yes
Hospital FE	No	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Hospital Linear trends					Yes
N	900	900	622	900	900
R2	0.767	0.962	0.960	0.962	0.962

Notes: Robust standard errors clustered at county level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column 1 reports on the results of equation 3.1 without including hospital fixed effects while the specification in column 2 includes hospital fixed effects. The specification in column 3 is estimated on the sample of 101 public hospitals for which budget data is available and controls for a proxy for the annual budget. Column 4 reports on the results of the event study in equation 3.3 and controls for total staff hired in 2007 interacted with the year. Column 5 shows the results of the parametric event study in equation 3.2.

3.8 Conclusions

In this paper, I evaluate the impact on in-hospital mortality of a reform that made all health professionals working part-time switch to full-time contracts at public hospitals in Ecuador. The reform was born from the need to homogenize wages at all public institutions, not in response to deficiencies in certain areas of the public health system. Adoption was progressive according to the resolutions issued by the Ministry of Labor which established a particular adoption date for each of the different groups of hospitals that constitute the public health system (Ministry of Health, IESS, and others), and a more specific schedule that set the order of adoption for hospitals within those groups. I argue that the timing of adoption was exogenous to hospital performance. This statement is based on the evidence from parallel trend analysis showing that three years before adoption, mortality was not different across hospitals that adopted the reform earlier and later. Furthermore, the results from the event study were robust to the inclusion of pre-treatment characteristics which could have been used by policymakers to decide the order of implementation.

The results from the event study for the complete sample of admissions show that the reform caused a reduction in mortality of 0.06% on the adoption year and of 0.1% one year later. For the sample of admissions to the emergency department, results are bigger in magnitude and remain statistically significant even after including additional controls like other hospital expenditures and pre-reform characteristics. Mortality decreased by 0.1% on the adoption year and by 0.2% one year later. The reduction in mortality was driven by a drop in mortality rates within the first 48 hours of admission, while changes in mortality after 48 hours of admission were not statistically significant.

The results were stronger for the subsample of admissions to the emergency department possibly because a second objective of the reform was to guarantee continuous attention in emergencies and hospitalization. So the emphasis was placed on increasing the availability of physicians in the emergency and surgery departments, which would also explain that effects were driven by a reduction in mortality within the first 48 hours of admission. Unfortunately, the information on hospital budgets does not allow me to identify the funds that were allocated to each department, so I can not present more evidence in this regard.

These results are also consistent with nonrandom selection of patients to hospitals, which leads to the underestimation of the intention to treat effects of the labor reform on the complete sample of admissions. It is possible that sicker patients normally seek treatment at more specialized hospitals, which would counteract the beneficial effect of an increase in the number of doctor hours due to the reform in those hospitals. Consequently, the true effects of the reform should be closer to the observed on the subsample of admissions to the emergency department.

I tested several specifications to tackle possible concerns of endogeneity in adoption and the results remained robust to the inclusion of pre-treatment characteristics and other time-variant controls. More importantly, I provided evidence that the effects of the labor reform reported in this paper can confidently be attributed to the labor reform and not to changes in other quality indicators at the hospital level like the length of stay or by changes in the patient mix.

All in all, the evidence provided in this study is consistent with the hypothesis that increasing staffing levels at hospitals reduce in-hospital mortality. The reform reduced the number of part-time professionals which is also expected to have a negative effect on in-hospital mortality rates (Bartel et al., 2014) resulting from reducing disruptions to team functioning that are more common in places with a high number of part-time and on-call workers. Results are also consistent with the literature that documents that job security increases the value of employment for the workers which in turn increases workers' unobservable effort, improving firms' outcomes (Griffith and Macartney, 2014). Unfortunately, data on the productivity per hour of work is not available at the individual level so it is not possible to disentangle the effect of an increase in the number of doctor-hours from changes in effort on hospital performance.

A labor reform like the one studied in this paper is also expected to have other welfare increasing effects. With more doctor-hours, hospitals can reduce waiting times, and treat more patients. This would also cause a reduction in out of pocket health care expenditures on the side of patients. However, the elimination of labor flexibility in the public sector is also expected to affect the labor supply in the private sector (Cheng et al., 2018; Faggio and Overman, 2014; Propper and Van Reenen, 2010). I have requested individual-level data that will allow me to identify the health professionals that were affected by the reform, and how they reacted to the reform, by changing their working hours or leaving or joining the public sector. This information will allow me to study in the future the spillover effects of the labor reform on private hospitals' performance.

References

Aiken, Linda H, Sean P Clarke, Douglas M Sloane, and Julie Sochalski (2017). "Hospital Nurse Staffing and Patient Mortality , Nurse Burnout, and Job Dissatisfaction". In: *Journal of the American Medical Association* 288.16, pp. 1987–1993.

Autor, David (2003). "Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing". In: *Journal of Labor Economics* 21.1, pp. 1–42.

Bartel, By Ann P, Nancy D Beaulieu, Ciaran S Phibbs, and Patricia W Stone (2014). "Human Capital and Productivity in a Team Environment: Evidence from the Hea...: DISCOVER: all subjects". In: *American Economic Journal: Applied Economics* 6.2, pp. 231–259.

Bell, Chaim M. and Donald A. Redelmeier (2001). "Mortality among Patients Admitted to Hospitals on Weekends as Compared with Weekdays". In: *New England Journal of Medicine* 345.9, pp. 663–668.

Besley, Timothy and Anne Case (2000). "Unnatural Experiments? Estimating the Incidence of Endogenous Policies". In: *The Economic Journal* 110.467, F672–F694.

Blundell, Richard and Thomas Macurdy (1999). *Chapter 27 Labor supply: A review of alternative approaches*. Vol. 3 PART. 1. Elsevier Masson SAS, pp. 1559–1695.

Broadway, Barbara, Guyonne Kalb, Jinhu Li, and Anthony Scott (2017). "Do Financial Incentives Influence GPs' Decisions to Do After-hours Work? A Discrete Choice Labour Supply Model". In: *Health Economics (United Kingdom)* 26.12, e52–e66.

Cappellari, Lorenzo, Carlo Dell'Aringa, and Marco Leonardi (2012). "Temporary Employment, Job Flows and Productivity: A Tale of Two Reforms". In: *The Economic Journal* 122.562, F188–F215.

Card, David, Carlos Dobkin, and Nicole Maestas (2009). "Does Medicare save lives?" In: *The Quarterly Journal of Economics* 124.2, pp. 597–636.

Carneiro, Pedro, Emanuela Galasso, and Rita Ginja (2018). "Tackling Social Exclusion: Evidence from Chile". In: *The Economic Journal*.

Chay, Kenneth, Patrick McEwan, and Miguel Urquiola (2005). "That Use Interventions The Central Role of Noise in Evaluating Test Scores to Rank Schools". In: *The American Economic Review* 95.4, pp. 1237–1258.

Cheng, Terence C., Guyonne Kalb, and Anthony Scott (2018). "Public, private or both? Analyzing factors influencing the labour supply of medical specialists". In: *Canadian Journal of Economics* 51.2, pp. 660–692.

Cingano, Federico, Marco Leonardi, Julian Messina, and Giovanni Pica (2016). "Employment Protection Legislation, Capital Investment and Access to Credit: Evidence from Italy". In: *The Economic Journal* 126.595, pp. 1798–1822.

Cook, Andrew, Martin Gaynor, Melvin Stephens, and Lowell Taylor (2012). "The effect of a hospital nurse staffing mandate on patient health outcomes : Evidence from California ' s minimum staffing regulation". In: *Journal of Health Economics* 31.2, pp. 340–348.

Demetriades, D, J Murray, K Chralambides, K Alo, G Velmahos, P Rhee, and L Chan (2004). "Trauma fatalities: time and location of trauma deaths". In: *Journal of the American College of Surgeons* 198.1, pp. 20–26.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew Notowidigdo (2018). “The Economic Consequences of Hospital Admissions”. In: *American Economic Review* 108.2, pp. 308–352.

Dolton, Peter and Vikram Pathania (2016). “Can increased primary care access reduce demand for emergency care? Evidence from England’s 7-day GP opening”. In: *Journal of Health Economics* 49, pp. 193–208.

Doyle, Joseph J., John A. Graves, Jonathan Gruber, and Samuel A. Kleiner (2015). “Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns”. In: *Journal of Political Economy* 123.1, pp. 170–214.

Duflo, Esther (2001). “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment”. In: *American Economic Review* 91.4, pp. 795–813.

Evans, William N. and Beomsoo Kim (2006). “Patient outcomes when hospitals experience a surge in admissions”. In: *Journal of Health Economics* 25.2, pp. 365–388.

Faggio, Giulia and Henry Overman (2014). “The effect of public sector employment on local labour markets”. In: *Journal of Urban Economics* 79.December 2011, pp. 91–107.

Fisher, Elliott S., David E. Wennberg, Therese A. Stukel, Daniel J. Gottlieb, F. L. Lucas, and Etoile L. Pinder (2003). “The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care”. In: *Annals of Internal Medicine* 138.4, pp. 273–287.

Fisher, Jessica M and Timothy B Gardner (2012). “The “Golden Hours” of Management in Acute Pancreatitis”. In: *The American Journal of Gastroenterology* 107.8, pp. 1152–1156.

Frakes, Michael (2013). “Impact of Medical Liability Standards on Regional Variations in Physicians Behavior : Evidence from Adoption of National-Standard Rules”. In: *American Economic Review* 103.1, pp. 257–276.

Freyaldenhoven, Simon, Christian Hansen, and Jesse M. Shapiro (2018). “Pre-event trends in the panel event-study design”.

Giovanella, Ligia, Oscar Feo, Mariana Faria, and Sebastián Tobar (2012). *Health Systems in South America : Challenges to the universality, integrality and equity*. Tech. rep. May, pp. 1–835.

Gowrisankaran, Gautam and Robert J. Town (1999). “Estimating the quality of care in hospitals using instrumental variables”. In: *Journal of Health Economics* 18.6, pp. 747–767.

Griffith, Rachel and Gareth Macartney (2014). “Employment protection legislation, multinational firms, and innovation”. In: *Review of Economics and Statistics* 96.1, pp. 135–150.

Herlitz, Johan, Birgitta Wireklintsundström, Angela Bång, Annika Berglund, Leif Svensson, and Christian Blomstrand (2010). “Early identification and delay to treatment in myocardial infarction and stroke: differences and similarities.” In: *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 18, p. 48.

Jayawardena, Suriya, Joseph R. Lauro, Jacob Eisdorfer, Shalaka Indulkar, Anomadarshi Barua, and Sundara Sridhar (2007). “Death within 48 hours of admission to the emergency department: The value of autopsy”. In: *American Journal of the Medical Sciences* 334.2, pp. 87–91.

Karoshi, Mahantesh and Louis Keith (2009). “Challenges in managing postpartum hemorrhage in resource-poor countries”. In: *Clinical Obstetrics and Gynecology* 52.2, pp. 285–298.

Latense, Barbara A. (2009). “Critical Care of the Burn Patient: The First 48 Hours”. In: *Critical Care Medicine* 37.10, pp. 2819–2826.

Lin, Haizhen (2014). “Revisiting the relationship between nurse staffing and quality of care in nursing homes: An instrumental variables approach”. In: *Journal of Health Economics* 37.1, pp. 13–24.

Matsudaira, Jordan D. (2014). “Government Regulation and the Quality of Healthcare.” In: *Journal of human resources* 49.1, pp. 32–72.

Merino, Cristina (2006). “Empleo de los recursos humanos de salud en instituciones del sector público”. In: *Serie Observatorio de Recursos Humanos de Salud. CONASA* 3, pp. 1–119.

Needleman, Jack, Peter Buerhaus, Soeren Matteke, Maureen Stewart, and Katya Selevinsky (2002). “Nurse staffing levels and the quality of care in hospitals”. In: *N Engl J Med* 346.22, pp. 1715–1722.

Needleman, Jack, Peter Buerhaus, Shane Pankratz, Cynthia L. Leibson, Susanna R. Stevens, and Marcelline Harris (2011). “Nurse staffing and inpatient hospital mortality.” In: *The New England journal of medicine* 364.25, pp. 1037–1045.

Pham, Julius Cuong, Mary Andrawis, Andrew D. Shore, Maureen Fahey, Laura Morlock, and Peter J. Pronovost (2011). “Are temporary staff associated with more severe emergency department medication errors?” In: *Journal for Healthcare Quality* 33.4, pp. 9–18.

Pronovost, Peter J., Mollie W. Jenckes, Todd Dorman, Elizabeth Garrett, Michael J. Breslow, Brian A. Rosenfeld, Pamela A. Lipsett, and Eric Bass (1999). “Organizational Characteristics of Intensive Care Units Related to Outcomes of Abdominal Aortic Surgery". In: *JAMA: The Journal of the American Medical Association* 281.14, p. 1310.

Propper, Carol and John Van Reenen (2010). "Can Pay Regulation Kill? Panel Data Evidence on the Effect of Labor Markets on Hospital Performance". In: *Journal of Political Economy* 118.2, pp. 222–273.

Sircar, Padmini, Darshan Godkar, Shmuel Mahgerefteh, Karinn Chambers, Selva Niranjan, and Robert Cucco (2007). "Morbidity and mortality among patients with hip fractures surgically repaired within and after 48 hours". In: *American Journal of Therapeutics* 14.6, pp. 508–513.

Sochalski, Julie, R. Tamara Konetzka, Jingsan Zhu, and Kevin Volpp (2008). "Will mandated minimum nurse staffing ratios lead to better patient outcomes?" In: *Medical Care* 46.6, pp. 413–452.

Swami, Megha, Hugh Gravelle, Anthony Scott, and Jenny Williams (2017). "Hours worked by general practitioners and waiting times for primary care". In: *Health Economics* 27.May, pp. 1–25.

Tong, Patricia K. (2011). "The effects of California minimum nurse staffing laws on nurse labor and patient mortality in skilled nursing facilities". In: *Health Economics* 20.7, pp. 802–816.

Valdez, Carrie, Babak Sarani, Hannah Young, Richard Amdur, James Dunne, and Lakhmir S. Chawla (2016). "Timing of death after traumatic injury. A contemporary assessment of the temporal distribution of death". In: *The Journal of Surgical Research* 200.2, pp. 604–609.

3.9 Appendix

Table 3.9: Characteristics of hospitals in the sample of 101 public hospitals for which budget data is available

	2009	2010	2011	2012	2013	2014
Admissions	12487	13005	13564	13191	13019	12225
Mortality	0.01	0.01	0.01	0.01	0.01	0.01
Mortality within 48 hours	0.00	0.00	0.00	0.00	0.00	0.00
Mortality after 48 hours	0.01	0.01	0.01	0.01	0.01	0.01
Mortality (ED)	0.02	0.03	0.03	0.02	0.02	0.03
Mortality within 48 h. (ED)	0.01	0.01	0.01	0.01	0.01	0.00
Mortality after 48 h. (ED)	0.02	0.02	0.02	0.01	0.02	0.03
Full-time staff	61.39	79.20	183.59	248.49	251.25	285.23
Part-time staff	160.70	157.57	71.75	32.37	18.96	28.97
Occasional staff	0.72	0.68	0.51	0.63	0.33	0.06
Full-time equiv. doctors	168.74	182.83	229.00	268.49	262.97	303.31
Patients - staff ratio	79.30	78.23	67.56	56.74	60.22	52.21
Length of stay (days)	4.41	4.40	4.36	4.31	4.41	4.59
Case mix index	1.42	1.42	1.43	1.42	1.43	1.43
Imaging equipment	7.09	7.83	8.50	8.41	9.45	9.61
Diagnostic equipment	8.27	8.90	9.38	10.40	10.79	14.15
Treatment equipment	14.53	13.11	16.88	13.52	16.21	17.18
Operating rooms	9.70	9.53	9.49	7.13	8.37	7.27

Notes: The sample corresponds to the 101 public hospitals for which budget data is available.

Chapter 4

THE EFFECTS OF MALPRACTICE REFORM ON PROCEDURE CHOICE AND PATIENT OUTCOMES

Abstract

This paper studies the impact of increased liability risk facing physicians, on the use of c-sections and on indicators of maternal and infant health. I take advantage of a legal reform that led to the hardening of sentences for cases of professional malpractice in Ecuador. I use a difference-in-difference strategy that compares the outcomes of two neighboring countries, Ecuador and Colombia, before and after the introduction of the reform and perform several parallel trend tests on the outcomes of interest and test for the stability of the demographic composition of both countries to support my identification strategy. During the five quarters following the reform, Ecuadorian doctors reduced the c-section rate by 1.1% among women aged 15 to 24 years, and by 0.9% among women aged 25 to 34 years. The c-section rate remained unaffected for women aged 35 to 44 years, possibly because doctors have less discretion over riskier births. Interestingly, the observed reduction in the c-section rate did not affect the health outcomes of mothers or newborns.

4.1 Introduction and literature review

Financial incentives play an important role in doctors decisions (McGuire, 2000). This argument has been identified as one of the main reasons behind rising health care costs in many countries (Kessler and McClellan, 1996; Shurtz, 2014). Doctors value their own income as well as patients' health, so when procedures are elective and relatively low-risk, they tend to prescribe the most profitable treatment (Shurtz, 2014). In this scenario, the medical malpractice liability system has two main objectives: providing redress to individuals who suffer negligent injuries, and creating incentives for doctors to provide appropriately careful treatment to their patients (Kessler and McClellan, 1996).

A large literature has found that the liability system is an important determinant of health care provision (Baicker et al., 2006; Frakes, 2013; Kerschbamer and Sutter, 2017; Kessler and McClellan, 1996). In the field of obstetrics, Baicker et al. (2006) argue that mother's characteristics only explain 8.6% of the geographic variation in c-section rates, while county socio-economic characteristics account for 27.3% of the variation, followed by medical malpractice liability that accounts for 14.8% of the variation. This paper evaluates the impact of increased liability risk on procedure choice in obstetrics and linked to this, it also looks at the impacts on infant health and maternal mortality. I focus on c-sections because this is a commonly used elective procedure that has been raising in the last decade in both developed and developing countries (Schulkind and Shapiro, 2014). Furthermore, obstetrics is a specialty that is particularly exposed to malpractice risk due to possible complications during labor. Some of the risks associated to c-sections include infections, hemorrhage, injury to organs for the mother, complications with the anesthesia and brain damage in newborns due to oxygen deprivation arising from delays in performing a c-section (Dubay et al., 1999).

The theoretical models of procedure choice in response to changes in malpractice pressure predict that a legal change that increases physician liability will decrease procedure use if and only if the error rate when the procedure is chosen is higher than the error rate when the procedure is not chosen (Currie and MacLeod, 2008). Then, if the marginal procedures are being performed on a low (high) risk patient for whom the risk of performing the procedure is higher (lower) than not performing it, it is expected that an increase in the liability pressure will decrease (increase) procedure use.

The early literature on the effects of liability fears on procedure use reports mixed evidence (Tussing and Wojtowycz, 1992; Baldwin et al., 1995; Sloan et al., 1997; Localio et al., 1993). These studies are problematic because are based on the analysis of cross section data and are prone to omitted variable bias. More recently, Dubay et al. (1999) uses the variation in the timing of implementation of tort reforms across states to instrument malpractice premiums and found higher c-section use when malpractice claims risk was higher.[1]

On the other hand, Currie and MacLeod (2008) argue that while the probability of a suit can respond quickly to tort reforms, malpractice premiums are affected by tort reforms only with a relatively long and uncertain lag. For instance, in the US, the mean time between an injury and the settlement of a claim is six years. Consequently, Currie and MacLeod (2008) directly evaluate the impact of different types of tort reforms on procedure use and the health outcomes of mothers and their infants. They find that reform of the rule of joint and several liability, which increased liability risk, reduced c-sections, but caps on noneconomic damages, that decreased liability risk, increased them. They also found that the observed increases in procedure use induced by caps on noneconomic damages did not affect infant

[1] In another paper, Dubay et al. (2001) study whether malpractice liability pressure causes physicians to limit their supply of obstetric services including prenatal care. They found that malpractice liability pressure was associated with greater prenatal care delay and fewer prenatal care visits but that such pressure did not affect infant health.

health.

In this paper, I exploit the exogenous variation on malpractice pressure induced by a reform of the Penal Code in Ecuador that led to the hardening of sentences for cases of professional malpractice. The reform is an exogenous event to the health system since it was a consequence of a major reform to the Penal Code which had not been reformed since 1971. Furthermore, there is no evidence suggesting that the inclusion of the articles related to malpractice in the new Penal Code responded to an unusual growth in the number of malpractice lawsuits in the country.

My analysis of the impact of changes in malpractice pressure on doctors' decisions and ultimately on patients' outcomes is more in line with the study by Currie and MacLeod (2008) that takes a reduced form approach and directly evaluates the impact of different tort reforms on c-section rate using panel data methods. I use a difference-in-difference strategy that compares the outcomes of two neighboring countries, Ecuador and Colombia, before and after the introduction of the malpractice reform. Both countries are very similar in terms of several socio-demographic characteristics of their populations and in terms of access to health services for pregnant women. Furthermore, their malpractice legislations were very similar before the passing of the new Penal Code in Ecuador.

The high c-section rate observed in Ecuador and Colombia in the period prior to the reform suggests an excess of unnecessary c-sections. The malpractice reform studied in this paper creates a unique opportunity to asses whether the excess c-sections can be deemed as socially wasteful or not. For this, I test whether the observed reduction in c-sections is accompanied by a deterioration of infant or maternal health, in which case the extra c-sections are deemed necessary, otherwise they are deemed wasteful (Kessler and McClellan, 2002; Dubay et al., 2001).

Better decision making in response to the reform would result in a reallocation of procedures from fewer low-risk to high-risk cases (Currie and MacLeod, 2017; Currie and MacLeod, 2008). Along these lines, the results from this paper show that doctors responded to the reform by reducing the c-section rate among women aged 15 to 34 years but not among women aged 35 to 44 years. Furthermore, the reduction in the number of c-sections did not affect the health outcomes of mothers or newborns. Conversely, the share of low birth-weight babies decreased driven by the reduction in c-sections one quarter after the reform; and, early maternal mortality remained at the same pre-reform rates.

The contributions of this paper are threefold: First, it is to the best of my knowledge, the first study looking at the impact of malpractice reforms in a developing country. Second, is one of the few studies that evaluate a policy that effectively tackles the problem of overprovision of elective procedures in developing countries. Another study that addresses a similar question is the one by Cohen et al. (2015) that studies ways to better target subsidized drugs in Kenya. Third, this paper contributes more generally to the literature on over-treatment, which is one of the main drivers of health costs and a source of concern throughout the world (Kessler and McClellan, 1996; Adhvaryu, 2014).

The rest of the paper is organized as follows: Section 4.2 describes the Ecuadorian and Colombian Health Systems as well as their respective malpractice legislation. Section 4.3 discusses the empirical strategy which is based on a difference-in-difference model that compares the two neighboring countries before and after the Ecuadorian malpractice reform. Section 4.4 explains the process to homogenize the vital statistics and death certificates data of the two countries, to finally build a balanced panel at the municipality level which constitutes the main sample for this study. Section 4.5 presents a careful assessment of the validity of the identification strategy by showing evidence of parallel trends on the outcomes of interest before the reform in the two countries and of the stability in the demographic composition of both countries before and after the reform. Section 4.6 present the main results of the paper, followed by additional robustness checks in section 4.7, and section 4.8 concludes.

4.2 Background

4.2.1 Ecuador's Health system and malpractice legislation

The health system in Ecuador is mixed, composed by the public and the private sector. The public health system operates a fee for service model. Health care services are financed by general taxation and are provided free of charge at the point of use. The Ministry of Health (MSP) and the Social Security Institution (IESS) control most public hospitals and clinics and provide health services to the majority of the population.

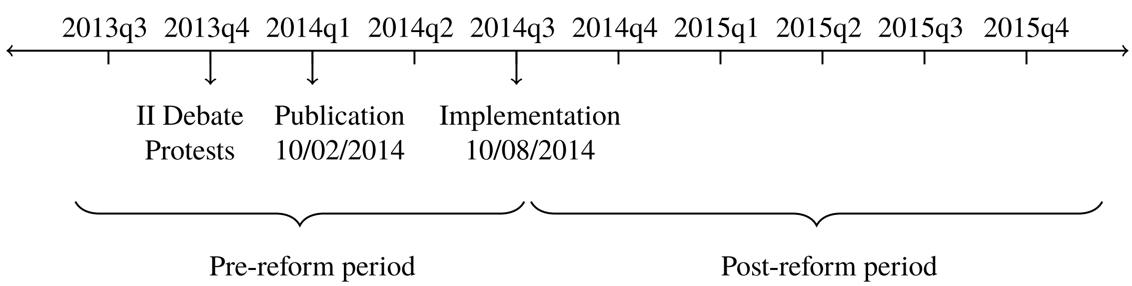
The referential price of a vaginal delivery (for reimbursement purposes) at a public hospital is around 78% the price of a c-section (Ministerio de Salud Pública del Ecuador, 2014). This difference is much higher in the private sector and would explain the high c-section rates observed in the private sector (Gruber, Kim, et al., 1999; Gruber and Owings, 1996; Currie, Gruber, et al., 1995).

A recent study about the prevalence of c-sections in Ecuador concluded that between 2001 and 2013, c-sections accounted for 57.5% of births in the private sector and 22.3% of births in the public sector but, only 51% of c-sections in the public sector were justified by parallel reports of prenatal maternal or fetal complications, including prenatal hypoxia, multiple pregnancies, or labor dystocia; while, in the private sector, 22% appeared to have such justification (Ortiz-Prado et al., 2017).

In Ecuador, as in many other countries, obstetricians and surgeons are particularly exposed to malpractice risk. Yet, the level of litigiousness is lower compared to countries with well developed malpractice liability systems like the US. Proof of this is that until August 10, 2014, the Ecuadorian Penal Code did not consider an autonomous criminal type for the cases of professional malpractice. The Penal Code of 1971 only typified unintentional homicide and unintentional injuries without distinguishing the cases when the author was a doctor. For this reason, most of the malpractice trials that did not involve the death of a patient used to be resolved in civil courts with sanctions involving reparation for damages but not jail time. This situation changed after the enactment of the new Penal Code.

The first and second debates to approve the new Penal Code occurred in June 2012, and October 9th, 2013, respectively. News about the latter was widely publicized by the media and on October 24, 2013, tensions reached a peak with doctors protesting in the streets threatening with quitting their jobs if the articles related to malpractice were included in the final version of the new Penal Code. Despite opposition from health-care professionals, the new Penal Code was published in the Official Registry with the status of law on February 10th, 2014, setting August 10, 2014, as the official implementation date. Figure 4.1 depicts the timeline of the reform.

Figure 4.1: Timeline of the reform



The new Penal Code contains two articles that define the punishments for cases of unintentional manslaughter and injuries linked to professional malpractice. Art.146 states that unintentional manslaughter due to professional malpractice shall be punished with imprisonment of one to three years. Aggravating circumstances could increase the punishment to five years. However, Art. 146 also describes a list of conditions that must be met in order to bring a case of professional malpractice, each of which has to be assessed by a medical specialist. Similarly, Art. 152 of the new Penal Code establishes that the punishment for a professional who injures a patient ranges from thirty days to seven years in jail.

According to the Office of the Prosecutor, in 2007, 89 malpractice complaints were filed in the country. In 2008, the number of complaints raised to 147, followed by 112 complaints in 2009 and 166 complaints in 2010.[2] All these figures before 2014 should be taken with caution though. It was only with the new classification of the criminal types, that it became much easier to distinguish cases of unintentional manslaughter due to professional malpractice from the other multiple cases of unintentional manslaughter[3] For that reason, official statistics are only available since August 2014. From that date, the number of claims has actually dropped from 176 in 2015 to 158 in 2016 and 110 in 2017.

The malpractice reform was enacted in August 2014 at the national level, eliminating the possibility of exploiting variation across areas that adopted the reform at different points in time. In similar circumstances, several papers have used a difference-in-difference strategy that compares changes in the outcomes of a treated country relative to a comparable country or a set of states that serve as a valid control group to evaluate the impact of a policy or reform in the treated country (Banerjee et al., 2002; Lindo and Packham, 2017). In the next section, I explain the empirical strategy and the reasons for choosing Colombia as a control group for Ecuador.

4.3 Empirical strategy

The difference-in-difference strategy makes the counterfactual assumption that the outcome of interest would have grown at the same rate in the treated and in the control

[2]Source: Proyecto de ley reformatoria al Código Penal para la tipificación de los delitos de mala práctica y falta de atención médica.

[3]This did not happen with injuries. All types of injuries are described in Art. 152. For that reason the Office of the Prosecutor does not keep statistics on the claims related to injuries caused by doctors.

country had the reform or policy change did not occur. While this assumption is not directly testable, it is possible to test if the outcome of interest was growing at the same rate in the pre-reform period in both groups. A second condition requires that the composition of both groups remain stable before and after the policy change (Blundell and Macurdy, 1999).

Identification also hinges on the exogeneity of the reform that is being studied. I argue that the malpractice reform in Ecuador was an exogenous event to the health system since it was a consequence of the reform to the Penal Code which had not been reformed since 1971. Furthermore, there is no evidence that indicates that the introduction of Arts. 146 and 152 responded to an unusual growth in the number of malpractice lawsuits in Ecuador.

To evaluate the malpractice reform, I compare municipalities in Ecuador and Colombia before and after the reform. Instead of estimating an average effect in the post-reform period, I estimate a flexible specification that allows the estimated effects to vary across time with a set of quarter indicators relative to the reform date. I estimate the following regression:

$$Y_{jt} = \alpha_j + \lambda_t + \sum_{k=-4}^{5} (Treat_j \times \delta_{jt}^{k})\gamma_k + X_{jt}\beta + \epsilon_{jt} \tag{4.1}$$

where Y_{jt} is the outcome of interest in municipality j, and quarter t, α_j are municipality fixed effects to account for any systematic difference across municipalities. λ_t are quarter fixed effects that control for time shocks that affect municipalities in the two countries in the same way. $Treat_j$ is an indicator which equals one if the municipality belongs to Ecuador and zero if belongs to Colombia.

Since the outcomes of interest are driven by fertility decisions and woman's preferences about the type of delivery, there is no reason to expect a discrete jump in the first quarter after the reform took place (Lindo and Packham, 2017). I allow the estimated effects to vary across time with a set of interaction terms between the treatment and time indicators relative to the third quarter of 2014 ($Treat_j \times \delta^k_{jt}$). For that let $\tilde{t} = 0$ on the third quarter of 2014, when the reform took place, and $\delta^k_{jt} \equiv 1\{t = \tilde{t} + k\}$.

Given that different municipalities will react in different ways to the malpractice reform, γ_k report on the intention to treat effects of the reform across municipalities at different points in time. The coefficients of the leads included in the model, allow testing the parallel trends assumption required for identification. The reference category corresponds to the second quarter of 2014. For the parallel trends test to be successful, the coefficients of the pre-treatment interactions relative to the omitted term should not be significantly different from zero. If this is true, then the coefficients of the 5 interaction terms $\sum_{k=1}^{5}(Treat_j \times \delta^k_{jt})$ after the reform correspond to the dynamic effects of the reform.

X_{jt} is the vector of time-varying controls that include the proportion of the black and indigenous population (the omitted category is mestizo) in the municipality, the share of women in different age groups and the average gross value added per capita in the municipality. In all the specifications, the errors ϵ_{jt} are clustered at the municipal level. To account for zeros in the data, the log transformation log(variable +1) is used for the main outcomes and for the control variables.

As a robustness check, I add municipality-specific linear trends, in order to address concerns about different preexisting time trends in the municipalities of the two countries which may introduce bias to the estimates. In said specification, the coefficients of interest γ_k, reflect the change in the outcome relative to any preexisting time trend.

To choose a valid control group for Ecuador, I first discarded other South American countries that have a federal organization and consequently different structures of their health systems compared to Ecuador; these include Argentina, Brazil, and Venezuela. I also discarded Chile because it is a high-income country, while Ecuador remains an upper-middle income country according to the classification of the World Bank. Besides, Chile and more recently Uruguay (via Law 18,211 of 2011) are countries with integrated insurance systems were social security coverage exceeds 70% of the population (Giovanella et al., 2012).

For geographic proximity and cultural similarities, the natural candidates are Colombia and Peru. The main difference between the two is that Peru had already an autonomous criminal type for cases of professional malpractice by the time of the reform in Ecuador, while Colombia did not (Asociación Nacional de Clínicas y Hospitales Privados del Ecuador, 2013).[4] Furthermore, the c-section rate in Peru in 2015 was 35.4%,[5] far below the c-section rate of 51% in Ecuador, while in Colombia the c-section rate was 45% in 2014. I also discarded Bolivia as the control group, because it registers an even lower c-section rate compared with Ecuador (33% in 2016).[6] Furthermore, by 2014 Bolivia also had a specialized criminal type for wrongful injuries linked to professional malpractice (Asociación Nacional de Clínicas y Hospitales Privados del Ecuador, 2013).

The identifying assumption in the difference-in-difference model is not violated if the countries that are being compared have different levels of the outcome of interest. It is only necessary that the outcomes grow at the same rate in both countries before the reform. However, a country with very different levels of the outcome of interest is likely to have different unobservable characteristics, for

[4] Another issue with Peru was the lack of available microdata on death certificates by the time I was working on the main analysis in this paper. In face of a deficient system to register deaths. In 2015, MINSA and RENIEC developed SINADEF, a web application that allows doctors to prepare an online death certificate.

[5] Source: Statistical bulletin of births Peru (2015)

[6] Source: https://www.minsalud.gob.bo/images/Documentacion/EDSA-2016.pdf

example, restrictions in medical protocols or little taste for surgical procedures that make the c-section rate low. Fixed unobserved characteristics are controlled for by country fixed effects, but, if the difference is explained by time variant unobservable characteristics or a different demographic composition in the country, this would violate the parallel trends assumption, since it is necessary that both countries have a similar composition, so that both countries respond to common shocks in the same way.

Based on the previous discussion and given that Ecuador and Colombia are very similar in terms of socio-demographic characteristics, Colombia was selected as the control group for Ecuador. Access to health care is similar in both countries because the Constitutions of both countries guarantee citizens the unalienable right to health care as a fundamental right (Escobar-Córdoba, 2012). Furthermore, it is against the law to deny access to health care at public or private establishments in case of emergencies in both countries. While in Ecuador, all citizens have unrestricted access to health care at public sector hospitals and clinics; in Colombia, there is a universal health insurance scheme whereby all citizens, irrespective of their ability to pay, are entitled to a comprehensive health benefits package.

In Colombia, the General Insurance System established the Subsidized Regime (SR) to cover people identified as being poor through a proxy means test, and the Contributory Regime (CR) to cover formal workers and their families (Giedion and Villar, 2009). The contributory regime in Colombia is equivalent to the Social Security regime for formal workers in Ecuador, while the subsidized regime, requires that poor individuals pay a small share of the basic premium, which costs approximately 137 US dollars per person per year (Giedion and Villar, 2009). Like in Ecuador, the Colombian government provides insurance for free for the poorest individuals and coverage for diseases that are not included in the Obligatory Health Plan through public hospitals. The Territorial Entities are responsible for the poor uninsured population (PNNA), who receive health care for free with charge to the resources of the General System of Participations.[7]

In Ecuador, workers with Social Security affiliation have maternity coverage and uninsured individuals can seek attention at public hospitals free of charge at every level of care. Similarly, in Colombia, the Obligatory Health Plan covers the provision of health services in prenatal care, delivery care and postpartum control in the subsidized and contributory regimes. The insurance rate in Colombia is close to 90% while in Ecuador by 2012, it was close to 45% (Naranjo, 2014). However, this does not mean that the uninsured population in Ecuador have no access to health care services. The public health system guarantees unrestricted access to health care to the uninsured population. Hence, a fairer comparison between the two countries would contrast the share of the Colombian population in the contributory regime (41.45% in 2013) to the share of the Ecuadorian population with social insurance (41.2% in 2012)[8]; and, to assess the differences in terms of access to health care between the two countries, one should look at indicators like the share of births at a hospital or clinic or the share of births assisted by a doctor or obstetrician, which are very similar in both countries.

Regarding the malpractice legislation in Colombia, the current Penal Code does not consider an autonomous criminal type for the cases of professional malpractice and when a patient dies, as a result of malpractice, the legal figure of unintentional manslaughter is used, just as it used to be in Ecuador before the passing of the new Penal Code. Furthermore, when dealing with injuries, sanctions in Colombia range from one to three years of imprisonment for injuries that cause inability to work beyond thirty days without exceeding ninety days, and from two to five years for injuries that cause inability to work for more than 90 days.[9] Similarly, in Ecuador,

[7] By 2014, 859,090 individuals were in this category.

[8] Source: Giedion and Villar (2009) and Naranjo (2014)

[9] Source: Colombian Penal Code published in Law 599 published in July 2000.

before the passing of the new Penal Code, injuries causing inability to work below 30 days used to be punished with imprisonment from six months to two years and injuries causing inability to work beyond 90 days used to be sanctioned with imprisonment from one to three years.[10]

There is very limited information about the number of malpractice claims filled in a year in Colombia because this information is not shared in a systematic way with the public (Tamara et al., 2012). However, the study by Tamara et al. (2012) mentions that between January 1st, 2006 and December 31st, 2010, the authors were able to collect information on 535 reports related to 402 cases, corresponding to 82% of the reports written during the study period. Once Again, the number of malpractice claims is very similar to the one observed in Ecuador, which gives an idea that the level of litigiousness in both countries is also comparable.

Ecuador and Colombia are also similar in terms of socio-demographic characteristics. Table 4.1 compares the characteristics of women of fertile age who gave birth four quarters before the reform and five quarters later in Ecuador and in Colombia. In both countries, the share of mixed-race population is above 90% before and after the reform and the percentage of mothers in each age bracket is also fairly balanced across the two countries. Furthermore, in terms of education, in both countries around 20% of the population has achieved some level of postsecondary education.

Regarding social security coverage among women of fertile age, 57% of the Colombian women in the pre-reform period reported being uninsured or having subsidized insurance, while the share of women with subsidized insurance or uninsured in Ecuador was 52%. Access to health services was also very similar in the two countries. In the pre-reform period, 99% of Ecuadorian and Colombian mothers were assisted by a doctor or obstetrician during childbirth and the average number of prenatal checks was 6.96 in Ecuador and 6.37 in Colombia. Another indicator that

[10] Source: Ecuadorian Penal Code of 1971 published in the Official Registry 147.

Table 4.1: Descriptive statistics for Ecuador and Colombia

	Pre-reform		Post-reform	
Variables	Ecuador	Colombia	Ecuador	Colombia
Number of counties or municipalities	758	192	758	192
Percent black	0.05	0.02	0.05	0.03
Percent indigenous	0.02	0.05	0.02	0.05
Percent mixed-race	0.93	0.92	0.92	0.92
Percent with less than high school education	0.78	0.78	0.75	0.79
Percent with more than high school education	0.22	0.21	0.24	0.20
Percent female 15-19 years old	0.22	0.18	0.21	0.21
Percent female 20-24 years old	0.30	0.28	0.29	0.27
Percent female 25-29years old	0.22	0.24	0.23	0.23
Percent female 30-34 years old	0.16	0.18	0.16	0.17
Percent female 35-39 years old	0.08	0.09	0.08	0.09
Percent female 40-44 years old	0.02	0.03	0.02	0.02
Percent female 35 years or older	0.10	0.12	0.10	0.11
Percent of births assisted by a doctor	0.99	0.99	0.99	0.99
Born at a hospital	0.99	0.98	0.99	0.98
Born at home	0.01	0.02	0.01	0.02
Contributory regimen	0.43	0.48	0.45	0.44
Subsidized regimen or uninsured	0.57	0.52	0.55	0.56
Mean prenatal checks	6.37	6.96	6.47	6.94
Percent babies born male	0.51	0.51	0.51	0.51
C-section rate	0.45	0.51	0.46	0.50
Percent of low birth-weight babies	0.08	0.08	0.08	0.08
Gross value added per capita	7,762.41	6,270.66	6,335.89	6,223.86
Maternal mortality rate	0.05	0.06	0.05	0.06
Percent female 35 years or older who died	0.44	0.43	0.44	0.43

Source: The sample is women of fertile age in Vital Statistics data and death certificates data. The period of analysis is four quarters before the reform and five quarters later. The weights reflect the number of births in the municipality as a share of the total in the country or, for variables related to maternal mortality, the number of early maternal deaths in the municipality as a share of the total deaths among women of fertile age in the country.

is very similar in both countries is the share of low birth-weight babies, which is almost identical in both countries and is close to 8%. Finally, the maternal mortality rate was 5% in Colombia and 6% in Ecuador.

The information from the Health Nutrition and Population Statistics of the World Bank allow establishing other similarities between the two countries. For instance, by 2014, the prevalence of anemia among pregnant women in Colombia was 27.5% and it was 26.1% in Ecuador. Furthermore, the number of Hospital beds per 1,000 people was 1.4 in 2011 for Colombia and 1.6 for Ecuador. Maternal leave benefits – measured as a percentage of wages paid in the covered period – are covered in its entirety in both counties, while the health expenditure per capita, in current US dollars was $855.96 in Colombia and $994.40 in Ecuador in 2014 (Purchasing Power Parity adjusted). Finally, age at first marriage for women in Colombia was 22.3 years in 2010 and 21.8 years in Ecuador.

All in all, both countries are very similar in terms of socio-demographic characteristics, in terms of health care coverage for pregnant women and in their malpractice legislation before the passing of the new Penal Code in Ecuador. On top of this, the National Statistics Offices in both countries collect Vital Statistics and Death certificates data in a very comparable fashion and with the same periodicity.

4.4 Data and sample

In this section, I describe the two sources of administrative data used to evaluate the impact of the malpractice reform on the c-section rate and on indicators of maternal and infant health. These are vital statistics and death certificates data for Ecuador and Colombia. I explain the outcomes that have been selected as proxies for infant and maternal health and describe the samples that have been chosen for the different analyses.

4.4.1 Vital Statistics data

I use Vital Statistics data for Ecuador and Colombia to investigate the impact of the malpractice reform on c-section utilization and infant health. The latest waves of microdata available are for 2016 for both countries. The Ecuadorian data is compiled by the National Statistics Office (INEC) and contain individual-level information of all births registered in the country. The data contain the precise date of birth, birth-weight, the height of the newborn, type of delivery, age, race, education level of the mother, number of prenatal checkups among other variables. There is also information about the place where the birth took place (province, county, and parish) and the type of health-care establishment (public, private). Income information is not recorded on Vital Statistics data, so I used information about gross value added per capita at the municipality level generated by the Ecuadorian Central Bank and by the National Statistics Office of Colombia.

Vital Statistics data for Colombia is compiled by the National Statistics Office (DANE) and contains almost the same information with some differences. First, the type of delivery began to be recorded in Ecuador in 2011 while a longer series is available for Colombia. Second, some variables are continuous in the Ecuadorian data but for Colombia are recorded as categorical variables, examples are birth-weight and height of the newborn and the age of the mother. Third, DANE collects the Apgar score of newborns and the health insurance status of the mother, while, INEC does not. Fourth, DANE publishes definitive data so the birth date of the baby always corresponds to the data collection year, while the Ecuadorian data presents several discrepancies between the birth date and the registration date. For this reason, for Ecuador, I kept only babies born and registered in the same year or one year later. Since the last available wave is for 2016, I do not use that wave in any of the analyses. Despite these differences in data processing time,[11] by 2015, the indicator that measures the completeness of birth registration – the percentage of children under age 5 whose births were registered at the time of the survey – in Colombia was 98.6% and for Ecuador was 94%.[12]

Finally, DANE records the type of health insurance of the mother and the hospital ID where the birth took place but not the type of hospital (public or private), while INEC does not contain information on health insurance or hospital IDs but it records the type of hospital where the birth took place which can be used to infer the type of insurance of the mother. I classified mothers into two groups according to their insurance status: (i) women in the contributory regime and (ii) women

[11]The reason is that in Colombia the process to register births and deaths is electronic, while in Ecuador it is not.

[12]Source: Health Nutrition and Population Statistics of the World Bank.

in the subsidized regime or uninsured. Women in the first group are Colombian women in the contributory regime or special regimes (teachers, police officers, oil companies, etc) and Ecuadorian women who gave birth at an establishment that belongs to IESS, the private sector, non-profit organizations or in special regimes (police officers, army, etc). Women in the second group are Colombian women who belong to the subsidized regime or reported being uninsured and Ecuadorian women who gave birth at home or a hospital that belongs to the Ministry of Health. It is worth noting that women who choose to give birth at home are commonly uninsured indigenous women.

The first outcome of interest to this study is the type of delivery. INEC began collecting this information in 2011. Because of this, the c-section rate for Ecuador has an artificially steep slope until mid-2012 when more mothers started reporting that information. Hence, to study the effects of the reform on c-section utilization, I restrict the sample to births that occurred four quarters before the third quarter of 2014 and five quarters after that date. In the robustness checks section, I extend the period of analysis to eight quarters before and five quarters after the reform date.

The other outcome I am interested in is birth-weight, which is a measure of infant health as well as the Apgar score. The Apgar score is known to be a reliable indicator of acute infant health at birth that has been demonstrated to be correlated with both morbidity and mortality (Dubay et al., 2001). Both birth-weight and Apgar score are expected to be responsive to obstetrician behavior whereas other measures of infant health like infant mortality are associated with the behavior of pediatricians (Dubay et al., 2001).[13]

Unfortunately, the Apgar score is only recorded in the Colombian Vital Statistics dataset and is not available for Ecuador, so I focus on birth-weight. birth-weight

[13] APGAR stands for activity, pulse, grimace, appearance, and respiration. The infant is given a maximum score of 2 for each attribute, for a maximum score of 10.

is reported as a categorical variable for Colombia and as a continuous variable for Ecuador, so I build a low-birth-weight indicator equal to one if birth-weight is below 2500 grams and zero otherwise (Baicker et al., 2006).[14]

4.4.2 Death certificates data

To study the impact of the malpractice reform on maternal mortality, I use death certificates data for Ecuador and Colombia. These datasets contain similar information in both countries. Late registration of deaths occurs in Ecuadorian data, so for any particular year, I only keep deaths registered in the same year or the next year. The data contain a specific set of questions to identify and study maternal mortality. It records the exact date of death and distinguishes between maternal deaths that happened during pregnancy or within 42 days of childbirth and late maternal deaths, which occurred within 11 months of childbirth. I focus on maternal deaths within 42 days of childbirth and ignore late maternal deaths because those could be related to other factors unrelated to the physician's work.

4.4.3 Analytical samples

The final sample from Vital Statistics contains women who gave birth within four quarters before the enactment of the new Penal Code and five quarters after that date. I kept only babies born and registered in the same year or one year later, so, considering that the last wave of Vital Statistics data available for Ecuador corresponds to 2016, the final sample includes data up to the fourth quarter of 2015.[15] The sample includes women aged 15 to 44 years, as this is considered the standard age range for fertility.[16] I also restricted the sample to singletons as

[14] The categories in Colombian data are: 1) below 1000 g; 2) from 1000 g to 1499 g; 3) from 1500 g to 1999 g; 4) from 2000 g to 2499 g; 5) from 2500 g to 2999 g; 6) from 3000 g; to 3499 g; 7) from 3500 g to 3999; 8) above 4000 g.

[15] This left me with around 6% fewer observations corresponding to late registrations.

[16] Because Vital Statistics data for Colombia only report the age of the mother in categories it was not possible to select women aged 15 to 45 but aged 15 to 44.

c-section is more common in cases of twins or triplets.

The equivalent for Ecuadorian counties are municipalities in Colombia.[17] There are 221 counties in Ecuador (hereafter municipalities), and 1,122 municipalities in Colombia. I aggregated the data at the municipal level to build a balanced panel of municipalities. After dropping the municipalities that did not register births in all the periods, the final sample preserved 99.57% of the records in both countries during the period of analysis. The final panel contains 192 municipalities in Ecuador and 758 municipalities in Colombia.

Regarding Death Certificates data, the final sample also includes women aged 15 to 44 years, but since there are significantly fewer observations about reported deaths, and it is expected that not every municipality registers maternal deaths in every period, I kept all the observations in the unbalanced panel.

4.5 Assessing the validity of the identification strategy

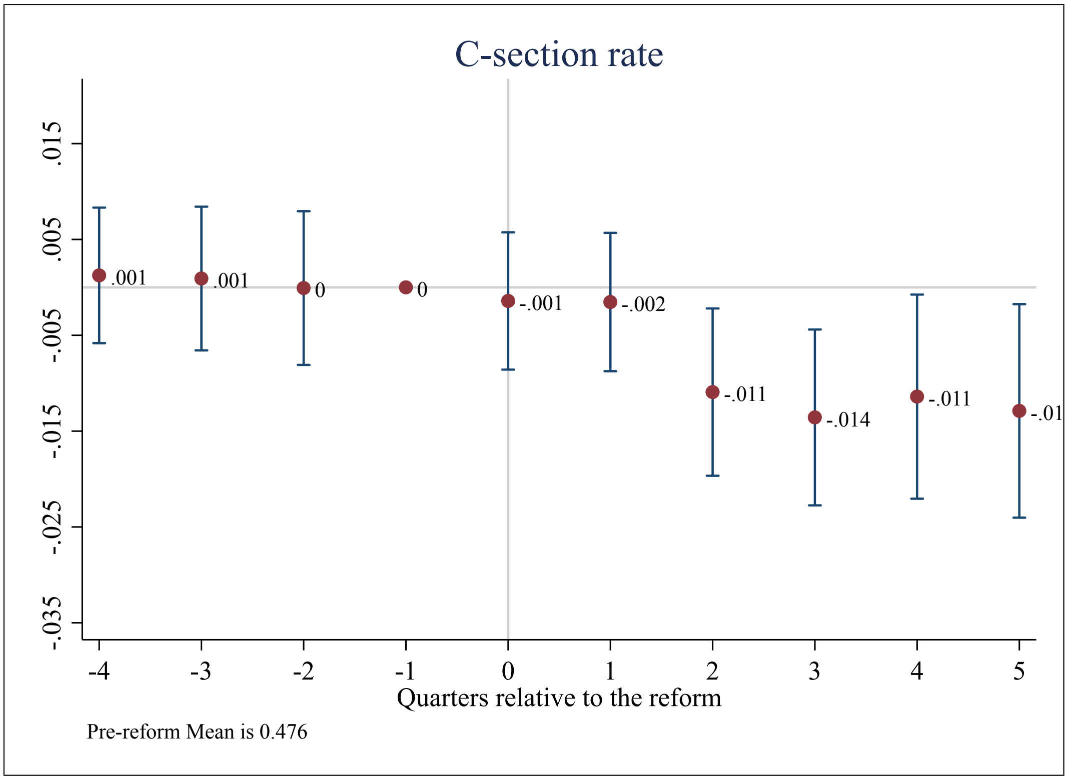
The evidence presented so far shows that Colombia is a good control group for Ecuador. In this section, I show further evidence to back this argument. For this, I show that the outcomes of interest follow similar trends before the reform in the two countries. I also present evidence of the stability in the demographic composition of both countries before and after the reform.

4.5.1 Parallel trends analysis

To formally test the parallel trends assumption on the three outcomes of interest. I estimate the difference-in-difference model in equation 4.1 and check whether the coefficients of the leads in the models are statistically different from zero. This is equivalent to test that there were no differential trends in the outcomes of Ecuador

[17]The term county and municipality are used interchangeably in the Ecuadorian Municipal Regime Law published in the Official Registry 331 in 1971; however, INEC refers to municipalities as counties in the Ecuadorian Vital Statistics data.

Figure 4.2: γ_k estimates from the difference-in-difference regression in equation 4.1 when the outcome is the c-section rate

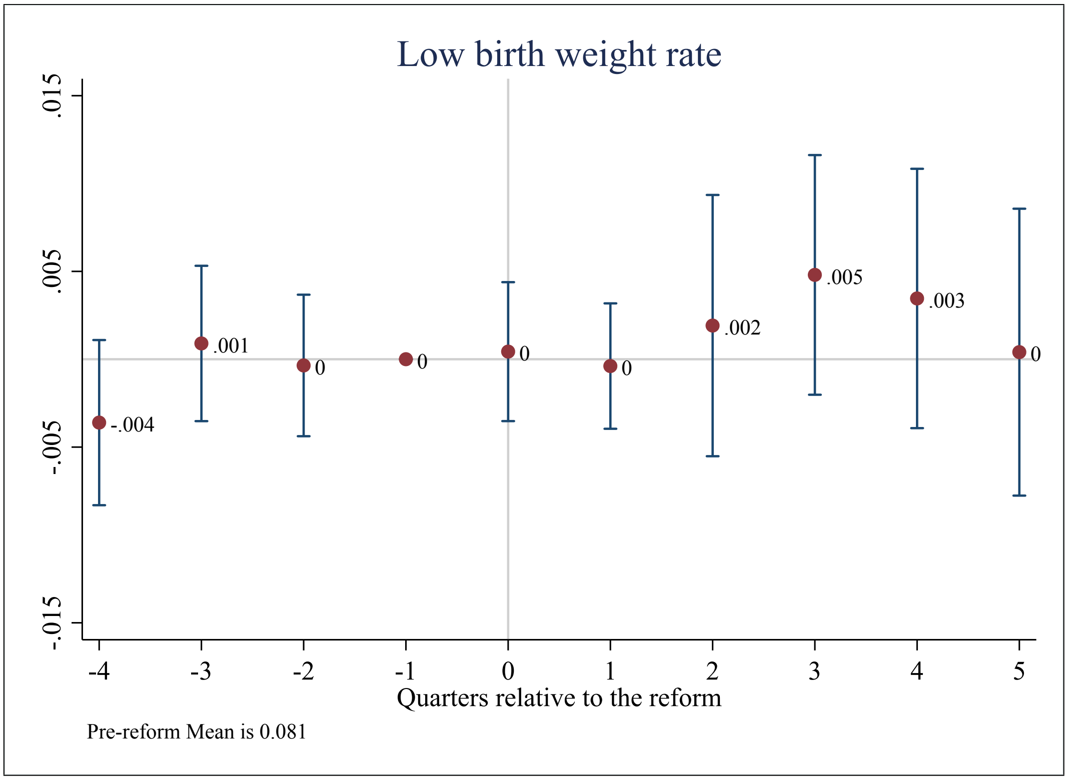


Notes: The sample is the balanced panel of municipalities and the period of analysis is four quarters before the reform and five quarters later in both countries. The points in each figure represent the γ_k coefficients from the difference-in-difference model in equation 4.1. The range plot with capped spikes present the 95 percent confidence intervals. The reference period is the second quarter of 2014. The specification includes municipality fixed effects. Controls include the proportion of black and indigenous population (the omitted category is mestizo), the share of women in different age groups and the average gross value added per capita in the municipality. Municipality linear time trends are not included.

and Colombia before the reform date. The reference period in all the analyses is the second quarter of 2014, which corresponds to the quarter immediately before the passing of the new Penal Code. The specifications include municipality fixed effects and time variant controls to account for differences in racial composition, age composition and average gross value added per capita across municipalities.

Figure 4.2 shows the pattern in the coefficients of the leads and lags in the model where the dependent variable is the logarithm of the c-section rate. It shows that

Figure 4.3: γ_k estimates from the difference-in-difference regression in equation 4.1 when the outcome is the low-birth-weight rate

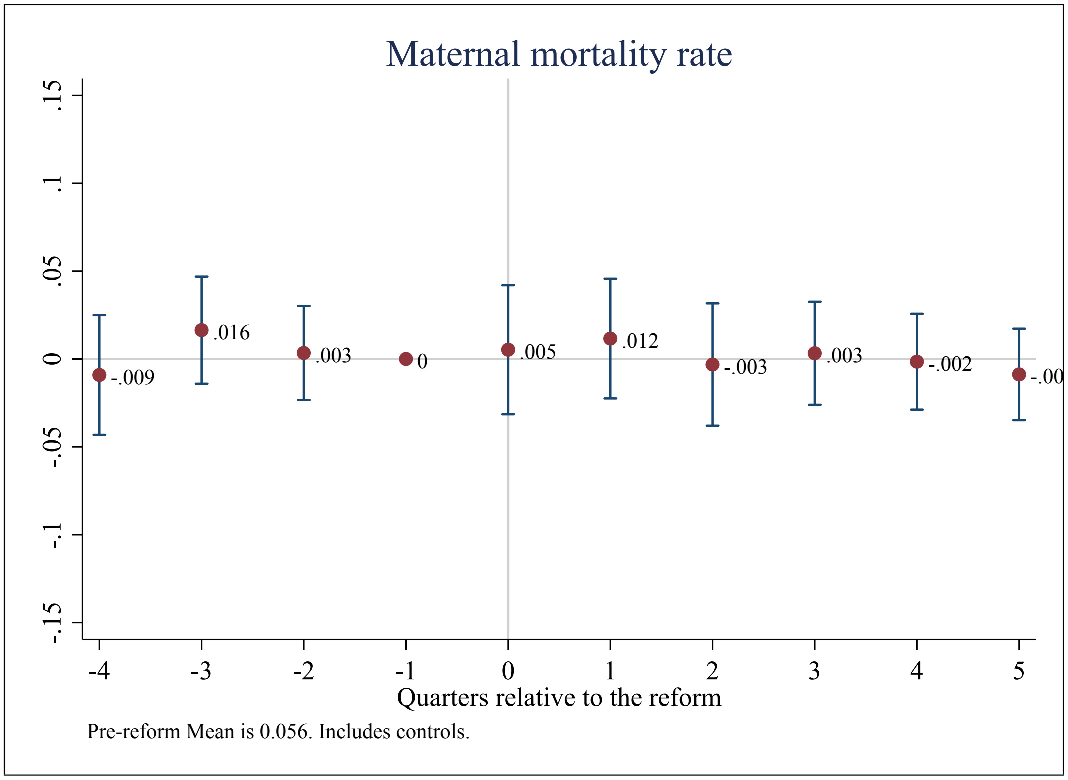


Notes: The sample is the balanced panel of municipalities and the period of analysis is four quarters before the reform and five quarters later in both countries. The points in each figure represent the γ_k coefficients from the difference-in-difference model in equation 4.1. The range plot with capped spikes present the 95 percent confidence intervals. The reference period is the second quarter of 2014. The specification includes municipality fixed effects. Controls include the proportion of black and indigenous population (the omitted category is mestizo), the share of women in different age groups and the average gross value added per capita in the municipality. Municipality linear time trends are not included.

there are no statistically significant differences between the two countries in the four quarters before the reform. The logarithm of the c-section rate starts to decrease significantly relative to Colombia's rate in the second quarter after the reform, which corresponds to the first quarter of 2015.

Figure 4.3 also shows the γ_k coefficients from the difference-in-difference regression in equation 4.1. The outcome of interest is the logarithm of the share of low birth-weight babies in the municipality. The figure shows small and not statistically

Figure 4.4: γ_k estimates from the difference-in-difference regression in equation 4.1 when the outcome is the maternal mortality rate



Notes: The sample is the balanced panel of municipalities and the period of analysis is four quarters before the reform and five quarters later in both countries. The points in each figure represent the γ_k coefficients from the difference-in-difference model in equation 4.1. The range plot with capped spikes present the 95 percent confidence intervals. The reference period is the second quarter of 2014. The specification includes municipality fixed effects. Controls include the proportion of black and indigenous population (the omitted category is mestizo), the share of women in different age groups and the average gross value added per capita in the municipality. Municipality linear time trends are not included.

significant differences between Ecuador and Colombia before and after the reform.

Finally, Figure 4.4 shows the dynamic effects of the reform on the logarithm of maternal mortality rates. It shows, like in the previous graphs small differences between Ecuador and Colombia before the reform which are not significantly different from zero. Furthermore, there is no evidence of treatment effects in any of the five quarters that followed the reform.

4.5.2 Checking for stability in the composition of the two countries

The parallel trends assumption is violated if the groups that are being compared are not similar or if their composition changes differentially across time. In Table 4.1, I showed that women in Colombia and Ecuador have a very similar demographic composition. Here I show that the demographic composition in both countries followed similar paths before and after the reform. Table 4.2 shows the results of the model in equation 4.1. The specifications in columns 1 to 3 include controls for the age composition and gross value added per capita in the municipality, while the specifications in columns 4 to 6 control for the ethnic composition and value added per capita at the municipality level. I added municipality specific time trends only in the regressions where the dependent variable is the share of mothers in the different age groups because those variables did not follow parallel trends before the reform. To show the complete pattern of coefficients before and after the reform date, the third quarter of 2014 was chosen as the reference period.

The joint F-tests for the leads in the models suggest that the share of black, indigenous and mestizo women followed parallel trends before and after the reform in both countries. However, the variables for the different age groups did not follow strictly a parallel trend before the reform. The deviations in the different age groups are small though, always close to 1% or below 1%. What is interesting is that there was a statistically significant reduction in the share of Ecuadorian women aged 35 to 44 years that registered giving birth right after the reform which persisted several quarters later.

Given that c-sections are more common among this age group because these women are old enough to be having their second or third child and because the likelihood of a c-section increases with age, a fertility drop in this group may explain part of the observed reduction in c-sections after the reform. To test that the results are not

Table 4.2: Stability in the demographic composition of Ecuador and Colombia

	(1) Black	(2) Indigenous	(3) Mestizo	(4) 15-24 years	(5) 25-34 years	(6) 35-44 years
Effect in t-1	0.00219	-0.000489	-0.00166	-0.0114***	0.00471*	0.00911***
	(0.00237)	(0.00107)	(0.00245)	(0.00289)	(0.00263)	(0.00236)
Effect in t-2	0.000603	-0.00107	-0.000851	-0.0125***	0.00266	0.0131***
	(0.00468)	(0.00182)	(0.00394)	(0.00370)	(0.00361)	(0.00264)
Effect in t-3	0.00193	-0.000585	-0.000952	-0.00346	0.00205	0.00203
	(0.00447)	(0.00188)	(0.00392)	(0.00301)	(0.00364)	(0.00377)
Effect in t-4	-0.000449	-0.00289	0.00114	0.00727*	-0.00610	-0.00128
	(0.00440)	(0.00225)	(0.00403)	(0.00384)	(0.00443)	(0.00452)
Effect in t+1	-0.00166	-0.000356	0.000876	0.0140***	-0.00794**	-0.00912***
	(0.00152)	(0.00127)	(0.00135)	(0.00304)	(0.00326)	(0.00175)
Effect in t+2	-0.00849*	-0.000715	0.00523	0.00809**	0.000231	-0.0107***
	(0.00441)	(0.00210)	(0.00334)	(0.00407)	(0.00373)	(0.00265)
Effect in t+3	-0.00277	-0.000498	0.00149	0.00545*	0.00241	-0.00999***
	(0.00433)	(0.00223)	(0.00339)	(0.00278)	(0.00255)	(0.00212)
Effect in t+4	-0.00465	-0.000560	0.00248	0.00838***	0.000839	-0.0122***
	(0.00367)	(0.00310)	(0.00350)	(0.00213)	(0.00249)	(0.00212)
Effect in t+5	-0.00366	0.00177	0.00145	0.00180	0.00545**	-0.00873***
	(0.00413)	(0.00233)	(0.00330)	(0.00292)	(0.00257)	(0.00235)
P-value (leads)	0.736	0.495	0.740	0.000	0.001	0.000
Average effect (five lags)	-0.004	-0.000	0.002	0.008	0.000	-0.010
P-value (Average effect)	0.130	0.968	0.344	0.001	0.928	0.000
Municip. and quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Municip. linear time trend	No	No	No	No	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Weights	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8865	8865	8865	8865	8865	8865

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The sample is women of fertile age in Vital Statistics data. Robust standard errors clustered at municipality level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of analysis is four quarters before the reform and five quarters later. The Table shows the results of the model in equation 4.1 and includes municipality and time fixed effects as well as controls for age and race composition at municipality level (depending on the dependent variable) and the logarithm of the gross value added per capita in the municipality. The reference period is the third quarter of 2014. Municipality specific time trends are included in the models in columns 4 to 6.

driven by the fertility drop among women aged 35 to 44 years, in the robustness checks section, I present results by age groups. It is reassuring though, that the effects found after excluding these women are almost identical to the ones observed for the complete sample, which will be discussed in the next section.

Table 4.3 presents additional evidence of parallel trends on other determinants of the c-section rate that have been discussed so far. The analysis is done separately for women ages 15 to 24 years, 25 to 34 years and 35 to 44 years. The joint F-tests for the leads in the models suggest that despite the differences in the social insurance systems in Ecuador and Colombia, the share of births that were assisted by a doctor during the four quarters prior to the reform followed a parallel trend in both countries. This is true also for the share of babies that were born at a hospital or at home across all age groups. Interestingly, the share of mothers aged 35 to 44 years with subsidized or contributory insurance also followed parallel trends in the two countries.

Table 4.3: Parallel trends tests on other characteristics by age group

Age Group	Assisted by a doctor (1)	Born at hospital (2)	Born at home (3)	Subsidized regime (4)	Contributory regime (5)
Mothers aged 15 to 24 years					
P-value (leads)	0.369	0.623	0.731	0.0390	0.0300
Observations	8099	8099	8099	8099	8099
Mothers aged 25 to 34 years					
P-value (leads)	0.201	0.601	0.200	0.0650	0.0430
Observations	7836	7836	7836	7836	7836
Mothers aged 35 to 44 years					
P-value (leads)	0.501	0.715	0.610	0.474	0.183
Observations	5814	5814	5814	5811	5811
Municip. and quarter FE	Yes	Yes	Yes	Yes	Yes
Municip. linear time trend	No	No	No	No	No
Controls	Yes	Yes	Yes	Yes	Yes
Weights	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors clustered at municipality level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample is mothers aged 15 to 44 years old. The period of analysis is four quarters before the reform and five quarters later. I report the results of my preferred specification for each of the outcomes of interest. The weights in the regressions reflect the number of births in the municipality in each age range as a share of total births in each age range. All the dependent variables are in logarithms. Controls include the proportion of black and indigenous population (the omitted category is mestizo), the share of women in different age groups and the average gross value added per capita in the municipality.

4.6 Results

In this section, I present the results on the effects of the malpractice reform using the non-parametric specification in equation 4.1. I study the effects of the reform on the logarithm of the c-section rate in the municipality, the logarithm of the share of low birth-weight babies and the logarithm of early maternal mortality rate also at the municipality level. I report on the effects of the malpractice reform up to five quarters after it's implementation (lags) and the cumulative average effect of the five lags in the model. Results are accompanied by an F-test of the joint significance of the leads in the model, to formally test for parallel trends and adds to the graphical evidence discussed in section 4.5.

4.6.1 C-section rate

Table 4.4 reports results of the effect of the malpractice reform on the logarithm of the c-section rate in the municipality. Column 1 reports the results of the difference-in-difference model in equation 4.1. Column 2 adds a municipality specific linear time trend to the model and column 3 adds time variant controls to the specification in column 2. All the models use weights that reflect the number of births in the municipality as a share of the number of births in the country.

In Column 1, the F-test of the joint significance of the leads indicates that the c-section rate followed parallel trends in both countries before the reform (p-value=0.972). For this reason, my preferred specification is the one reported in column 1, which corresponds to the estimates presented in Figure 4.2. The estimates in column 1 show that the reform had a negative effect on the c-section rate of Ecuadorian municipalities. All the lags in the model have a negative sign, yet the effect is statistically significant starting in the second quarter after the reform. It is worth noting that the negative effects are stronger three quarters after the passing of the new Penal Code (-1.4% significant at 1% level). The reason may be that

Table 4.4: Effects of the Malpractice Reform on c-section rate in Ecuador

	(1) C-section	(2) C-section	(3) C-section
Effect in t	-0.00142	-0.00159	-0.00172
	(0.00365)	(0.00368)	(0.00367)
Effect in t+1	-0.00153	-0.00213	-0.00215
	(0.00368)	(0.00342)	(0.00341)
Effect in t+2	-0.0109**	-0.0116***	-0.0119***
	(0.00445)	(0.00412)	(0.00409)
Effect in t+3	-0.0136***	-0.0146***	-0.0148***
	(0.00467)	(0.00417)	(0.00416)
Effect in t+4	-0.0114**	-0.0128***	-0.0130***
	(0.00543)	(0.00399)	(0.00397)
Effect in t+5	-0.0129**	-0.0146***	-0.0148***
	(0.00568)	(0.00364)	(0.00362)
P-value (leads)	0.972	0.960	0.961
Average effect (five lags)	-0.0101	-0.0111	-0.0113
P-value (test Average effect=0)	0.0122	0.0000815	0.0000566
Municip. and quarter FE	Yes	Yes	Yes
Municip. linear time trend	No	Yes	Yes
Controls	Yes	No	Yes
Weights	Yes	Yes	Yes
Observations	9487	9487	9487

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at municipality level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of analysis is four quarters before the reform and five quarters later. Column 1 reports the results of the difference-in-difference model in equation 4.1 using weights that reflect the number of births in the municipality as a share of the total in the country. Column 2 adds a municipality specific linear time trend to the model and column 3 adds some time variant controls related to race, the share of women in different age groups and the average gross value added per capita in the municipality.

women who got pregnant after the reform were exposed to it during the 9 months that followed, giving doctors the chance to do more to avoid an unnecessary planned c-section or complications leading to an emergency c-section. The estimates reported in columns 2 and 3 look almost identical to the ones reported in column 1.

4.6.2 Share of low birth-weight babies

Table 4.5 shows the effects of the reform on the logarithm of the share of low birth-weight babies at the municipal level. In line with figure 4.3, the F-test of the joint significance of the leads in the model shows that the share of low birth-weight babies followed parallel trends in the two countries before the reform (p-value=0.118). Hence, my preferred specification is in column 1. Results in column 1 show a negative but not statistically significant effect of the reform on the quarter after it's implementation.

The sign of the effect is in line with the results reported in Table 4.4, since a reduction in c-sections is associated with an increase in the birth-weight of newborns and hence with a reduction in the share of low birth-weight babies in t+1. The observed negative effects were not statistically significant in the specification that did not include municipality time trends and turned statistically significant in the specification that included municipality time trends reported in column 2. However, after controlling for time-variant controls (in column 3) like age and race composition in the municipality and the logarithm of value added per capita, the estimates of the time indicators that were closer to the reform date became not statistically significant.

4.6.3 Maternal mortality

Consistent with the pattern of γ coefficients shown in Figure 4.4, the p-value of the F-test reported in column 1 of Table 4.6 shows that early maternal mortality followed parallel trends in both countries in the pre-reform period (p-value=0.293).

The coefficients of the five lags included in the model are not statistically significant in any of the specifications presented in columns 1 to 3 and neither was the coefficient of the cumulative average effect in the five quarters following the reform date. The results reported in column 1 of Table 4.6 suggest that the reform had a small negative effect on the logarithm of early maternal mortality in the second, fourth

Table 4.5: Effects of the Malpractice Reform on low birth-weight in Ecuador

	(1) Low birth weight	(2) Low birth weight	(3) Low birth weight
Effect in t	0.000425	-0.00225	-0.000810
	(0.00202)	(0.00224)	(0.00240)
Effect in t+1	-0.000453	-0.00440**	-0.00296
	(0.00181)	(0.00205)	(0.00266)
Effect in t+2	0.00174	-0.00466	-0.00471
	(0.00378)	(0.00296)	(0.00298)
Effect in t+3	0.00469	-0.00395*	-0.00323
	(0.00347)	(0.00225)	(0.00236)
Effect in t+4	0.00331	-0.00737***	-0.00586**
	(0.00374)	(0.00186)	(0.00251)
Effect in t+5	0.000332	-0.0126***	-0.0103***
	(0.00415)	(0.00170)	(0.00306)
P-value (leads)	0.118	0.149	0.230
Average effect (five lags)	0.00192	-0.00659	-0.00541
P-value (test Average effect=0)	0.507	0.0000473	0.00858
Municip. and quarter FE	Yes	Yes	Yes
Municip. linear time trend	No	Yes	Yes
Controls	Yes	No	Yes
Weights	Yes	Yes	Yes
Observations	8894	9422	8894

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at municipality level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of analysis is four quarters before the reform and five quarters later. Column 1 reports the results of the difference-in-difference model in equation 4.1 using weights that reflect the number of births in the municipality as a share of the total in the country. Column 2 adds a municipality specific linear time trend to the model and column 3 adds some time variant controls related to race, the share of women in different age groups and the average gross value added per capita in the municipality.

and fifth quarter after the reform date, but none of the coefficients were statistically significant. Furthermore, once linear time trends were added to the model, the sign of the coefficients became positive but still not statistically significant.

Table 4.6: Effects of the Malpractice Reform on early maternal mortality in Ecuador

	(1) Maternal mortality	(2) Maternal mortality	(3) Maternal mortality
Effect in t	0.00526	0.0106	0.0155
	(0.0187)	(0.0211)	(0.0222)
Effect in t+1	0.0116	0.0145	0.0322
	(0.0174)	(0.0208)	(0.0240)
Effect in t+2	-0.00316	0.00269	0.000708
	(0.0177)	(0.0189)	(0.0184)
Effect in t+3	0.00326	0.00705	0.0167
	(0.0149)	(0.0130)	(0.0147)
Effect in t+4	-0.00153	0.000299	0.0231
	(0.0139)	(0.0154)	(0.0195)
Effect in t+5	-0.00879	-0.00548	0.0258
	(0.0133)	(0.00929)	(0.0203)
P-value (leads)	0.293	0.443	0.352
Average effect (five lags)	0.000	0.004	0.020
P-value (test Average effect=0)	0.983	0.769	0.214
Municip. and quarter FE	Yes	Yes	Yes
Municip. linear time trend	No	Yes	Yes
Controls	Yes	No	Yes
Weights	Yes	Yes	Yes
Observations	4183	4511	4183

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Robust standard errors clustered at municipality level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of analysis is four quarters before the reform and five quarters later. Column 1 reports the results of the difference-in-difference model in equation 4.1 using weights that reflect the number of early maternal deaths in the municipality as a share of the total deaths among women of fertile age in the country. Column 2 adds a municipality specific linear time trend to the model and column 3 adds some time variant controls related to race, the share of women in different age groups and the average gross value added per capita in the municipality.

4.6.4 Heterogeneous effects

The c-section rate is expected to be higher among older mothers than among younger mothers (Walker et al., 2016). In fact, the research in the area has found that the risks of perinatal death, hypertensive disease, diabetes mellitus, among others, are higher among women aged 35 years or older (Walker et al., 2016). Older women also face a higher risk for preterm labor and low birth-weight infants ($< 2500g$) (Reddy et al., 2006).

Taking this into consideration, Table 4.7 reports on the effects of the reform for three different age groups. None of the models include municipality time trends so that they are comparable to the preferred specifications discussed in the results section. In terms of the c-section rate, the effects found among mothers aged 15 to 34 years were very similar to those reported in Table 4.4 for the complete sample. However, the reform did not have a significant effect on mothers aged 35 to 44 years. It seems that doctors reduced the use of c-sections only among younger women, but kept the c-section rate unchanged among older mothers. This finding is in line with the theory that argues that doctors tend to have less discretion over riskier pregnancies.

Regarding the effects on the share of low-birth-weight babies, it is possible to see that the reform caused a statistically significant reduction in the share of low-birth-weight babies only among mothers aged 25 to 34 years, while for the other age groups the effect remained not statistically significant. Finally, regarding the effects on maternal mortality, the results by age groups were not different from the results for the complete sample in the sense that the effects remained insignificant across all the age groups, however, it is possible to see that the reform had a positive effect on maternal mortality among the younger group and a negative effect among the older group.

Table 4.7: Effects of the Malpractice Reform on the c-section rate among mothers aged 15 to 34 years and 35 to 44 years

	Mothers Aged 15-24 years			Mothers Aged 25-34 years			Mothers Aged 35-44 years		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	C-section	Low birth weight	Maternal mortality	C-section	Low birth weight	Maternal mortality	C-section	Low birth weight	Maternal mortality
Effect in t	-0.000579	0.000329	0.0496	-0.00383	0.000390	-0.00933	0.00765	0.00217	-0.00759
	(0.00471)	(0.00236)	(0.0374)	(0.00329)	(0.00273)	(0.0381)	(0.00628)	(0.00608)	(0.0141)
Effect in t+1	0.00235	0.00322	0.0442	0.00493	-0.00443**	0.00344	0.00314	-0.00151	-0.00262
	(0.00464)	(0.00232)	(0.0315)	(0.00430)	(0.00224)	(0.0306)	(0.00692)	(0.00539)	(0.0234)
Effect in t+2	-0.0119*	0.00307	0.0357	-0.0134**	0.00189	-0.0128	0.000633	-8.74e-05	-0.0246
	(0.00695)	(0.00387)	(0.0489)	(0.00646)	(0.00351)	(0.0323)	(0.00646)	(0.00820)	(0.0231)
Effect in t+3	-0.0171**	0.00528	0.0287	-0.0110*	0.00539	0.0154	0.00197	0.00616	-0.0260
	(0.00730)	(0.00355)	(0.0382)	(0.00603)	(0.00453)	(0.0285)	(0.00657)	(0.00823)	(0.0183)
Effect in t+4	-0.0150**	0.00182	0.0233	-0.0131*	0.00585	0.00151	0.00312	0.00997	-0.0185
	(0.00741)	(0.00437)	(0.0340)	(0.00684)	(0.00411)	(0.0347)	(0.00586)	(0.00840)	(0.0263)
Effect in t+5	-0.0159**	-0.00189	0.0309	-0.0124*	0.00338	-0.0113	0.00984	0.00408	-0.0315**
	(0.00716)	(0.00405)	(0.0452)	(0.00648)	(0.00492)	(0.0337)	(0.00658)	(0.00902)	(0.0146)
P-value (leads)	0.816	0.249	0.720	0.470	0.548	0.208	0.0310	0.750	0.667
Average effect (five lags)	-0.0110	0.00200	0.0330	-0.00900	0.00200	-0.00100	0.00400	0.00400	-0.0210
P-value (test Average effect=0)	0.0450	0.409	0.289	0.0700	0.445	0.975	0.381	0.580	0.232
Municip. and quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municip. linear time trend	No	No	No	No	No	No	Yes	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8728	8628	1998	8476	8359	2074	6396	6228	2509

Notes: Robust standard errors clustered at municipality level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample is mothers aged 15 to 44 years old. The period of analysis is four quarters before the reform and five quarters later. I report the results of my preferred specification for each of the outcomes of interest. Controls include the proportion of black and indigenous population (the omitted category is mestizo), the share of women in different age groups and the average gross value added per capita in the municipality.

4.7 Robustness checks

In this section, I test the robustness of the results to changes in the window around the reform date. More specifically, I check if the results hold when I expand the period of analysis to include eight quarters before the reform and five quarters later. I also investigate whether the results found so far change if I select Peru instead of Colombia as control group.

4.7.1 Sensitivity to changes in the window around the reform date

As a robustness check, I estimate the model in equation 4.1 considering the period between eight quarters before the reform and five quarters later. Table 4.8 reports the results of my preferred specification for each of the outcomes of interest. The joint F-test in Column 1 shows that eight quarters before the reform, the c-section rate in Ecuador followed a parallel trend with respect to Colombia's rate. In general, the results in Table 4.8 are similar to the ones presented in Table 4.4 and show a statistically significant reduction of 0.9% on the logarithm of the c-section rate in Ecuadorian municipalities in the five quarters that followed the passing of the new Penal Code (in average). The results reported in columns 2 and 3 about the impacts on the share of low birth-weight babies and maternal mortality are also in line with the results reported in Tables 4.5 and 4.6, respectively.

4.7.2 Using Peru as control group

In this section, I compare the results that were discussed in section 4.6 with the results of an alternative specification that uses Peru as the control group for Ecuador. I got access to Vital Statistics data for Peru through a special petition to the National Statistics Office, however, microdata on death certificates was not readily available. Hence, the analysis in this section focuses on the effects of the reform on the c-section rate and on the share of low-birth-weight babies at the municipality level.

Table 4.8: Effects of the Malpractice Reform. Period of analysis eight quarters before the reform and five quarters later

	(1) C-section	(2) Low birth weight	(3) Maternal mortality
Effect in t	-0.000954	-0.000311	0.00793
	(0.00349)	(0.00194)	(0.0195)
Effect in t+1	0.00349	-0.00227	0.00989
	(0.00363)	(0.00185)	(0.0191)
Effect in t+2	-0.0126**	-0.00181	-0.00256
	(0.00640)	(0.00264)	(0.0177)
Effect in t+3	-0.0137**	0.000532	0.000688
	(0.00629)	(0.00217)	(0.0124)
Effect in t+4	-0.0132**	-0.00205	-0.00775
	(0.00642)	(0.00240)	(0.0144)
Effect in t+5	-0.0100	-0.00390*	-0.0155
	(0.00699)	(0.00227)	(0.00991)
P-value (leads)	0.101	0.0830	0.375
Average effect (five lags)	-0.00900	-0.00200	-0.00300
P-value (test Average effect=0)	0.0880	0.268	0.807
Municip. and quarter FE	Yes	Yes	Yes
Municip. linear time trend	No	Yes	Yes
Controls	Yes	No	No
Weights	Yes	Yes	Yes
Observations	12621	13165	6344

Notes: Robust standard errors clustered at municipality level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The period of analysis is eight quarters before the reform and five quarters later. Column 1 reports the results from the model in equation 4.1 while the specifications in columns 2 and 3 include municipality specific linear time trends. Controls include the proportion of black and indigenous population (the omitted category is mestizo), the share of women in different age groups and the average gross value added per capita in the municipality.

Columns 1 and 4 show the results reported in section 4.6 for comparison purposes while columns 2 and 5 show the results of the specification in equation 4.1 using Peru as the control group. The results found when Peru was used as the control group for Ecuador are in line with the results found so far (shown in column 1). Column 2 reports a negative effect of the reform on the c-section rate that starts in the reform quarter and turns statistically significant in the second quarter after the reform. The reduction in c-sections was slightly higher in the fourth and fifth quarter possibly because Peru experienced a slight increase in the c-section rate by the end of 2015 as evidenced in the results presented in column 3 that compares

Colombia and Peru.[18]

The p-value of the joint F-test of the leads in the model reported in column 3 shows that Colombia and Peru followed similar trends in the c-section rate before the reform. This was expected considering that both countries followed parallel trends with respect to the c-section rate of Ecuador. Since none of these countries experienced a reform, I expected that they followed parallel trends in the c-section rate also after the third quarter of 2014. However, Peru experienced a slight increase in the c-section rate in the two last quarters of 2015. This can be due to the implementation of “Plan Bienvenidos a la Vida" that was launched in April 2015 after the passing of the Resolution 997-2014/MINSA.[19] The aim of “Plan Bienvenidos a la Vida" was to provide mothers with a newborn kit, counseling regarding newborn care and to register babies in the Health Insurance System (SIS). More importantly, the program targets mothers in the first and second poverty quintiles who gave birth at a health facility. Since giving birth at a health facility increases the likelihood of having a c-section, this may explain the rise in c-sections in Peru observed in the last quarters of 2015.

Columns 4 to 6 show the results for the share of low birth-weight babies. The three countries followed parallel trends before the reform as shown by the p-value of the F-test of the leads in the models (p-value=0.118, 0.475, 0.563, respectively). In line with the results obtained when Colombia was the control group (shown in column 4), when Peru was used as control group, Ecuador also experienced a decrease in the share of low birth-weight babies on the first quarter after the malpractice reform (as shown in column 5), even though the difference was not statistically significant.

[18]The rise in the c-section rate of Peru is likely to make the difference between Ecuador’s and Peru’s c-section rates larger, making also the reduction in the c-section rate of Ecuador look larger.

[19]For more details see https://es.slideshare.net/FresiaIsmelda/plan-bienvenidos-a-la-vida

Table 4.9: Results using Perú as control group

	C-section rate			Low birth-weight		
	(1) Ecuador vs Colombia	(2) Ecuador vs Perú	(3) Perú vs Colombia	(4) Ecuador vs Colombia	(5) Ecuador vs Perú	(6) Perú vs Colombia
Effect in t	-0.00142	-0.00612	0.00513	0.000425	-0.00172	0.00134
	(0.00365)	(0.00372)	(0.00316)	(0.00202)	(0.00200)	(0.00122)
Effect in t+1	-0.00153	-0.00204	0.00575	-0.000453	-0.00266	0.00199
	(0.00368)	(0.00419)	(0.00436)	(0.00181)	(0.00183)	(0.00142)
Effect in t+2	-0.0109**	-0.0131***	0.00537	0.00174	0.000821	0.000959
	(0.00445)	(0.00494)	(0.00454)	(0.00378)	(0.00271)	(0.00140)
Effect in t+3	-0.0136***	-0.0153***	0.00579	0.00469	0.00226	0.00236
	(0.00467)	(0.00453)	(0.00421)	(0.00347)	(0.00215)	(0.00168)
Effect in t+4	-0.0114**	-0.0219***	0.0126**	0.00331	0.000614	0.00243
	(0.00543)	(0.00512)	(0.00530)	(0.00374)	(0.00241)	(0.00152)
Effect in t+5	-0.0129**	-0.0274***	0.0176***	0.000332	-0.00243	0.00240
	(0.00568)	(0.00545)	(0.00556)	(0.00415)	(0.00286)	(0.00158)
P-value (leads)	0.972	0.503	0.146	0.118	0.475	0.563
Average effect (five lags)	-0.0101	-0.0160	0.00942	0.00192	-0.000279	0.00203
P-value (Average effect)	0.0122	3.23e-05	0.0276	0.507	0.878	0.0879
Municip. and quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Municip. linear time trend	No	No	No	No	No	No
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Weights	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9487	6867	12540	8894	6811	12525

Notes: Robust standard errors clustered at municipality level shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The sample is mothers aged 15 to 44 years old. The period of analysis is four quarters before the reform and five quarters later. I report the results of the specification in equation 4.1 for each of the outcomes of interest. For the specifications where Peru is used as control group, the control variables included in the models are the share of women in the different age groups at the municipality level, since race composition variables were not available neither measures of value added per capita at the district level.

4.8 Conclusions

Health systems that provide health coverage at no cost for the greatest part of the population, necessarily face high health care costs because neither patients nor doctors internalize the costs of health services. Malpractice laws may reduce these costs but can also encourage defensive medicine so, the final result is an empirical question. Costs may drop if, in response to increased malpractice pressure, doctors reduce the use of invasive and expensive procedures in low-risk patients who do not need them. However, costs may increase if in response to the reform doctors begin to perform more preventive exams that do not have a positive effect on health outcomes, in other words, costs may increase if the reform induces defensive medicine.

This paper shows that policies that increase the malpractice pressure faced by doctors and health care providers can be effective in reducing the overprovision of elective procedures. In line with previous literature, I show that doctors respond to an increase in malpractice pressure by reducing the c-section rate. The observed reduction in the number of c-sections did not affect the health outcomes of mothers or newborns, instead, the share of low birth-weight babies decreased as expected (driven by the reduction in c-sections). Furthermore, early maternal mortality remained at the same pre-reform rates.

During the five quarters following the reform, doctors reduced the c-section rate by 1.1% among women aged 15 to 24 years, and by 0.9% among women aged 25 to 34 years. The c-section rate remained unaffected for women aged 35 to 44 years, possibly because these women have more risky pregnancies and doctors are expected to have less discretion over high-risk births (Currie and MacLeod, 2008). These results are consistent with excessive procedure use among women aged 15 to 34 years before the reform, meaning that doctors were performing too many unnecessary procedures. The theory predicts that when the marginal procedures are unnecessary, the probability of a medical error leading to liability is higher with surgery than without surgery (Currie and MacLeod, 2008) which explains why the malpractice reform led to a decrease in procedure use in Ecuador.

Given that the reduction in c-sections did not affect the health outcomes of mothers or newborns, we can conclude that the difference between the pre-reform and post-reform c-section levels correspond to socially wasteful procedures. Back of the envelope calculations allow estimating the aggregate savings caused by the reform. Considering that the number of registered newborns in Ecuador in the second quarter of 2014 (the reference period in all the analyses) was 63,671 and that the average cost of a c-section in 2014 in the public sector was 892.64 US dollars,[20] reducing the c-section rate in average by 1% in the five quarters following the reform implies aggregate savings in the order of 1,420,882.04 US dollars (0.5 x 0.01 x 63,671 x 892.64 x 5).

A better assessment of the overall benefits of increased malpractice pressure would also factor in the costs related to defensive medicine by estimating the costs of low-benefit treatments triggered by fear of prosecution. Further checks should use doctor-level data to check whether doctors decided to leave the least profitable sector (possibly the public sector) in response to the malpractice reform, leading to specialization within sectors.

[20] According to the information provided by the National Director of Articulation of the Public and Complementary Health System in document MSP-DNARPCS-2017-0295-O.

References

Adhvaryu, Achyuta (2014). “Learning, misallocation, and technology adoption: Evidence from new malaria therapy in Tanzania”. In: *Review of Economic Studies* 81.4, pp. 1331–1365.

Asociación Nacional de Clínicas y Hospitales Privados del Ecuador (2013). *Informe sobre responsabilidad Penal médica en Ecuador*. Tech. rep.

Baicker, Katherine, Kasey Buckles, and Amitabh Chandra (2006). “Geographic variation in the appropriate use of cesarean delivery”. In: *Health Affairs* 25.5, pp. 355–367.

Baldwin, Laura-Mae, Gary Hart, Meredith Fordyce, Roger Rosenblatt, and Michael Lloyd (1995). “Defensive medicine and obstetrics”. In: *The Journal of the American Medical Association* 274, pp. 1606–1610.

Banerjee, Abhijit V, Paul J Gertler, and Maitreesh Ghatak (2002). “Empowerment and efficiency: Tenancy reform in West Bengal”. In: *Journal of Political Economy* 110.2, pp. 239–280.

Blundell, Richard and Thomas Macurdy (1999). *Chapter 27 Labor supply: A review of alternative approaches*. Vol. 3 PART. 1. Elsevier Masson SAS, pp. 1559–1695.

Cohen, Jessica, Pascaline Dupas, and Simone Schaner (2015). “Price subsidies, diagnostic tests, and targeting of malaria treatment : Evidence from a randomized controlled trial”. In: *The American Economic Review* 105.2, pp. 609–645.

Currie, Janet, Jonathan Gruber, and Michael Fischer (1995). “Physician payments and infant mortality: Evidence from Medicaid fee policy”. In: *The American Economic Review* 85.2, pp. 106–111.

Currie, Janet and W. Bentley MacLeod (2008). “First do no harm? Tort reform and birth outcomes”. In: *Quarterly Journal of Economics* 123.2, pp. 795–830.

Currie, Janet and W. Bentley MacLeod (2017). “Diagnosing expertise: Human capital, decision making, and performance among physicians”. In: *Journal of Labor Economics* 35.1, pp. 1–43.

Dubay, Lisa, Robert Kaestner, and Timothy Waidmann (1999). “The impact of malpractice fears on cesarean section rates”. In: *Journal of Health Economics* 18.4, pp. 491–522.

– (2001). “Medical malpractice liability and its effect on prenatal care utilization and infant health”. In: *Journal of Health Economics* 20.4, pp. 591–611.

Escobar-Córdoba, Franklin (2012). “Medical liability of the psychiatrist”. In: *Revista Colombiana de Anestesiología* 40.1, pp. 17–20.

Frakes, Michael (2013). “Impact of Medical Liability Standards on Regional Variations in Physicians Behavior : Evidence from Adoption of National-Standard Rules”. In: *American Economic Review* 103.1, pp. 257–276.

Giedion, Ursula and Manuela Villar (2009). “Colombia’ s universal health insurance system”. In: *Health Affairs* 28.3, pp. 853–863.

Giovanella, Ligia, Oscar Feo, Mariana Faria, and Sebastián Tobar (2012). *Health Systems in South America : Challenges to the universality, integrality and equity*. Tech. rep. May, pp. 1–835.

Gruber, Jon, John Kim, and Dina Mayzlin (1999). “Physician fees and procedure intensity: The case of cesarean delivery”. In: *Journal of Health Economics* 18.4, pp. 473–490.

Gruber, Jonathan and Maria Owings (1996). “Physician financial incentives and cesarean section delivery”. In: *The RAND Journal of Economics* 27.1, pp. 99–123.

Kerschbamer, Rudolf and Matthias Sutter (2017). “The economics of credence goods - A survey of recent lab and field experiments”. In: *CESifo Economic Studies* 63.1, pp. 1–23.

Kessler, D. and M. McClellan (1996). “Do doctors practice defensive medicine?” In: *The Quarterly Journal of Economics* 111.2, pp. 353–390.

Kessler, Daniel and Mark McClellan (2002). “Malpractice law and health care reform: Optimal liability policy in an era of managed care”. In: *Journal of Public Economics* 84.2, pp. 175–197.

Lindo, Jason M. and Analisa Packham (2017). “How much can expanding access to long-acting reversible contraceptives reduce teen birth rates?” In: *American Economic Journal: Economic Policy* 9.3, pp. 348–376.

Localio, A. Russell, Ann G. Lawthers, Joan M. Bengtson, Liesi E. Hebert, Susan L. Weaver, Troyen A. Brennan, and J. Richard Landis (1993). “Relationship between claims and cesarean malpractice delivery”. In: *The Journal of the American Medical Association* 269.3, pp. 366–373.

McGuire, Thomas G. (2000). “Physician agency”. In: *Handbook of Health Economics*. Ed. by A. J. Culyer and J. P. Newhouse. Vol. 1. Chap. 9, pp. 461–536.

Ministerio de Salud Pública del Ecuador (2014). *Tarifario de prestaciones para el sistema nacional de salud*.

Naranjo, Mariana (2014). *Social protection systems in Latin America and the Caribbean: Ecuador*. Tech. rep. ECLAC.

Ortiz-Prado, Esteban, Tamara Acosta Castillo, Mauricio Olmedo-López, Luciana Armijos, Darío Ramírez, and Ana Lucia Iturralde (2017). “Cesarean section rates in Ecuador: a 13-year comparative analysis between public and private health systems”. In: *Pan American Journal of Public Health* 41, pp. 1–17.

Reddy, Uma M., Chia Wen Ko, and Marian Willinger (2006). "Maternal age and the risk of stillbirth throughout pregnancy in the United States". In: *American Journal of Obstetrics and Gynecology* 195.3, pp. 764–770.

Schulkind, Lisa and Teny Maghakian Shapiro (2014). "What a difference a day makes: Quantifying the effects of birth timing manipulation on infant health". In: *Journal of Health Economics* 33.1, pp. 139–158.

Shurtz, Ity (2014). "Malpractice law, physicians' financial incentives, and medical treatment: How do they interact?" In: *The Journal of Law and Economics* 57.1, pp. 1–29.

Sloan, Frank A., Stephen S. Entman, Bridget Reilly, Cheryl Glass, Gerald Hickson, and Harold Zhang (1997). "Tort liability and obstetricians' care levels". In: *International Review of Law and Economics* 17, pp. 245–260.

Tamara, Liliana, Sofía Jaramillo, and Luis Muñoz (2012). "Forensic expert report on alleged medical liability in Bogotá". In: *Revista Colombiana de Anestesiología* 39.4, pp. 489–505.

Tussing, A. Dale and Martha A. Wojtowycz (1992). "The cesarean decision in New York State, 1986. Economic and noneconomic aspects". In: *Medical Care* 30.6, pp. 529–40.

Walker, Kate F., George J. Bugg, Marion Macpherson, Carol McCormick, Nicky Grace, Chris Wildsmith, Lucy Bradshaw, Gordon C.S. Smith, and James G. Thornton (2016). "Randomized trial of labor induction in women 35 years of age or older". In: *New England Journal of Medicine* 374.9, pp. 813–822.

Chapter 5

CONCLUSIONS

Market-driven economic growth is likely to be the main driver of poverty reduction in most countries, however, markets alone rarely ensure efficient and equitable outcomes. Public policy plays a central role in providing the institutional foundations within which markets operate, in providing public goods, and in correcting market failures (Fiszbein et al., 2009). In the recent years, the focus of social protection systems in Latin America has been on social assistance programs like non-contributory pensions and cash transfer programs (Barrientos, 2011). The success of these programs has made countries aware of the need to put in place complementary actions to achieve a greater impact. Some of them include expanding access to public services and exploring reforms to social assistance programs focused on improving the targeting of beneficiaries as well as the timing and size of the benefits (Fiszbein et al., 2009).

The objective of this thesis has been to analyze the effectiveness of government interventions in the form of cash transfers to the poor and health care regulations. Considering the high costs involved in maintaining cash transfer programs in the long-run, this thesis contributes with new evidence about the persistence of effects in the long-run. This thesis also analyzes the role of government regulation in the health care sector, a topic that has not been sufficiently explored in the developing world. It provides first-hand evidence on the effects of labor market regulations on hospital quality in a developing country context and adds to the literature that looks at possible solutions to mitigate the problem of rising health care costs in developing countries.

Another important contribution of this thesis is the method proposed in Chapter 2 to evaluate the persistence of the effects of cash transfers in the long-run, which is relevant to other Latin American countries and more broadly to other developing countries that have programs with similar characteristics to BDH, in particular, programs that use Social Registry data and a poverty index to select beneficiaries.

The interventions studied in this thesis took place at different points in time in the same country. Hence, each identification strategy was chosen carefully to prevent the effects of a certain policy from contaminating the effects of another policy. For instance, the labor reform that began in 2009 and increased the number of doctor-hours in public hospitals may have had a positive effect on the health status of BDH beneficiary children, which in turn, may have boosted their education outcomes during the second phase of expansion of the BDH. However, given that soon after its implementation in 2003, it became clear that BDH was an unconditional transfer, I expect that children in treatment and control households would benefit from improved health attention in the same way. Hence, any difference in education and labor market outcomes should reflect only the effects of BDH.

Furthermore, I do not expect that the labor reform that affected health professionals at public health facilities would contaminate the effects of the malpractice reform on the c-section rate, maternal mortality, and birth-weight. First, because an increase in staffing levels would imply higher procedure use instead of the observed reduction in the c-section rate. Second, given that mortality is inversely related to staffing levels, the labor reform would reinforce a reduction on maternal mortality, yet the results were not statistically significant probably because very few hospitals (18% of the total) adopted the labor reform in 2014. Furthermore, there was enough time between the two reforms so that any change in the outcomes due to the first reform would be clearly distinguishable from the changes caused by the second reform, precisely around the implementation dates. To dispel any remaining doubts, as a robustness check, I estimated the models excluding hospitals that adopted the labor reform after 2013 and confirmed that the results were robust to the exclusion of those hospitals (not shown).

5.0.1 Summary of the Findings

The findings in Chapter 2, suggest that BDH became inefficient in raising children's schooling after several years targeting the same children. Most of the gains in enrollment and schooling were achieved in the short-run among children that began treatment when they were about to start elementary school, high school or Baccalaureate. However, an extended exposure to BDH was not enough to keep raising children's education.

The lack of differential effects on education and labor market outcomes among children that were treated during the two phases versus those treated during just the first phase is not explained by an attenuation of BDH effects for all the age groups in phase two. In fact, the analysis of the short-term effects at the end of phase two showed positive and important effects in years of education at juncture ages. This contradicts the hypotheses that educational gratuity or lack of monitoring of the conditions attenuated the effects of BDH during phase two among all age groups. A more plausible reason for the lack of differential effects is that once children reach the education level they have planned to achieve, or alternatively, once they reach a certain age when the opportunity cost of schooling becomes sufficiently high, an unconditional transfer does not provide enough incentive to keep them at school.

The only group that experienced positive differential effects in years of education after being exposed during the two phases of the program were children aged 0 when BDH began in 2003. Children who began treatment at an older age did not experience similar gains despite being treated with similar intensity possibly because once cognitive gaps appear, the process cannot be reversed with CTs. It is likely that older children were already lagging behind perhaps with low grades and attendance problems, which explains why BDH did not have positive effects on their education. These results are in line with the notion that human capital investment exhibits both self-productivity and complementarity and explains why late investments that attempt to compensate for the lack of adequate early investments can be very costly and ineffective (Cunha, Heckman, Lochner, et al., 2006; Carneiro et al., 2002; Cunha, Heckman, and Schennach, 2010; Heckman, 2000).

Regarding labor market outcomes in the short-run, results were not conclusive about whether the negative effect on the likelihood of having a job among treated 17-year-olds at the end of phase two was caused by a concurrent raise in enrollment for this age group. Moreover, being exposed to BDH for two phases versus just one did not give treated children an advantage in the labor market by the end of phase two. The reason could be that children exposed during the two phases of the program did not achieve more years of education after all.

Chapter 3 contributes to the literature that looks at the effects of changes in staffing levels on hospital quality and patient outcomes, by analyzing for the first time a reform that progressively made all health professionals working part-time switch to full-time contracts. The reform was successful in reducing in-hospital mortality, especially within the first 48 hours of admission. This is important because the literature on the distribution of time to death shows that most deaths in the emergency department, occur within the first 24 to 48 hours (Fisher and Gardner, 2012; Herlitz et al., 2010; Karoshi and Keith, 2009; Latense, 2009; Sircar et al., 2007; Demetriades et al., 2004).

Since a second objective of the reform was to guarantee continuous attention in emergencies and hospitalization, more emphasis was placed on increasing the availability of physicians in the emergency and surgery departments, which explains that the biggest drop in mortality was observed in the sample of admissions to the emergency department and in the indicator of mortality within the first 48 hours of admission.

Chapter 4, provides evidence that policies that increase the malpractice pressure facing doctors can be effective in reducing the overprovision of elective procedures, which would, in turn, reduce aggregate health care costs. Cost reduction is desirable if a reduction in procedure use does not affect patient's health meaning that doctors reduce the use of invasive and expensive procedures in low-risk patients who do not need them but not on high-risk patients that require them. The results show that this was the case in Ecuador. Doctors reduced the c-section rate among women aged 15 to 34 years but not among women aged 35 to 44 years who are expected to have riskier pregnancies.

Furthermore, the reduction in c-sections did not affect the health outcomes of mothers or newborns, so it is possible to conclude that the difference between the pre-reform and post-reform c-section levels correspond to socially wasteful procedures. Consequently, the reform caused important savings to the government and consumers and bigger savings would have been expected had we considered more types of elective procedures.

Limitations

This thesis has some limitations. In Chapter 2, limitations are related to attrition in the panel of individuals in the Registro Social. Even though most of the individuals reported a valid ID in the second and third waves, the sample in the three-wave panel was bounded by the number of individuals that reported a valid ID in the first wave. If children that experienced higher gains in the outcomes of interest are the ones that I am not able to follow in the panel, the results would underestimate the true effects of the program.

I took several steps to correct for this. First, I constructed sample weights so that the panel's totals on key characteristics[1] match the totals of the corresponding wave of the Registro Social. Second, to validate the results obtained with the data of Registro Social, I used administrative data from the ENES exam for the years 2013 and 2014 to estimate the impact of different lengths of exposure to BDH on the likelihood of taking the exam, which is a good proxy for high school graduation, and found similar results to the ones obtained using the three-waves panel of Registro Social.

In Chapter 3, the main limitation for the study of the effects of the labor reform on in-hospital mortality has to do with the lack of information about the official implementation schedule. As an alternative, I adopted a data-driven approach based on the assumption that treatment started in the first year a reduction in the number of part-time contracts above 35% was observed. This drop in part-time contracts had to occur within the adoption period determined by the resolutions of the Ministry of Labor for the different groups of public hospitals.

I plan to use the dates in the official implementation schedule as soon as I get access to it, but I do not expect to see significant changes in the results of the study, considering that I got access to the resolutions of the Ministry of Labor that guided in a more general way the order of adoption in the different groups of public hospitals. Every hospital had to obey the dates established in those resolutions, so I can be confident that any big drop in the number of part-time contracts within the adoption period set by the resolutions has to signal the implementation of the reform. Moreover, since the time unit for the event studies in Chapter 3 is the year, there is no need to have the exact adoption date for each hospital, just the year.

In Chapter 4, one limitation is that the Ecuadorian National Statistics Office began collecting information on the type of birth (c-section or vaginal delivery) in 2011.

[1]These variables include gender, education level, highest grade completed, the number of years of education, birth year, employment, province, county and type of house

Because of this, the c-section rate for Ecuador has an artificially steep slope until mid-2012 when more mothers started reporting that information. I had to restrict the sample for the main analysis to births that occurred four quarters before the third quarter of 2014 (reform date) and five quarters after that date. However, as a robustness check, I expanded the period of analysis to include eight quarters before the reform and five quarters later. The fact that I do not have a longer series does not affect my identification strategy though, because there are enough quarters before and after the reform to conduct the parallel trends analysis and to evaluate the impacts of the malpractice reform.

Policy Implications and Recommendations

The results from Chapter 2 stress the need for a redesign of BDH. BDH became an unconditional transfer in the first place partly due to the lack of individual-level information that could be accessed in a centralized and simple way to monitor the conditions. However, since 2007, there is a Master File of Educational Institutions which collects data of students from public and private institutions at the beginning and at the end of the school year. Additionally, since 2014, the Ministry of Health collects information at the patient-level about medical visits at outpatient health centers. With this information, in theory, it would be possible to effectively monitor compliance with BDH's conditions.

Given that age at treatment matters and considering that BDH was effective in the short-run raising schooling among children that were close to begin elementary school as well as the end of basic education. To ensure more children remain at school long enough to approach the end of high school, the government should make the transfer conditional on grade progression. By making BDH a conditional transfer, every additional year in school is rewarded, not only the last years of basic education and baccalaureate that have a higher return in the labor market. Furthermore, the size of the transfer should increase for older children to compensate for the wages they could get in the labor market if they choose to work.

To help treated children enter the labor market, first BDH should become effective in raising the years of education of children that remain in the program for long periods of time. However, even this might not be enough considering my results showing that BDH had important short-term effects in years of education and still did not help young adults to enter the labor market. It may be that these young adults lack the skills to find and keep a job. One way to remedy early skill deficits is to make compensatory investments, which may be an expensive option. A more efficient solution could be to offer guidance to motivate them to pursue tasks that do not require the skills that deprived early environments do not produce. Finally, the government could also create a program similar to the General Educational Development (GED) program in the US that allows high school dropouts to get a high school degree, which would improve their chances in their labor market (Cunha, Heckman, Lochner, et al., 2006).

The results from Chapter 3 show that a reform that increases the number of doctor-hours at hospitals with the intention of guaranteeing continuous attention at the emergency and surgical departments can be effective in reducing in-hospital mortality, in particular within the first 48 hours of admission. However, is responsibility of the government to consider all the implications of a reform of this kind, for instance, the existence of any negative spillover effects to the private sector, since it is likely that doctors had to leave their secondary jobs in the private sector to meet the regulations in the public sector or simply because one sector becomes more profitable than the other once part-time and on-call contracts are banned (Cheng et al., 2018). Moreover, doctors working long hours can suffer from fatigue and this could also affect their performance (Landrigan et al., 2004; Collewet and Sauermann, 2017; Pencavel, 2014).

Finally, the results in Chapter 4 show that countries with high health care costs could implement malpractice reforms to tackle the overprovision of elective procedures. Two sufficient conditions have been identified in the literature for the efficient provision of credence goods like health care services, namely liability and verifiability (Kerschbamer and Sutter, 2017). In practice liability is easier to enforce than verifiability from a policy perspective. It follows that malpractice and tort reforms have been proposed and welcomed by patients in many countries. Another way of reducing the problem of over-provision and overcharging is to choose an appropriate price structure (Kerschbamer and Sutter, 2017). However, in the public sector services are subsidized, leaving marginal cost pricing as an instrument to ensure efficient outcomes only in private markets. Hence, another alternative to regulate the public provision of health services is to develop tailor-made malpractice legislation that corrects the problems stemming from asymmetric information without encouraging the practice of defensive medicine. This could involve tougher liability laws on health providers and less stringent malpractice laws for doctors.

More generally, a persistent issue throughout the thesis has to do with data quality. Good data are the first step towards increasing transparency and accountability. In fact, some administrative data like civil registration and vital statistics are by themselves a crucial tool to fight poverty and inequality and should receive special attention. For example, there is evidence that argues that children who do not hold a birth certificate are three percent less likely to be enrolled in school in Bolivia and 18 percent less likely to be enrolled in Brazil (Peters, 2016).

Currently, in Ecuador, vital statistics and death certificates are processed by the Civil Registry, from birth and death forms that are distributed by the National Statistics Office (INEC, 2018). This creates delays in the publication of micro-data and explains why in a specific year the microdata that is published by the National Statistics Office contains children that were born in other years but registered in the current year. By contrast, in Colombia, individuals have access to a web application that allows them to register births and deaths in real time. Other countries in the region have implemented similar systems. In Uruguay, for example, around 35 percent of death certificates are electronic (Peters, 2016).[2]

Regarding healthcare data, patient discharge data in Ecuador do not record information about patients after they leave the hospital. Hence, it is not possible to construct several risk-adjusted mortality rates and 30 days mortality indicators. These data do not distinguish between urgent admissions and emergency admissions either. The latter is information that hospitals have or could generate, but to the moment, the National Statistics Office – the institution that processes the information of all hospitals in the country – does not request this information from hospitals. Regarding data on hospital re-admissions, it would be best that the National Statistics Office generate these data by matching inpatient records over time using patients IDs, so that is possible to identify re-admissions even if the patient is admitted to a different hospital.

Finally, a common problem in Latin American countries is that no major efforts have been made to homogenize the information recorded in household surveys, vital statistics, hospital resources, and hospital activity data. Furthermore, very little information is available to the public, which makes it difficult to carry out investigations. Additionally, these countries should make greater efforts to generate longitudinal databases, which allow following the observation units over time.

Future research

Related to Chapter 2, future research could use the panel that links the three waves of social registry data and merge it with individual-level administrative data that the Ministry of Public Health collects from children below age five who attend regular

[2] However, an important aspect to consider is the trade-off between the goal of full digitization and the more important goal of complete coverage (Peters, 2016).

health checkups. These data contain the date of birth of the children, anthropometric measurements including weight, height and hemoglobin levels, as well as the ID of the mother or the child when available. Since the government began collecting these data in 2014, it is possible to estimate the short-term effects of BDH among children that were exposed only during the second phase of the program. Furthermore, one can use the original assignment to treatment in 2003 to study the long-term effects of BDH, by comparing the health outcomes of children who were close to the eligibility threshold set in 2003 ten years later.

Using the information on other outcomes that are available in the social registry panel, future research could look at the effects of a long exposure to BDH on the labor supply of women and how it evolved across time, as well as the effects of BDH on the marriage and divorce rates among beneficiary women. Other topics to be explored include the effects of Credito de Desarrollo Humano, a program that provides small nonrefundable loans to BDH recipients, on the likelihood of becoming self-employed and on consumption of durable goods.

To study the long-term effects of BDH on the likelihood of starting a new business, in theory, it is possible to match the three waves of social registry data to the datasets collected by the Superintendence of Companies using the names of the individuals. The latter contains information about the firms where individuals participate as shareholders and financial statements of the firms. Many of these firms are small and medium enterprises, and some of them were registered but do not generate profits. Hence, it is possible to analyze the effect of BDH and Credito de Desarrollo Humano in business creation and possibly in their growth.

Regarding Chapter 3, I have requested individual-level data that allows identifying the health professionals who switched to full-time contracts as well as their transitions to the public or private sector. These data would allow me to study the cost-effectiveness of the reform since I would be able to distinguish the health professionals who switched to full-time contracts from the new full-time employees hired in a year, which is not possible using hospital resources data alone, since the latter report the total number of full-time contracts per hospital in a year. Furthermore, I would be able to observe the wage increase of doctors that were affected by the reform.

These data would also allow me to study the spillover effects of the labor reform on private hospitals' performance. If it is true that specialists respond to changes in earnings by reallocating working hours to the sector with relatively higher wages, while leaving total working hours unchanged as recent evidence suggest (Cheng et al., 2018), the reform should have an effect on mortality at private hospitals, even though the direction of the effect is not completely clear because it ultimately depends on the number and characteristics of the doctors that decide to leave or remain in the private sector. Given that dual practice is widespread in many developed and developing countries and considering that it can have important implications for health care cost and quality (Cheng et al., 2018), this is a very relevant topic which has received little attention in the literature (McPake et al., 2016).[3]

Finally, linked to the analysis in Chapter 4, future research could study whether doctors facing increased malpractice pressure began to administer treatments with no apparent medical benefits by fear of prosecution – a practice known as defensive medicine – (Kessler and McClellan, 1996; Paik et al., 2017; Baldwin et al., 1995). This would provide a more complete assessment of the effectiveness of liability rules on aggregate health care costs reduction. I recently got access to administrative data for Ecuador that contains information that allows me to explore doctors' treatment patterns over time since 2013. These data contain the ID of the doctor and patients as

[3] Dual practice is defined as a situation where a doctor combines clinical practice in the public and private sector.

well as the diagnosis and treatment that each patient receives every time she registers a visit to an ambulatory health facility of the Ministry of Public Health. With these data, it would be possible to construct estimates of doctor productivity based on the number of patients treated in a year and identify changes in productivity and prescription patterns around the reform date using a model that controls for doctor and time fixed effects.

BIBLIOGRAPHY

Amarante, Verónica and Martín Brun (2016). "Cash transfers in Latin America effects on poverty and redistribution". In: *WIDER Working Paper* 136.

Angrist, Joshua D. and Alan B. Krueger (1999). "Empirical strategies in labor economics". In: *Handbook of Labor Economics*. Ed. by O. Ashenfelter and D. Card. Vol. 3. Chap. 23, pp. 1277–1366.

Baicker, Katherine, Kasey Buckles, and Amitabh Chandra (2006). "Geographic variation in the appropriate use of cesarean delivery". In: *Health Affairs* 25.5, pp. 355–367.

Baldwin, Laura-Mae, Gary Hart, Meredith Fordyce, Roger Rosenblatt, and Michael Lloyd (1995). "Defensive medicine and obstetrics". In: *The Journal of the American Medical Association* 274, pp. 1606–1610.

Barrientos, Armando (2011). "Social protection and poverty". In: *International Journal of Social Welfare* 20.3, pp. 240–249.

Behrman, Jere R., Susan W. Parker, and Petra E. Todd (2011). "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits?: A Five-Year Followup of PROGRESA/Oportunidades". In: *Journal of Human Resources* 46.1, pp. 203–236.

Behrman, Jere R., Piyali Sengupta, and Petra Todd (2005). "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico". In: *Economic Development and Cultural Change* 54.1, pp. 237–275.

Broadway, Barbara, Guyonne Kalb, Jinhu Li, and Anthony Scott (2017). "Do Financial Incentives Influence GPs' Decisions to Do After-hours Work? A Discrete

Choice Labour Supply Model". In: *Health Economics (United Kingdom)* 26.12, e52–e66.

Carneiro, Pedro, James J Heckman, David Bravo, Partha Dasgupta, Steve Levitt, Lance Lochner, Costas Meghir, Kathleen Mullen, and Casey Mulligan (2002). "The Evidence on Credit Constraints in Post-Secondary Schooling". In: *The Economic Journal* 112, pp. 705–734.

Cheng, Terence C., Guyonne Kalb, and Anthony Scott (2018). "Public, private or both? Analyzing factors influencing the labour supply of medical specialists". In: *Canadian Journal of Economics* 51.2, pp. 660–692.

Cohen, Jessica, Pascaline Dupas, and Simone Schaner (2015). "Price subsidies, diagnostic tests, and targeting of malaria treatment : Evidence from a randomized controlled trial". In: *The American Economic Review* 105.2, pp. 609–645.

Collewet, Marion and Jan Sauermann (2017). "Working hours and productivity". In: *Labour Economics* 47, pp. 96–106.

Cotlear, Daniel, Octavio Gómez-Dantés, Felicia Knaul, Rifat Atun, Ivana C.H.C. Barreto, Oscar Cetrángolo, Marcos Cueto, Pedro Francke, Patricia Frenz, Ramiro Guerrero, Rafael Lozano, Robert Marten, and Rocío Sáenz (2015). "Overcoming social segregation in health care in Latin America". In: *The Lancet* 385.9974, pp. 1248–1259.

Cruces, Guillermo, Carolina García, and Gasparini Leonardo (2014). *Inequality in Education: Evidence for Latin America*, pp. 318–339. ISBN: 9780199682676. DOI: 10.1093/acprof. arXiv: arXiv:1011.1669v3.

Cunha, Flavio, James J Heckman, and Susanne M Schennach (2010). "Estimating the Technology of Cognitive and Noncognitive Skill Formation". In: *Econometrica* 78.3, pp. 883–931.

Cunha, Flavio, James Heckman, Lance Lochner, and Dimitriy Masterov (2006). “Chapter 12. Interpreting the Evidence on Life Cycle Skill Formation”. In: *Handbook of the Economics of Education*. Vol. 1, pp. 697–812.

Currie, Janet and W. Bentley MacLeod (2008). “First do no harm? Tort reform and birth outcomes”. In: *Quarterly Journal of Economics* 123.2, pp. 795–830.

Demetriades, D, J Murray, K Chralambides, K Alo, G Velmahos, P Rhee, and L Chan (2004). “Trauma fatalities: time and location of trauma deaths”. In: *Journal of the American College of Surgeons* 198.1, pp. 20–26.

Dolton, Peter and Vikram Pathania (2016). “Can increased primary care access reduce demand for emergency care? Evidence from England’s 7-day GP opening”. In: *Journal of Health Economics* 49, pp. 193–208.

Dubay, Lisa, Robert Kaestner, and Timothy Waidmann (1999). “The impact of malpractice fears on cesarean section rates”. In: *Journal of Health Economics* 18.4, pp. 491–522.

– (2001). “Medical malpractice liability and its effect on prenatal care utilization and infant health”. In: *Journal of Health Economics* 20.4, pp. 591–611.

Edmonds, Eric V. and Norbert Schady (2012). “Poverty alleviation and child labor”. In: *American Economic Journal: Economic Policy* 4.4, pp. 100–124.

Fisher, Jessica M and Timothy B Gardner (2012). “The “Golden Hours” of Management in Acute Pancreatitis”. In: *The American Journal of Gastroenterology* 107.8, pp. 1152–1156.

Fiszbein, Ariel, Norbert Schady, Francisco Ferreira, Margaret Grosh, Nial Kelleher, Pedro Olinto, and Emmanuel Skoufias (2009). “Conditional Cash Transfers. Reducing present and future poverty”. In: *The World Bank*, pp. 1–361.

Frakes, Michael (2013). “Impact of Medical Liability Standards on Regional Variations in Physicians Behavior : Evidence from Adoption of National-Standard Rules”. In: *American Economic Review* 103.1, pp. 257–276.

Giovanella, Ligia, Oscar Feo, Mariana Faria, and Sebastián Tobar (2012). *Health Systems in South America : Challenges to the universality, integrality and equity*. Tech. rep. May, pp. 1–835.

Guzman, Eghon, Jack Ludmir, and Mark Defrancesco (2015). “High cesarean section rates in Latin America, a reflection of a different approach to labor?” In: *Open Journal of Obstetrics and Gynecology* 5.5, pp. 433–435.

Heckman, James J (2000). “Policies to foster human capital”. In: *Research in Economics* 54.December 1998, pp. 3–56.

Herlitz, Johan, Birgitta Wireklintsundström, Angela Bång, Annika Berglund, Leif Svensson, and Christian Blomstrand (2010). “Early identification and delay to treatment in myocardial infarction and stroke: differences and similarities.” In: *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 18, p. 48.

INEC (2018). *Registro estadístico de defunciones generales 2017. Estimación de la razón de mortalidad materna en el Ecuador*. Tech. rep.

Karoshi, Mahantesh and Louis Keith (2009). “Challenges in managing postpartum hemorrhage in resource-poor countries”. In: *Clinical Obstetrics and Gynecology* 52.2, pp. 285–298.

Kerschbamer, Rudolf and Matthias Sutter (2017). “The economics of credence goods - A survey of recent lab and field experiments”. In: *CESifo Economic Studies* 63.1, pp. 1–23.

Kessler, D. and M. McClellan (1996). "Do doctors practice defensive medicine?" In: *The Quarterly Journal of Economics* 111.2, pp. 353–390.

Kessler, Daniel and Mark McClellan (2002). "Malpractice law and health care reform: Optimal liability policy in an era of managed care". In: *Journal of Public Economics* 84.2, pp. 175–197.

Kugler, Adriana and Ingrid Rojas (2018). "Do CCTs improve employment and earnings in the very long-term? Evidence from Mexico". Cambridge, MA.

Landrigan, Christopher P., Jeffrey M. Rothschild, John W. Cronin, Rainu Kaushal, Elisabeth Burdick, Joel T. Katz, Craig M. Lilly, Peter H. Stone, Steven W. Lockley, David W. Bates, and Charles A. Czeisler (2004). "Effect of reducing interns' work hours on serious medical errors in intensive care units." In: *The New England Journal of Medicine* 351.18, pp. 1838–48.

Latense, Barbara A. (2009). "Critical Care of the Burn Patient: The First 48 Hours". In: *Critical Care Medicine* 37.10, pp. 2819–2826.

Leite, Phillippe, Tina George, Changqing Sun, Theresa Jones, and Kathy Lindert (2017). "Social registries for social assistance and beyond: A guidance note & assessment tool". In: *World Bank Social Protection and Labor Discussion Paper* 1704.

López-Cevallos, Daniel and Chunhuei Chi (2010). "Health care utilization in Ecuador: A multilevel analysis of socio-economic determinants and inequality issues". In: *Health Policy and Planning* 25.3, pp. 209–218.

Maeda, Akiko, Edson Araujo, Cheryl Cashin, Joseph Harris, Naoki Ikegami, and Michael Reich (2014). *Universal health coverage for inclusive and sustainable development. A synthesis of 11 country case studies*. Tech. rep. The World Bank.

McPake, Barbara, Giuliano Russo, David Hipgrave, Krishna Hort, and James Campbell (2016). “Implications of dual practice for universal health coverage”. In: *Bulletin of the World Health Organization* 94.2, pp. 142–146.

Naranjo, Mariana (2014). *Social protection systems in Latin America and the Caribbean: Ecuador*. Tech. rep. ECLAC.

OECD (2014). “Open government in Latin America”. In: *OECD Public Governance Reviews*.

– (2017). *Enhancing social inclusion in Latin America: Key issues and the role of social protection systems*. Tech. rep. OECD.

Paik, Myungho, Bernard Black, and David A. Hyman (2017). “Damage caps and defensive medicine, revisited”. In: *Journal of Health Economics* 51, pp. 84–97.

Pencavel, John (2014). “The productivity of working hours”. In: *Economic Journal* 125.589, pp. 2052–2076.

Peters, B. Guy (2016). “Civil registration and vital statistics as a tool to improve public management”. In: *IDB Discussion Paper* IDB-DP-473.

Pribble, Jennifer (2013). “Latin America’s left parties and the politics of poverty and inequality”. In: *Welfare and Party Politics in Latin America*. Chap. 8, pp. 172–183.

Rubalcaba, Luis, Stefka Slavova, Maria Kim, Fernando Merino, Ernesto Franco Temple, and Jessica Victor (2017). *Republic of Ecuador. Improving firms’ innovation to foster productivity and diversification*. Tech. rep. The World Bank.

Schady, Norbert, Maria Araujo, Ximena Peña, and Luis López-Calva (2008). “Cash Transfers, Conditions, and School Enrollment in Ecuador”. In: *Economía* 8.2, pp. 43–77.

Sircar, Padmini, Darshan Godkar, Shmuel Mahgerefteh, Karinn Chambers, Selva Niranjan, and Robert Cucco (2007). "Morbidity and mortality among patients with hip fractures surgically repaired within and after 48 hours". In: *American Journal of Therapeutics* 14.6, pp. 508–513.

Stampini, Marco and Leopoldo Tornarolli (2012). "The growth of conditional cash transfers in Latin America and the Caribbean: Did they go too far?" In: *IDB Policy Brief* IDB-PB-185.

Swami, Megha, Hugh Gravelle, Anthony Scott, and Jenny Williams (2017). "Hours worked by general practitioners and waiting times for primary care". In: *Health Economics* 27.May, pp. 1–25.

Appendix A

PROBABILISTIC RECORD LINKAGE FOR SOCIAL REGISTRIES

Abstract

With the increasing use and availability of ‘big’ data, linking data from multiple sources is becoming more common in academic research. In the absence of individual identifiers, probabilistic record linkage is an important tool to match individuals across data sets. Understanding probabilistic record linkage is essential for conducting robust longitudinal studies and assessing any potential biases due to differential linkage that comes from higher linkage rates across different subgroups of the population. In this paper, I contrast the performance of probabilistic record linkage versus phonetic encoding when generating a longitudinal data set that links several waves of Social Registry data recorded in Spanish.[1], In general, the results favor the use of probabilistic record linkage with a bi-gram string comparator over the alternative. Results improve considerably after the introduction of an algorithm that allows tracking individuals that left or joined the household on the next waves. Finally, I check for differential link rates and propose the use of sample-balancing or raking to correct for it.

[1]This paper was conceived as an independent paper but at this stage, it has been included as an appendix. Its objective is to document the procedure used to generate the database that I use in Chapter 2.

A.1 Introduction and literature review

With the increasing use and availability of 'big' data and administrative data linking information from multiple sources is becoming more and more common in academic research. Record linkage is the process of bringing together information from multiple sources using a common identifier or several common variables as match keys. It has other applications including building longitudinal profiles, de-duplication of individual records within a single database of records and case re-identification in capture-recapture studies (Sayers et al., 2015). I study record linkage in the context of linking data from multiple sources. There are two types of record linkage methods: (i) deterministic and (ii) probabilistic. In deterministic record linkage, records are matched if all match keys agree or unmatched if at least one disagree. On the other hand, probabilistic record linkage involves the estimation of a similarity score and leaves the researcher to decide the minimum similarity score needed for a matched pair to be considered a valid match (Sayers et al., 2015). With a unique numeric identifier, deterministic record linkage becomes a trivial task but, when the identifier is a string variable like names, the task becomes a bit more cumbersome and probabilistic record linkage is recommended.

The literature on probabilistic record linkage goes back to Fellegi and Sunter (1969) who developed the formal theory of probabilistic record linkage. Christen (2012) provides a more updated review of the methods related to record linkage for string variables. The two main approaches for string matching are phonetic encoding and pattern matching. Many techniques have been developed for both approaches and several techniques combine the two with the aim to improve the matching quality. Phonetic encoding techniques convert a name string into a code according to how a name is pronounced. Unfortunately, this process is language dependent and most techniques have been designed to work with anglicized names. The most commonly used phonetic encoding algorithms are Soundex, Phonex, Phonix, Double-Metaphone and Fuzzy Soundex. On the other hand, pattern matching techniques are used in approximate string matching with several applications from data linkage and duplicate deletion, information retrieval, correction of spelling errors to bio- and health informatics. The most used algorithms for pattern matching are Levenshtein or Edit distance, Damerau-Levenshtein distance, Bag distance, Smith-Watherman distance, Longest common sub-string, Q-grams, Positional q-grams, Skip-grams, Jaro and Jaro-Winkler algorithm (Christen, 2006).

Several measures have been proposed in the literature to assess the quality of alternative record linkage techniques and in general to evaluate the results of a data linkage process. Sensitivity is often referred to as the ‘matching rate’ and is defined as the proportion of true positives over the sum of the true positives and false negatives. Specificity is also known as the ‘true negative rate’ and is computed as the ratio of true negatives over the sum of true negatives and false positives. Finally, the positive predictive value is known as the ‘precision rate’ and is defined as the ratio of true positives divided by the sum of true positives and false positives (Harron, Goldstein, et al., 2015).

In this paper, I contrast the results obtained when using pattern matching and phonetic encoding independently and together to improve the quality of the data linkage of social registries recorded in Spanish. Social Registries are information systems that support registration and determination of potential eligibility for social programs (Leite et al., 2017). I use Ecuadorian administrative data from Registro Social (RS). The data in the RS is used to compute a poverty index called Selben which uses a principal components analysis that assigns numerical values to categorical variables. The Selben index is a proxy means test that takes values from 0 to 100 and is used for the allocation of most of the social programs in Ecuador, including cash transfers like Bono de Desarrollo Humano (BDH). To quantify the differences in terms of sensitivity and positive predictive value achieved with each technique, I first use a training subsample of the data that includes individuals who reported a valid ID in two consecutive waves of RS. The linkage is performed using blocking by household ID, which means that the program will look at matches only among the members of the same household in the following waves (Sadinle and Fienberg, 2013).

In general, the results favor the use of probabilistic record linkage with a bi-gram string comparator over the alternative which involves the use of phonetic encoding to perform the merge. The former mechanism produces a higher rate of successful linkages, but also a higher rate of false positive matches, with best results achieved after a thorough clerical review of the linked pairs. However, using phonetic encoding while conditioning the merge on the household identifier does not produce very substandard results either and the time needed to complete the merge is much lower compared to probabilistic record linkage using a bi-gram string comparator. Yet, the latter is preferred because it produces comparatively more successful matches which in turns reduces the sample bias that could arise by ignoring those extra matches even when it involves a more lengthy process. With longitudinal administrative data, one should always account for changes in household composition through time to successfully track the highest number of individuals. I propose an algorithm that allows me to track individuals that moved to other households or that joined the original households on the next wave. I show that the results of the preferred technique (probabilistic record linkage with a bi-gram string comparator) improves considerably after the introduction of the algorithm.

The contributions of this paper are twofold. First, is one of the few papers that document the best practices for conducting record linkage using longitudinal administrative data recorded in Spanish. Second, this study addresses the issues related to changes in household composition through time and is one of the few papers to introduce an application that lies outside the medical field using administrative data on social benefits. The rest of the paper is as follows. Section A.2 describes the data contained in the RS, Section A.3 presents an example that contrasts the performance of probabilistic record linkage versus a merge using phonetic encoding on a training subsample of the RS. Section A.4 builds upon the results of section A.3 and explains the process to build a panel that links the three waves of RS using the preferred method, ie. probabilistic record linkage. Section A.5 presents some descriptive statistics of the resulting 3-waves panel. Section A.6 checks for differential link rates and propose a technique called sample-balancing or raking to correct for it. Section A.7 concludes.

A.2 Data

The RS is considered as a census of the poor because by 2008 the total number of households in Ecuador was estimated in 3 392 851 and the second wave of the RS covers 2 393 377 of those households. Besides, the RS was conducted in 215 cantons of the 223 registered in the census of 2010(Ponce and Falconí, 2011). It contains relevant information of BDH recipients and potential recipients, namely, individual socio-economic information at the family and individual level, the ID number of the members (when available) and the Selben score assigned to each household.

To the moment, there are three waves, the first wave covers 6,303,352 individuals and was collected between 2001-2007, however, most of the information was collected before 2003 [2]. The second wave covers 8,068,957 individuals and was collected during 2007-2013 with most of the surveys completed in 2008. While the third wave was mostly collected in 2014 but data collection started in 2013 covering a total of 6,930,701 individuals. The reason why the last wave contains around 1,1 million

[2] Ecuador's population was 12,628,596 inhabitants by the year 2000, 14,447,600 inhabitants by 2008 and 15,661,312 inhabitants by 2013 according to the World Bank data (http://databank.worldbank.org/data/reports.aspx?source=2&series=SP.POP.TOTL&country=ECU)

fewer observations is that during the data collection period the last 287 counties were excluded from the process.

A.3 Comparison of alternative techniques

There are several ways to quantify the rate of linkage errors including: (i) comparison with a gold-standard sub-sample; (ii) sensitivity analysis; (iii) comparison of linked and unlinked data; and (iv) identification of implausible matches (Harron, Wade, et al., 2014). In this section, I use a training subsample of the data that includes individuals who reported a valid ID in two consecutive waves of RS (waves 2 and 3). With big administrative data sets like the RS, even this simplified example would result in 3,885,056 x 3,885,056 potential links. Blocking or stratification is used to split the database into smaller blocks or strata to restrict the number of comparisons to a subspace. All the matching keys have been previously preprocessed and I use blocking at the household level in all the cases, to improve the likelihood of success.

Table A.1 shows the performance of the two alternative record linkage techniques under consideration in terms of sensitivity and positive predictive value. It also shows the linkage rates achieved when the tracking algorithm is put in place. In general, using probabilistic record linkage renders the highest number of true matches, however, the results from the simple merge using phonetic encoding (Soundex) are not very different, particularly considering that the command was written to perform at its best in English. The biggest improvement in terms of sensitivity and positive predictive value happens with the introduction of an algorithm that allows tracking individuals that joined the household or moved to other households on the third wave. The sensitivity of the linkage in all the cases increases by nearly 9% when the algorithm is used, besides the positive predictive value improves slightly as well. These results confirm that using probabilistic record linkage with a bi-gram string comparator together with the tracking algorithm is the best strategy to achieve the

Table A.1: Performance for the alternative record linkage techniques

	Prob. rec. link. + bi-gram (A)	Simple merge + phonet. enc. (B)	A + tracking algorithm (C)	B + tracking algorithm (D)
True positives	3,500,654	3,473,139	3,848,617	3,835,092
False negatives	384,402	411,917	36,439	49,964
False positives	8,452	2	7,258	3
Sensitivity	0.901	0.894	0.991	0.987
Specificity	-	-	-	-
Pos. Pred. Val.	0.997	0.999	0.998	0.999
N	3,885,056	3,885,056	3,885,056	3,885,056

Notes: The table shows the sensitivity and positive predictive values for the alternative record linkage techniques. The sample is 3,885,056 individuals with valid IDs that were used to merge the second and third waves of RS. There are no results for specificity because the sample contains only true positives.

highest match rates. Building on these results, the next section documents the steps to link the three waves of RS and explains in more detail the tracking algorithm that allows me to track individuals that moved to other households or that joined the original households on the next wave.

A.4 Probabilistic record linkage

Building on the results of the previous section, this section documents the process to build a balanced panel that links the three available waves of the RS. The process uses probabilistic record linkage and takes into account the changes in household composition as times goes by to successfully track the highest number of individuals. The data linkage process involves three key steps: 1) preprocessing, 2) probabilistic record linkage, and 3) clerical review of machine-generated matched pairs.

A.4.1 Preprocessing

Data preprocessing consists of two steps: 1) parsing a field into the relevant subcomponents and 2) standardizing common character strings. Regarding the first step; in Ecuador and many Latin American countries, people report two names and two last names, so, I split the names into four new variables. The second step is particularly important when dealing with data recorded in Spanish, because of the common use accents and special characters that are not present in other languages. These should be replaced or dropped altogether to simplify the matching process that will use commands that are meant for data recorded in English. Cleaning the names and last names also involve removing all the common expressions that appear instead of the names.

A.4.2 Probabilistic record linkage

For the record linkage step, I selected the 2 names and 2 last names as match keys and used a common household ID as the blocking variable. This common household ID needs a bit more discussion because the correct construction of this variable determines the possibility to follow individuals who joined or left the original households in the next wave. To build it, I merged the first and second wave of the RS by individual ID and created a new variable by concatenating the household ID that was assigned to the individual in each of the two waves. I assigned this new household ID (hhold) to all the family members of the person with id. In that way, I just had to use names and last names to track the individuals inside those households (blocking). However, since individuals move to or from other households as times goes by, there were many household identifiers (hhold) at the family level in both waves, and I had to expand each wave creating duplicates and assigning to each one a different common household ID so that the record linkage command would look for the individuals in all their possible former households.

This can be explained with an example. Imagine that in wave 2 there is a household with two adults and both share the second part of the common household ID (hhold) but not the first part because they belonged to different households in wave 1. To track the rest of their family members (that did not report id) in wave 1, I had to check in the two former households. This is a simple example of a couple that gets married and possibly had children of their own before the marriage, but in the data there are cases where up to seven members of a household moved to other households and to track the dependent members, who did not have id, it was necessary to try all possible combinations of individuals and households.

I used the Stata command <reclink2> to perform the probabilistic record linkage and used the 2 names and 2 last names as matching keys. <reclink2> uses a bi-gram score which is computed from the ratio of the number of common two consecutive letters in the two strings and their average length minus one. The bigram score used in reclink (Blasnik, 2010) is a modified version where a pair of strings with up to four common prefix letters also gets extra credit (Wasi and Flaaen, 2015). An important difference between reclink and <reclink2> is that the latter allows for many to one matching, so I selected the ones with the higher match score.

To be sure of the results of the linking, I further compared the prefixes and suffixes of the matched names and last names as well as the sound of the prefixes. An additional restriction requiring that the difference between the age reported in wave1 and wave2 was not higher than 2 years was also included. The same procedure was used to match individuals in waves2 and 3. To get the complete panel, I performed a simple merge between the two linked datasets (waves 1-2 and waves 2-3) using an individual ID that I created since the beginning of the process for people in the second wave. Finally, to link individuals that were ignored in the process because nobody in their household had an id, I did a simple merge using the names, last names, gender, and birth year.

A.4.3 Clerical review and differential link rates

To assess the quality of the linking, I divided the linked pairs into smaller groups according to the similarity scores obtained from the probabilistic record linkage

Table A.2: Results of the clerical review process for the linkage of waves 1 and 2 of RS

Score	Not a match (A)	Maybe (B)	Very likely (C)	Definitely a match (D)	N
0.60-0.699	32	5	-	463	500
0.70-0.799	10	8	1	481	500
0.80-0.899	8	-	2	490	500
0.90-0.999	-	-	-	1250	1250
1	-	-	-	500	500
Total	50	13	3	3184	3250

Notes: The table shows the results of the clerical review process for the matched pairs obtained after probabilistic record linkage of waves 1 and 2 of RS. The process excludes the matched pairs that were obtained from deterministic record linkage using individual IDs.

Table A.3: Results of the clerical review process for the linkage of waves 2 and 3 of RS

Score	Not a match (A)	Maybe (B)	Very likely (C)	Definitely a match (D)	N
0.60-0.699	7	4	2	302	315
0.70-0.799	-	-	1	314	315
0.80-0.899	-	-	-	315	315
0.90-0.999	-	-	-	800	800
1	-	-	-	500	500
Total	7	4	3	2231	2245

Notes: The table shows the results of the clerical review process for the matched pairs obtained after probabilistic record linkage of waves 2 and 3 of RS. The process excludes the matched pairs that were obtained from deterministic record linkage using individual IDs.

step. I used the acceptance sampling approach and followed the ANSI AQL tables to choose the samples size and acceptance levels required to accept a group of randomly selected matched links or reject it. The linked pairs were manually checked using the Stata command <clrevmatch> (Wasi and Flaaen, 2015). Table A.2 shows the results of the match of waves 1 and 2. The results for the sensitivity measure surpasses 96% for all the groups, except for the first group which is the one that contains the matches with the lowest similarity score. However, even for this group sensitivity is around 92.6%. The linking is better between the second and third wave as shown in Table A.3 because starting with the second wave the MCDS put more emphasis on the collection and validation of peoples' birth dates and also more people reported IDs.

A.5 Results

I was able to build a balanced panel of 2 961 079 individuals across the three waves of RS. Tables A.4, A.5 and A.6 analyze the results in more detail. In the first wave of the RS, from a total of 6.3 million individuals 4,2 million (67%) belong to a household where at least one member has an ID number and due to that can be tracked into the second wave of the RS. Among those 4,2 million individuals 2 961 079 individuals have a valid match on the second and third wave of the RS, which correspond to 67% of the individuals who are likely to be matched and 47% of the total.

In terms of households, the link rate is higher and if we analyze the ratio between the number of households actually matched to the number of households that could potentially be matched it exceeds 100%. This is because some individuals leave their original households to form new ones and consequently the number of households that can be tracked through time increases if the individual who moves has an id.

In general, the link rates for individuals with ID in the three-waves panel exceed 67% in all the cases. However, there are at least other 1.26 million individuals that can be tracked through waves 1 and 2 but not through waves 1, 2 and 3; and other 1,48 million individuals that can be tracked only through waves 2 and 3. The reason is that in the first wave there are much fewer people reporting a valid ID so even when there are more matches between waves 2 and 3 (4,5 million individuals), the length of the panel is limited by the number of valid matches between waves 1 and 2.

Table A.4: Results for the first wave of the panel

	Total	Likely linked	%	Actually linked	%	Actual/Likely %
Households	1,583,617	1,022,164	65	1,068,188	67	105
Individuals	6,302,861	4,221,610	67	2,961,079	47	70

Table A.5: Results for the second wave of the panel

	Total	Likely linked	%	Actually linked	%	Actual/Likely %
Households	1,910,165	967,454	51	1,036,012	54	107
Individuals	8,068,957	4,447,300	55	2,961,079	37	67

Table A.6: Results for the third wave of the panel

	Total	Likely linked	%	Actually linked	%	Actual/Likely %
Households	1,758,401	984,356	56	1,179,668	67	120
Individuals	6,930,712	4,181,534	60	2,961,079	43	71

A.5.1 Some descriptive statistics

Before moving forward to the construction of weights, I present some descriptive statistics and compare the sample means in each wave of the panel with the population means in the Ecuadorian Census collected in 2010. This helps to make sense of the groups of the population that are more likely to be under-represented in the panel. It is important to remember that the first wave of the RS was collected between 2000 and 2007, which means that it covers most of the Ecuadorian population but its information is not directly comparable with that of other surveys which are collected on a yearly basis like for example the Ecuadorian Labor Force Survey, ENEMDU. Let me illustrate this with an example. A person who was born in 1989 is surveyed in the RS of 2003, she is 14 years and has achieved 9 years of schooling. This person is directly comparable to an individual surveyed in ENEMDU of 2003 who was born in 1989 and reports 9 years of schooling. However, this is not the case for an individual who was born in 1989 but was surveyed also in the first wave of the RS but in 2007. By 2007 this person may have cumulated 13 years of schooling or less if she had repeated a year or more.

Figures A.1 to A.3 show histograms describing the distribution of the population by year of birth for each wave of the RS versus the ones obtained using the 2010 census. For the sake of comparison, I focus on individuals born between 1900 and 2007, the latter is the last year of data collection of the first wave of the RS and also the year in which the youngest person in the panel could have been born. The RS covers a good share of the population in the census, this is why sometimes the RS is called a census of the poor. The left skewed curves are in line with the actual demographic pyramid in Ecuador. The panel covers an important share of the population of each wave, however, people born between 1985 and 1992 seem to be underrepresented in the three cases. These are children ages 11 to 18 years in 2003 and 21 to 28 years in 2013.

As I mentioned in more detail earlier, among the reasons why the panel is not able to cover a higher proportion of the people in wave 1 is the lack of IDs in some households who had to be left out of the probabilistic record linkage process. Besides, during the data collection of wave 3, at least 287 counties were excluded from the process. In the case of Ecuador, the conditions linked to the BDH could not be monitored due to lack of administrative capacity so most of the people who leave the program leave due to other reasons like migration, death or because they did not update their data. More importantly, there were two changes to the eligibility criteria, the first happened in 2009 when the second phase of the BDH program started and a new Selben score was computed. People with a Selben score higher

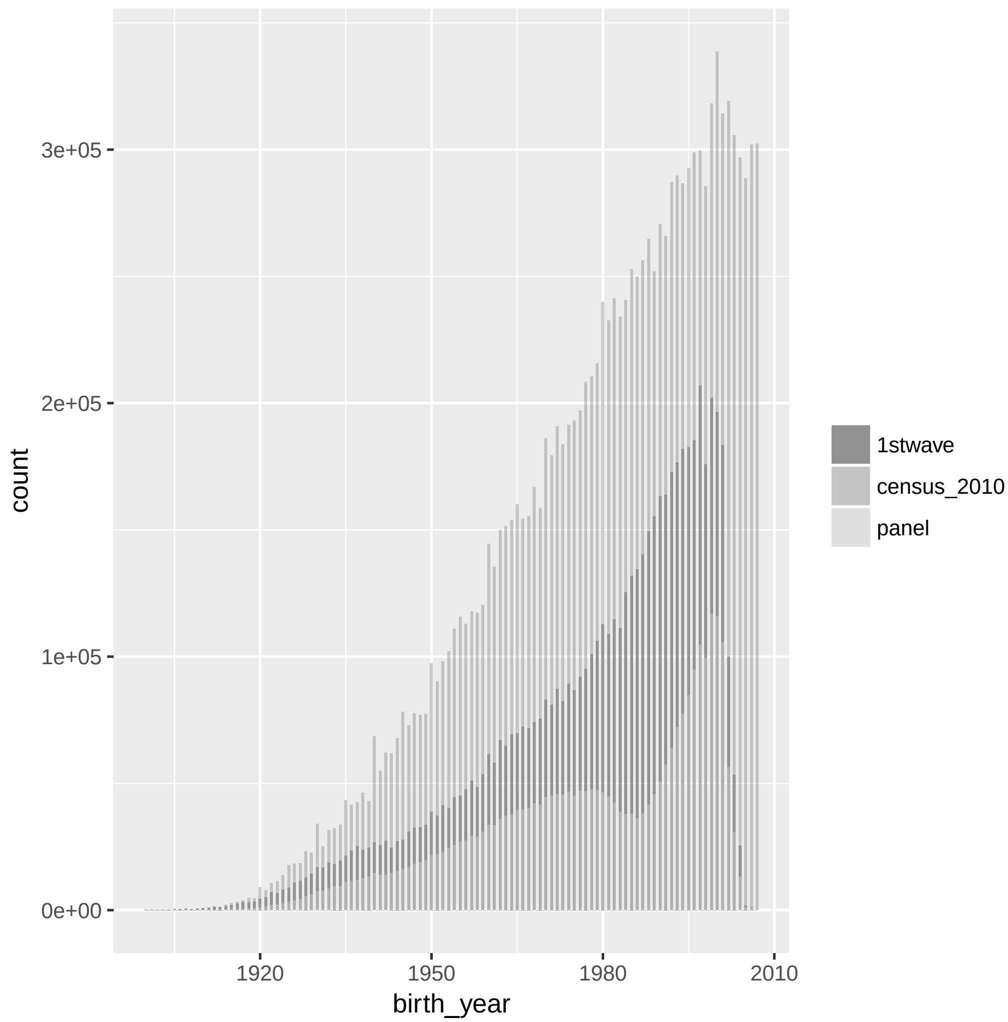


Figure A.1: Histogram of the distribution of the population by birth cohort in the census 2010, RS (first wave) and panel

than the new minimum score left the program and consequently left the RS. The second change was in 2013 when the government decided to reduce the pool of beneficiaries by moving the cutoff for selection; once again, these people left the RS. People that drop out for the latter reason can be identified in the administrative payment records of the BDH, while attrition due to migration can be accounted for by analyzing the variables in the RS related to the number of individuals who have migrated in the family or remittances.

There are around 1,9 million individuals who were in the first wave but left the RS

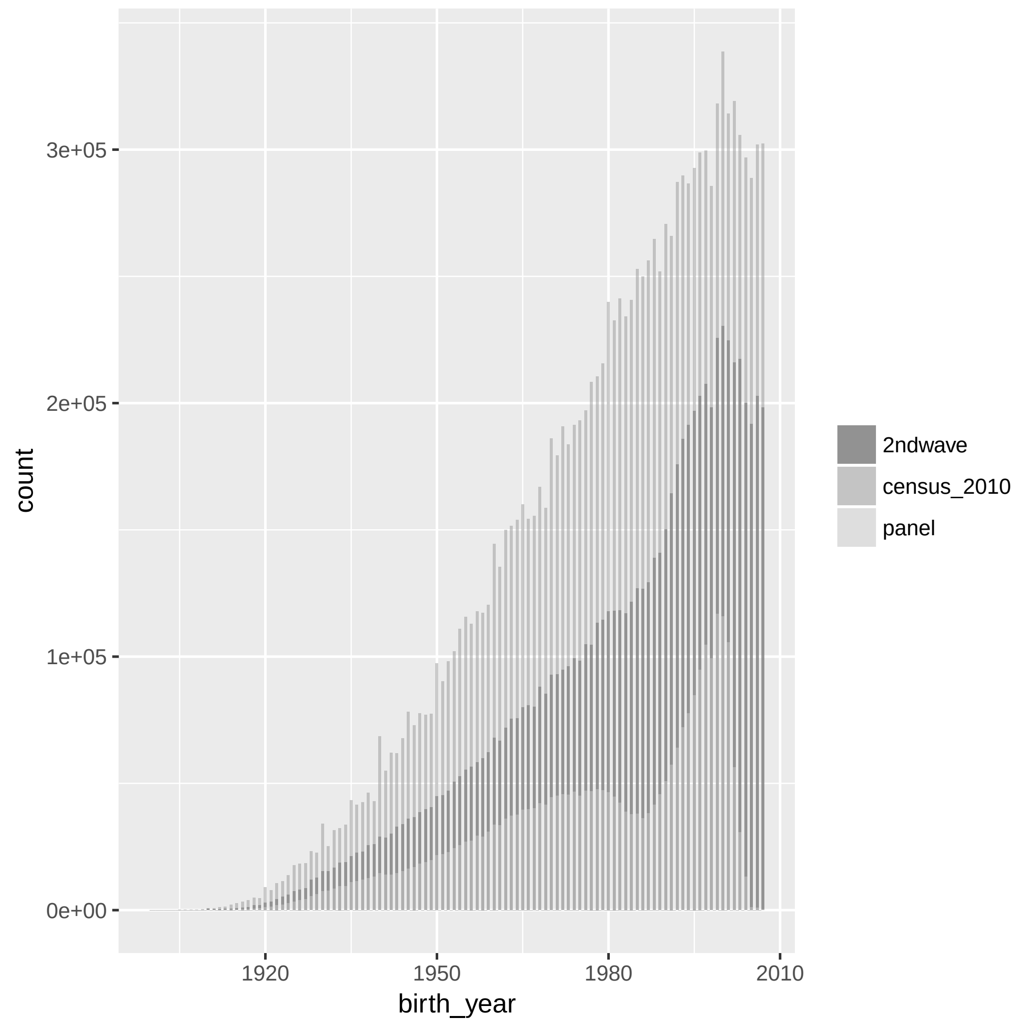


Figure A.2: Histogram of the distribution of the population by birth cohort in the census 2010, RS (second wave) and panel

after that. In general, these people tend to be richer as indicated by a higher Selben score when compared to the average score observed in the first wave. Also, it is more likely that men leave the RS. These results may be explained by the increasing efforts to refine targeting and coverage by the MCDS after the collection of the first wave, that led to dropping from the database an important number of individuals who were not poor enough.[3] The same happens with the people who were on the second wave of the RS but left the RS afterward. There are 2.6 million individuals

[3]The government performed several checks looking for people that appeared in the databases of the Social Security Institute or worked in the public sector, among others.

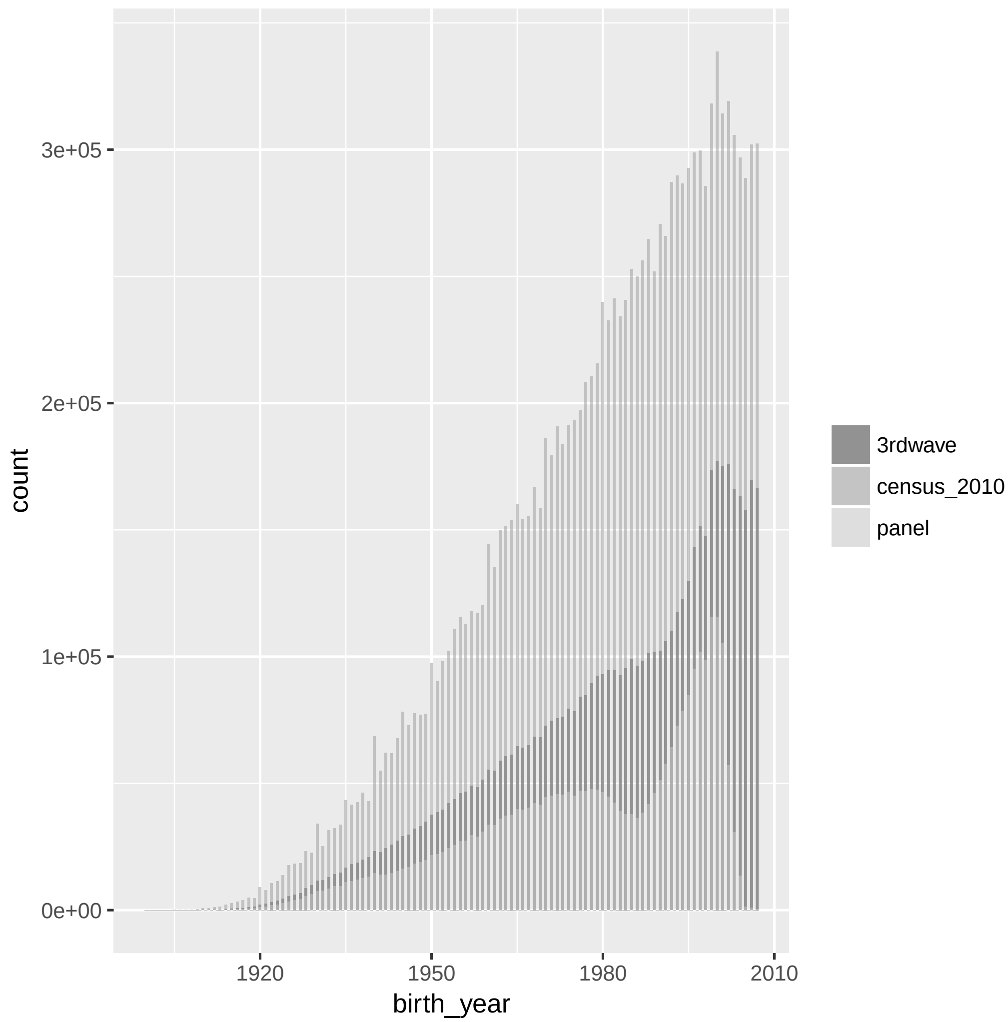


Figure A.3: Histogram of the distribution of the population by birth cohort in the census 2010, RS (third wave) and panel

that appear on the second wave of the RS but could not be observed on the third wave. These people are on average richer than the individuals who remain in the social registry data.

Table A.7: Number of individuals by age and gender in the first wave of the RS and the corresponding panel wave

Year	Panel Male	Panel Female	Total (A)	RS Male	RS Female	Total (B)	A/B %
1985	14628	23308	37936	65260	66675	131935	28.75
1986	15553	20728	36281	67802	66640	134442	26.99
1987	17910	20260	38170	70828	69469	140297	27.21
1988	20935	20722	41657	76315	73044	149359	27.89
1989	23579	22204	45783	79380	75987	155367	29.47
1990	26399	24480	50879	83736	79484	163220	31.17
1991	30019	27232	57251	83747	80090	163837	34.94
1992	33516	30450	63966	88036	84808	172844	37.01
1993	37498	34625	72123	89203	87496	176699	40.82
1994	40431	37223	77654	92431	89439	181870	42.7
1995	44339	40362	84701	92717	89797	182514	46.41
1996	50004	44875	94879	94212	90997	185209	51.23
1997	54982	49678	104660	105235	101684	206919	50.58
1998	51642	47657	99299	89668	86169	175837	56.47
1999	60312	56548	116860	103304	98731	202035	57.84
2000	59501	56423	115924	100393	96099	196492	59
2001	54030	51622	105652	93712	89701	183413	57.6
2002	28776	27674	56450	50996	48972	99968	56.47
2003	15625	15129	30754	27270	26183	53453	57.53

The panel covers over 40% of the children who were born between 1993 and 2003 in the first wave of the RS. The percentage is a bit lower for those born between 1985 and 1992. Almost the same happens with the second and third waves and this makes sense since by construction the panel follows the same people who entered the first wave in first place. Tables A.7 to A.9 show that the number of children born in 1985 is the same for all the waves (37 936) and the same happens with children born in the other years. However, the percentages with respect to the totals change because there are differences in the age distribution and the total number of people surveyed in each wave. For instance, the second wave is the one that surveyed the most people among the three, almost 1,7 million more than the first wave and 1 million more than the third wave.

Figures A.4, A.5 and A.6 show the kernel density curves for the SELBEN score

Table A.8: Number of individuals by age and gender in the second wave of the RS and the corresponding panel wave

Year	Panel Male	Panel Female	Total (A)	RS Male	RS Female	Total (B)	A/B %
1985	14550	23386	37936	58054	68856	126910	29.89
1986	15498	20783	36281	59288	67529	126817	28.61
1987	17905	20265	38170	60801	68521	129322	29.52
1988	20908	20749	41657	66635	72325	138960	29.98
1989	23528	22255	45783	68062	72887	140949	32.48
1990	26345	24534	50879	74493	75693	150186	33.88
1991	29950	27301	57251	82697	81725	164422	34.82
1992	33429	30537	63966	89164	86595	175759	36.39
1993	37458	34665	72123	94265	91579	185844	38.81
1994	40398	37256	77654	97403	93938	191341	40.58
1995	44231	40470	84701	99906	97089	196995	43
1996	49812	45067	94879	103120	99800	202920	46.76
1997	54797	49863	104660	105618	101938	207556	50.42
1998	51348	47951	99299	100748	97525	198273	50.08
1999	60175	56685	116860	114780	110857	225637	51.79
2000	59084	56840	115924	117213	113293	230506	50.29
2001	53504	52148	105652	113797	110907	224704	47.02
2002	28552	27898	56450	109407	106591	215998	26.13
2003	15428	15326	30754	110169	107205	217374	14.15

for each wave of the RS and the corresponding panel wave. In general, the curves follow a normal distribution and almost overlap, especially for the first and third wave. Is important to bear in mind that 30 variables were involved in the estimation of the Selben index, among them, characteristics of the household head, features of the house, access to services, assets, etc. This is the reason why it is considered a good measure to characterize households. The fact that the distributions of both the complete sample and the panel sample are very similar says that there is little sample selection and that on average the missing children belong to all the socio-economic backgrounds. Only in the second wave is quite clear that the panel contains slightly poorer households.

Table A.9: Number of individuals by age and gender in the third wave of the RS and the corresponding panel wave

Year	Panel Male	Panel Female	Total (A)	RS Male	RS Female	Total (B)	A/B %
1985	14534	23402	37936	44908	53929	98837	38.38
1986	15521	20760	36281	43984	52455	96439	37.62
1987	17931	20239	38170	45291	53144	98435	38.78
1988	20941	20716	41657	46946	54506	101452	41.06
1989	23540	22243	45783	47776	54025	101801	44.97
1990	26388	24491	50879	48652	53715	102367	49.7
1991	29983	27268	57251	50871	55088	105959	54.03
1992	33498	30468	63966	53904	56219	110123	58.09
1993	37517	34606	72123	58374	59304	117678	61.29
1994	40400	37254	77654	61592	60924	122516	63.38
1995	44313	40388	84701	65810	63893	129703	65.3
1996	49922	44957	94879	74033	69221	143254	66.23
1997	54878	49782	104660	78227	73116	151343	69.15
1998	51495	47804	99299	76113	71461	147574	67.29
1999	60269	56591	116860	88944	84489	173433	67.38
2000	59225	56699	115924	90572	86493	177065	65.47
2001	53655	51997	105652	88933	86150	175083	60.34
2002	28579	27871	56450	89746	86237	175983	32.08
2003	15493	15261	30754	84099	81955	166054	18.52

Figure A.4: Kernel density graphs of the Selben score in the first wave of RS and and the corresponding panel wave

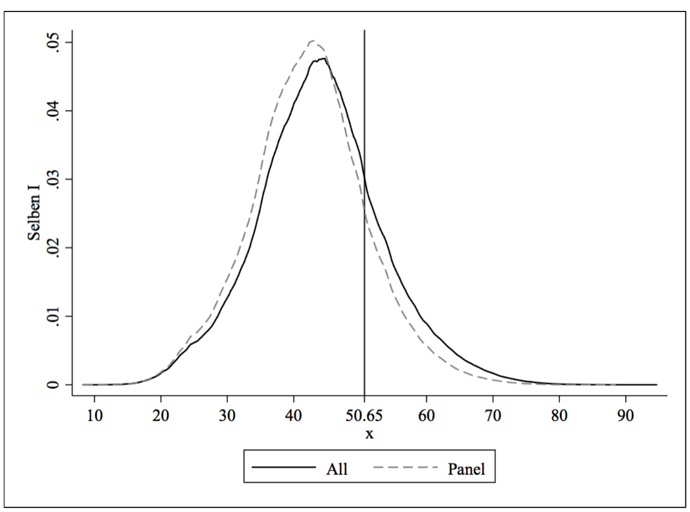


Figure A.5: Kernel density graphs of the Selben score in the second wave of RS and and the corresponding panel wave

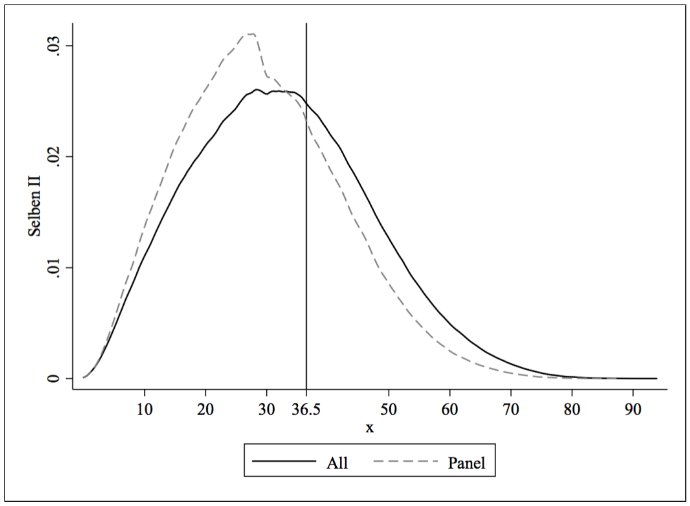


Figure A.6: Kernel density graphs of the Selben score in the third wave of RS and and the corresponding panel wave

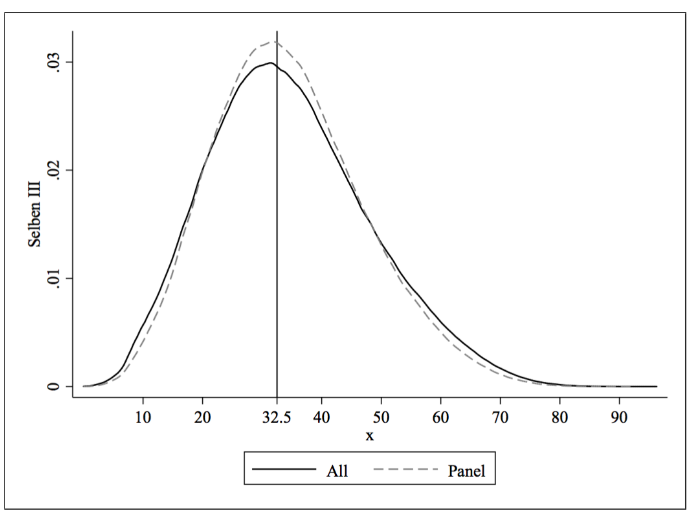


Table A.10: Number of years of education in the first wave of the RS and panel sample

Year	Panel	RS	Difference	t-test
1985	7.06	7.50	-0.44	0.00
1986	6.92	7.20	-0.28	0.00
1987	6.61	6.80	-0.19	0.00
1988	6.14	6.32	-0.18	0.00
1989	5.56	5.74	-0.18	0.00
1990	4.86	5.02	-0.16	0.00
1991	4.04	4.24	-0.20	0.00
1992	3.24	3.42	-0.18	0.00
1993	2.45	2.63	-0.18	0.00
1994	1.73	1.89	-0.16	0.00
1995	1.20	1.30	-0.11	0.00
1996	0.85	0.92	-0.07	0.00
1997	0.40	0.47	-0.06	0.00
1998	0.14	0.16	-0.02	0.00
1999	0.06	0.08	-0.02	0.00
2000	0.01	0.01	0.00	0.00
2001	0.00	0.00	0.00	0.00
2002	0.00	0.00	0.00	0.05
2003	0.00	0.00	0.00	0.40

Source: BDH Social Registry data.

Continuing with the descriptive statistics, Tables A.10 to A.12 show the average years of schooling by birth cohorts in each wave of the RS and its corresponding panel subsample. The differences are always below or equal to 0.45 years for children born between 1985 and 2003. The differences are in few cases not statistically significant and tend to be mostly negative for the first and second wave which indicates that on those waves the panel samples fail to identify and track some of the individuals with more years of schooling. In the third wave, the opposite is true meaning that people in the complete wave are on average less educated. This may be explained by the fact that in 2013, the government decided to graduate from the program to some of the beneficiaries with higher Selben scores.

All in all, the panel covers an important share of the population in each wave,

Table A.11: Number of years of education in the second wave of the RS and panel sample

Year	Panel	RS	Difference	t-test
1985	8.30	8.74	-0.44	0.00
1986	8.62	8.95	-0.33	0.00
1987	8.90	9.05	-0.14	0.00
1988	9.03	9.13	-0.10	0.00
1989	9.15	9.19	-0.04	0.03
1990	9.07	9.10	-0.02	0.12
1991	8.80	8.86	-0.06	0.00
1992	8.48	8.53	-0.05	0.00
1993	8.13	8.19	-0.06	0.00
1994	7.65	7.69	-0.04	0.00
1995	7.03	7.07	-0.04	0.00
1996	6.38	6.41	-0.03	0.00
1997	5.59	5.64	-0.05	0.00
1998	4.70	4.76	-0.07	0.00
1999	3.78	3.86	-0.08	0.00
2000	2.90	2.99	-0.08	0.00
2001	2.05	2.12	-0.07	0.00
2002	1.23	1.25	-0.02	0.00
2003	0.42	0.45	-0.02	0.00

Source: BDH Social Registry data.

however, people born between 1983 and 1995 are under-represented. These are children ages 8 to 20 years in 2003 and 19 to 31 years in 2013. Someone interested in the evaluation of the short-term effects of BDH on schooling using the information in the first and second waves of the panel would look at the effects on children ages 0 to 18 in 2003 who were born between 1985 and 2003. This group of interest overlaps with the children with lower link rates in the three-wave panel (born 1983 and 1995), for this reason, is very important to reweigh the observations. However, someone interested in the long-term effects of BDH would look at children that were exposed during the two phases of the program, namely, children between 0 and 18 years in 2008 who were also treated during the first phase. The group of interest would then be children ages 10 to 23 in the third wave who were born between

Table A.12: Number of years of education in the third wave of the RS and panel sample

Year	Panel	RS	Difference	t-test
1985	8.78	8.79	-0.01	0.62
1986	9.22	9.06	0.16	0.00
1987	9.61	9.26	0.35	0.00
1988	9.89	9.48	0.41	0.00
1989	10.17	9.72	0.45	0.00
1990	10.30	9.88	0.41	0.00
1991	10.29	9.99	0.31	0.00
1992	10.30	10.09	0.21	0.00
1993	10.24	10.09	0.15	0.00
1994	10.20	10.08	0.12	0.00
1995	10.01	9.91	0.10	0.00
1996	9.66	9.57	0.09	0.00
1997	9.20	9.12	0.08	0.00
1998	8.82	8.75	0.07	0.00
1999	8.30	8.22	0.08	0.00
2000	7.58	7.50	0.08	0.00
2001	6.81	6.72	0.09	0.00
2002	6.01	5.85	0.16	0.00
2003	5.07	4.94	0.13	0.00

Source: BDH Social Registry data.

1991 and 2004, which includes less under-represented children. Consequently, the long-term analysis would not be as much affected as the short term analysis by the low link rates for children born between 1983 and 1995. Yet, the use of raking weights would solve the problem for the age groups that are under-represented in the panel.

There are significant differences between the panel distribution and the RS distribution. This translates into statistically significant differences in terms of average years of schooling by birth cohort and is expected to happen as well with other variables that have not been presented here. In the following section, I explain the process to construct weights for the panel. The intention is to compensate for unequal probabilities of selection, non-coverage of some groups of the population, and non-response that lead to the observed departures between the panel and the RS.

A.6 Differential link rates and sample-balancing

As an additional check for differential link rates, I estimate a regression where the dependent variable is the binary variable that reflects the link status. A priori I know that is likely that people who turn 18 or people who are old enough to leave the household may be more difficult to find on subsequent waves of the RS despite my efforts to follow them. Also, some young children do not have IDs and there are some problems with the reported date of birth. However, after the clerical review of the results I can be quite confident that even though not all children report an ID, they are correctly matched based on names, last names, gender, and birth year.

Table A.13 shows that men have a lower linking probability than women and this effect is stronger when we focus on the linkage of the second and third wave, which suggests that more women started to be surveyed over time. Younger people with primary and secondary education are more likely to be correctly matched compared to people with higher levels of education, mostly because the latter are not the main target of the survey and are more likely to be dropped out of the RS. One of the main predictors of linkage is having an ID. The coefficient of this variable in the linkage of the first and second wave almost doubles the size of the coefficient in the linkage of the second and third waves. The reason may be that since more people report an ID in the second and third waves, having an ID is no longer a determinant factor of a good match, while in the match of the first and second waves, the lack of ID on the first wave makes that most of the good matches are due to ID linkage.

These results suggest that despite the high rate of successfully matched cases, being successfully linked is not random, new weights taking into account the linking bias need to be constructed to get a representative sample of the Ecuadorian population.

Table A.13: Probit regression to explain the link status between the first and second wave and between the second and third wave

	(1) Link wave 1 - wave 2		(2) Link wave 2 - wave 3
Gender	-0.0343***	Gender	-0.0670***
	(-29.53)		(-70.86)
Birth year	0.0158***	Birth year	0.000290***
	(410.15)		(10.33)
Score	-0.00789***	Score	-0.0221***
	(-128.88)		(-648.42)
Has ID	1.175***	Has ID	0.357***
	(686.22)		(309.55)
Constant	-30.60***	Constant	0.408***
	(-395.48)		(7.24)
N	6302758	*N*	8027317

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

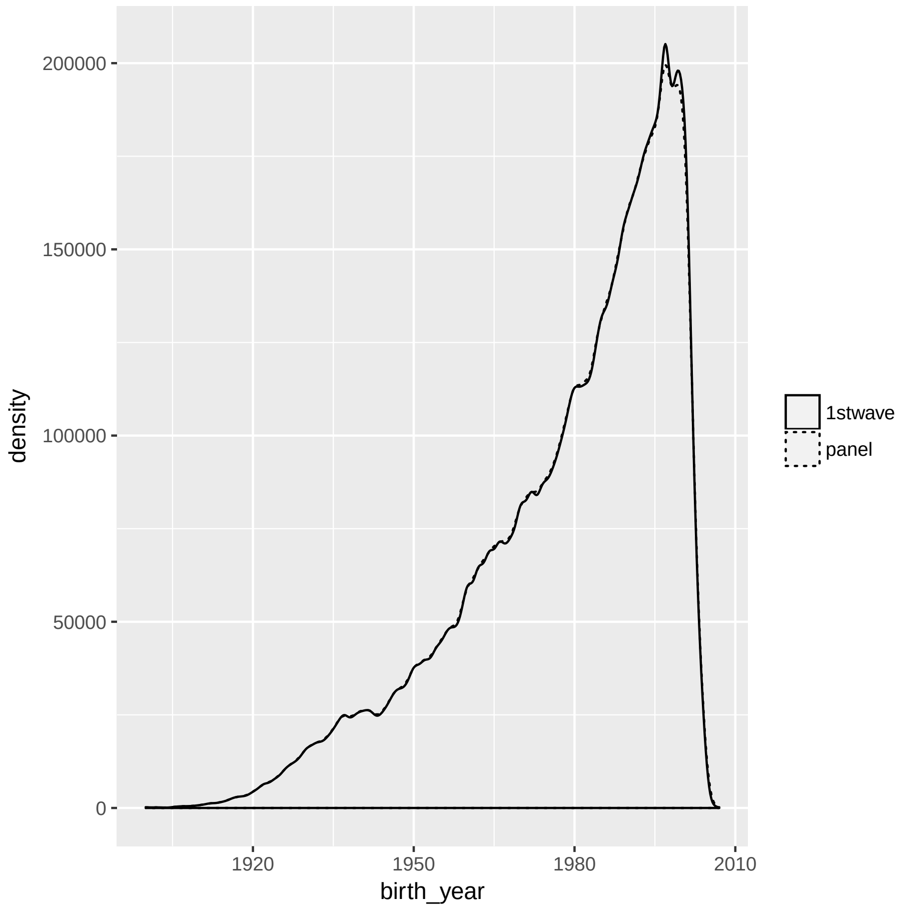
Notes: The first wave of RS has 6,3 million individuals, among them, 4,6 million or 72.44% could be followed in the wave 2. The second wave of RS has 8 million individuals, among them, 5,4 million or 67.37 % could be followed in wave 3.

A.6.1 Construction of weights

To make the sample distribution conform to a known population distribution it is necessary to adjust the sampling weights so that the marginal totals of the adjusted weights on key characteristics match the corresponding population totals. This process is known as raking or sample-balancing, and the key characteristics are known as auxiliary variables. Typical auxiliary variables include gender, age, years of schooling and geographic variables like province or county.

In sample-balancing is important to use as many auxiliary variables as possible. The reason is that if you make the response representative with respect to as many auxiliary variables as possible, it is likely that the response also becomes representative with respect to the other survey variables. I included gender, education

Figure A.7: Kernel density graphs of the distribution of birth year in the first wave of RS versus the first wave of the panel

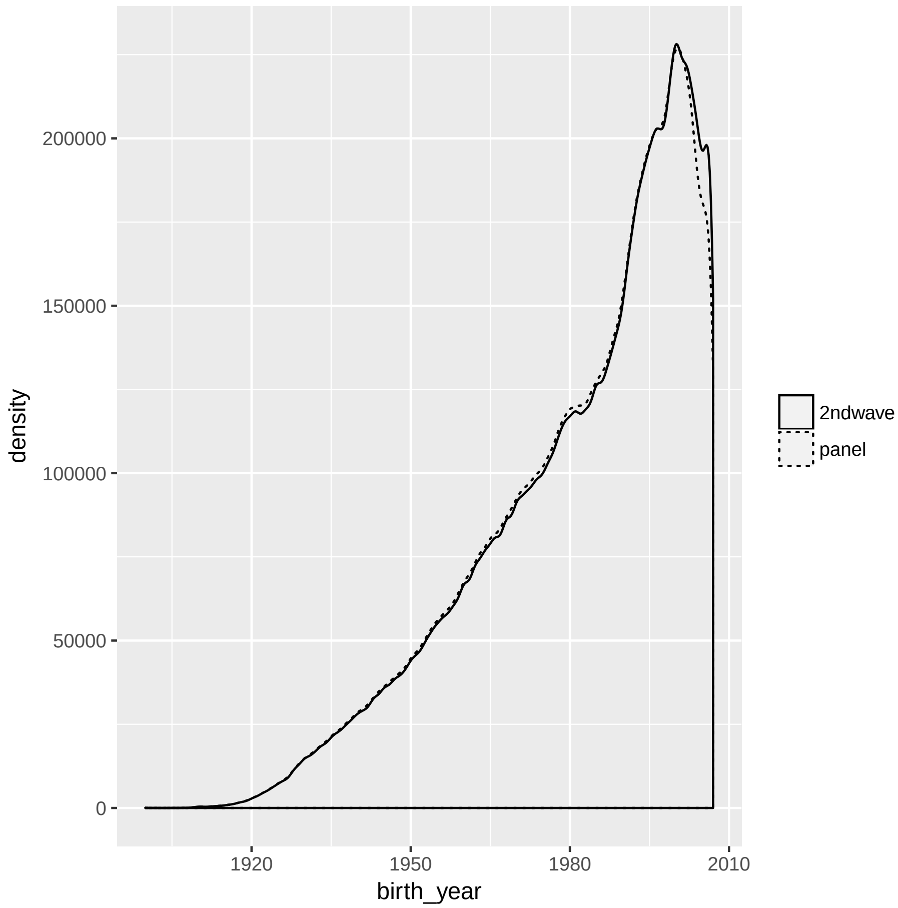


level, highest year of education approved, number of years of education, birth year, employment, province, county and type of house.

I construct sample weights so that the panel's totals on key characteristics match the totals of the corresponding wave of the Registro Social.[4] I use the Stata command <survwgt rake> (Winter, 2002). It creates sets of weights for replication-based variance estimation techniques for survey data including balanced repeated replication and several versions of the survey jackknife. These replication methods are alternates to the Taylor series linearization methods used by Stata's svy-based commands.

[4]These key characteristics include gender, education level, highest grade completed, the number of years of education, birth year, employment, province, county and type of house

Figure A.8: Kernel density graphs of the distribution of birth year in the second wave of RS versus the second wave of the panel



Figures A.7, A.8 and A.9 show that after correcting the sampling weights in the panel data, I obtain overlapping density curves for the date of birth and almost exact histograms for the variable years of education as shown in Figures A.10, A.11 and A.12.

Figure A.9: Kernel density graphs of the distribution of birth year in the third wave of RS versus the third wave of the panel

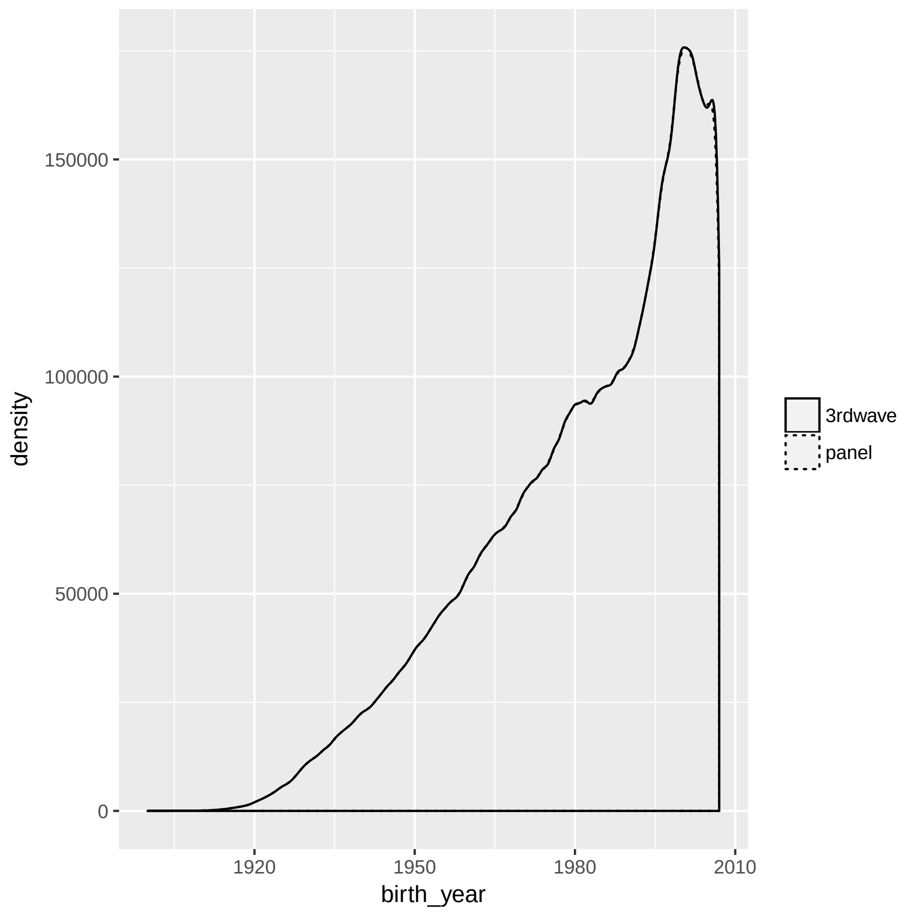


Figure A.10: Kernel density graphs of years of education in the first wave of RS versus the first wave of the panel

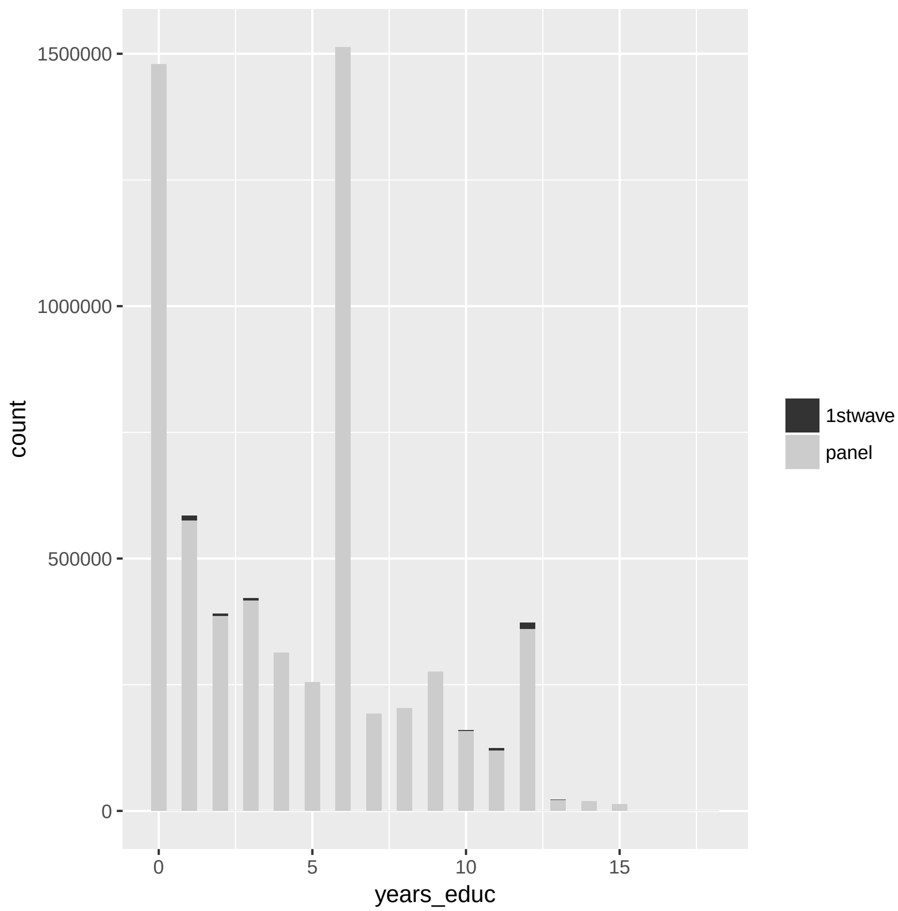


Figure A.11: Kernel density graphs of years of education in the second wave of RS versus the second wave of the panel

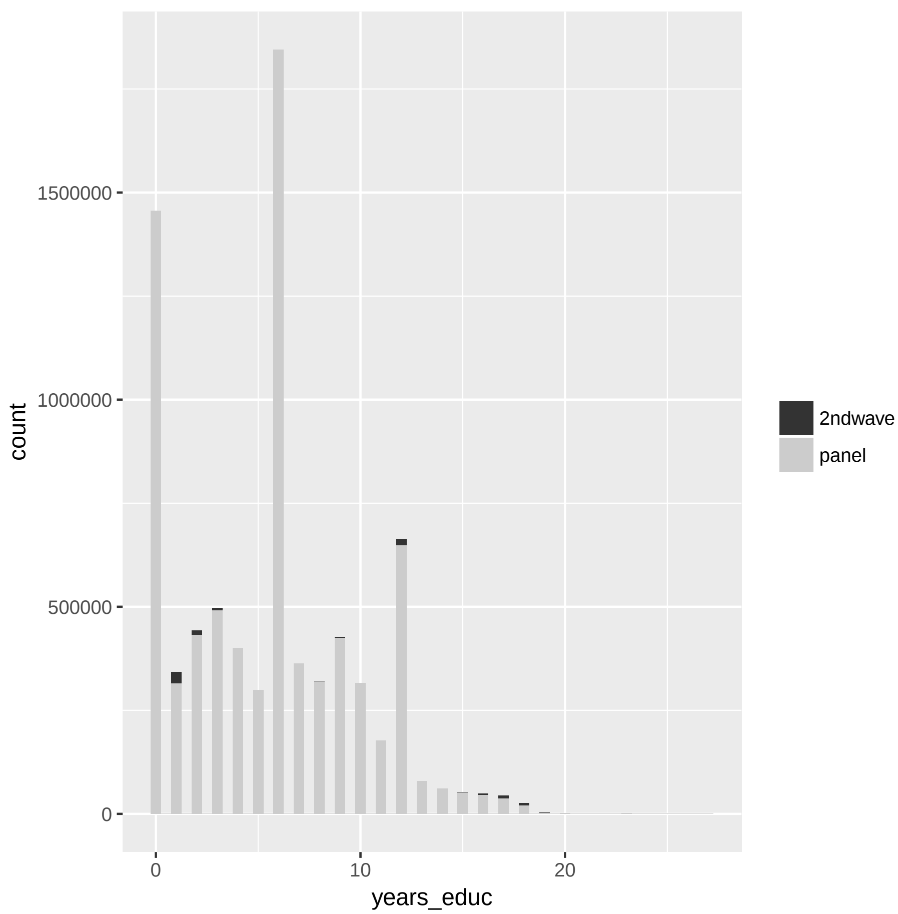
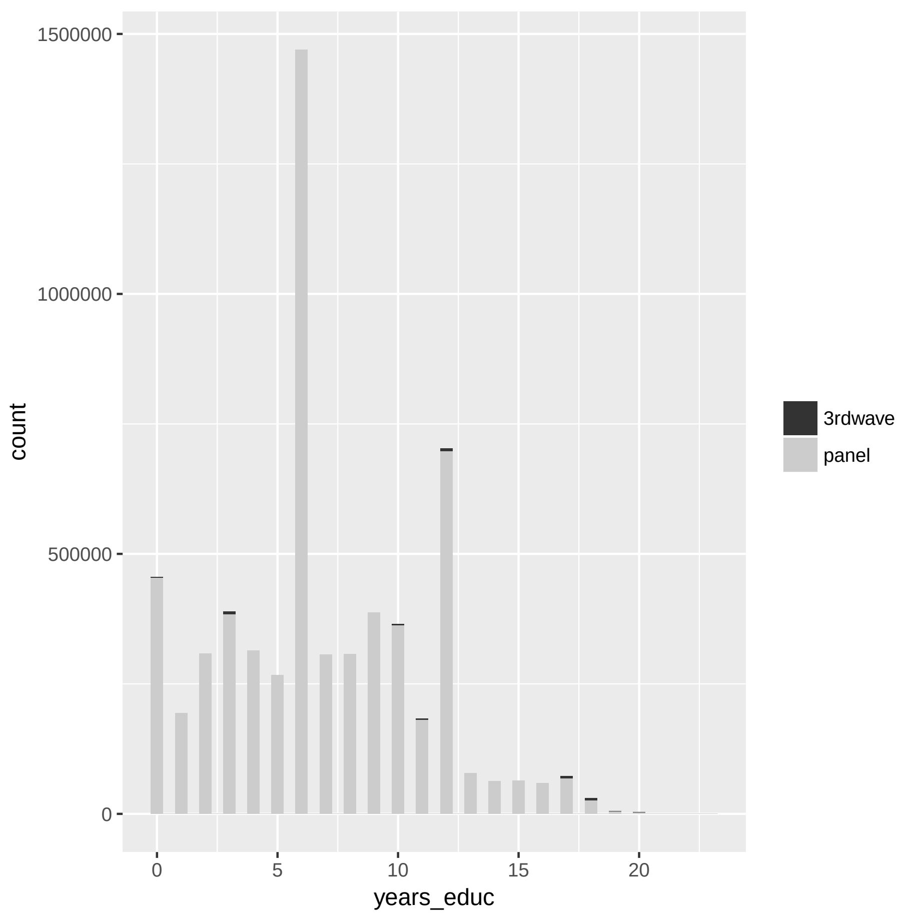


Figure A.12: Kernel density graphs of years of education in the third wave of RS versus the third wave of the panel



A.7 Conclusions

Latin American countries lag behind developed countries in terms of the availability of longitudinal data sets and longitudinal studies – where a defined population is interviewed over an extended period of time –. Considering that longitudinal data collection and analysis are critical to social research, policy, and practice, as well as a vital input for policy and program evaluation, governments must take actions to revert this reality. These actions should start by redoubling efforts in the collection of individual identifiers first through the improvement of the civil registration systems and secondly by implementing data collection protocols that contemplate revisiting people who do not report a valid ID until they do.

The reasons behind the lack of longitudinal studies and longitudinal datasets in the region are not only linked to the lack of funding or deficiencies of the Civil Registration Systems and data collection processes. Several technical and operational issues, as well as legal and ethical considerations, may hinder the use of datasets held by multiple organizations and constrain the extent to which data can be linked in practice. These problems must be solved through the establishment of inter-institutional agreements that settle the guidelines for the institutions to share information and generate new sources of information that are maintained over time.

There is also a lack of research about methods to link big administrative data sets, in particular, because the existing methods and software apply mostly to information recorded in English. This paper contributes to fill this gap in the literature and contrasts the performance of pattern matching and phonetic encoding to link several waves of Social Registry data recorded in Spanish. In general, the results of specificity and positive predictive values favor the use of probabilistic record linkage with a bi-gram string comparator over the alternative which involves the use of phonetic encoding to perform the merge. Probabilistic record linkage produces a higher rate of successful linkages, but also a higher rate of false positive matches, with best results achieved after a thorough clerical review of the linked pairs.

Other methods like machine learning can be implemented in the future to possibly improve the number of matched records. However, <reclink2> (Wasi and Flaaen, 2015) has proved to be a quite reliable tool and several public institutions use it to link surveys nowadays. In the UK, for instance, several waves of the Small Business Survey (SBS) are linked to the Inter-Departmental Business Register (IDBR) using the Stata command <reclink2> (Prabhat et al., 2015).

References

Blasnik, Michael (2010). "RECLINK: Stata module to probabilistically match records". Boston College Department of Economics.

Christen, P. (2006). "A comparison of personal name matching: Techniques and practical issues". In: *Sixth IEEE International Conference on Data Mining*. Computer Society.

Christen, Peter (2012). "A survey of indexing techniques for scalable record linkage and deduplication". In: *IEEE Transactions on Knowledge and Data Engineering* 24.9, pp. 1537–1555.

Fellegi, Ivan P. and Alan B Sunter (1969). "A theory for record linkage". In: *Journal of the American Statistical Association* 64.328, pp. 1183–1210.

Harron, K., H. Goldstein, and C. Dibben (2015). *Methodological Developments in Data Linkage*. Wiley Series in Probability and Statistics. Wiley, pp. 1–296.

Harron, Katie, Angie Wade, Ruth Gilbert, Berit Muller-Pebody, and Harvey Goldstein (2014). "Evaluating bias due to data linkage error in electronic healthcare records". In: *BMC Medical Research Methodology* 14, pp. 1–10.

Leite, Phillippe, Tina George, Changqing Sun, Theresa Jones, and Kathy Lindert (2017). "Social registries for social assistance and beyond: A guidance note & assessment tool". In: *World Bank Social Protection and Labor Discussion Paper* 1704.

Ponce, J and F Falconí (2011). *El trabajo infantil en Ecuador: marco institucional, evolución histórica y análisis costo beneficio de su erradicación*. Ed. by Gabriela Malo. Quito - Ecuador, pp. 1–87.

Prabhat, Vaze, Jonas Meldgaard, James Derbyshire, and Ben Davies (2015). "Small Business Survey: Linking 2006 and 2007 waves to the IDBR".

Sadinle, Mauricio and Stephen E. Fienberg (2013). "A Generalized Fellegi-Sunter framework for multiple record linkage with application to homicide record systems". In: *Journal of the American Statistical Association* 108.502, pp. 385–397.

Sayers, Adrian, Yoav Ben-Shlomo, Ashley W. Blom, and Fiona Steele (2015). "Probabilistic record linkage". In: *International Journal of Epidemiology* December 2015, pp. 954–964.

Wasi, Nada and Aaron Flaaen (2015). "Record linkage using Stata: Preprocessing, linking, and reviewing utilities". In: *The Stata Journal* 15.3, pp. 672–697.

Winter, Nick (2002). "SURVWGT: Stata module to create and manipulate survey weights". Boston College Department of Economics.