

Pay for Performance for Prenatal Care and Newborn Health: Evidence from a Developing Country

Maria Laura Alzua y Noemí Katzkowicz

Documento de Trabajo Nro. 272

Diciembre, 2020

ISSN 1853-0168

www.cedlas.econo.unlp.edu.ar

Cita sugerida: Alzua, M.L. y Katzkowicz, N. (2020). Pay for Performance for Prenatal Care and Newborn Health: Evidence from a Developing Country. Documentos de Trabajo del CEDLAS N° 272, Diciembre, 2020, CEDLAS-Universidad Nacional de La Plata.

Pay for performance for prenatal care and newborn health: Evidence from a developing country

Maria Laura Alzua & Noemí Katzkowicz

Abstract

Empirical literature analyzing the effect of pay-for-performance programs (P4P) for healthcare providers on maternal care and newborn health outcomes is scarce. In 2008, Uruguay's Ministry of Public Health implemented a P4P called *Metas Asistenciales* (Healthcare Goals), a country-wide program that grants healthcare providers an economic incentive for complying with certain maternal and newborn healthcare goals. Health organizations use these funds to provide maternal and child health services. Using administrative records and a difference-in-difference methodology, we evaluate the effect of the *Metas Asistenciales* program on maternal and newborn health outcomes. We find that in the institutions affected by the program, 10 percentage points more women received an adequate number of prenatal checkups and pregnancy detection in the first trimester improved by 5 percentage points. We also found better results among newborns for indicators related to birth weight, premature births, and stillbirths. In sum, the program had a positive, significant impact on the rate of pregnant women's utilization of health services and on newborn health outcomes. This study thus provides evidence supporting the idea that economic incentives are a promising tool for incentivizing healthcare providers to achieve better health services in developing countries.

Keywords: prenatal care, pay-for-performance programs, *Metas Asistenciales*, newborn health

JEL Classification: I1, O2, C5

1. Introduction

Investment in childhood development is a fundamental factor in achieving better outcomes in individual lives. Evidence suggests that early childhood interventions are more cost effective and achieve better returns than interventions at later stages in life (Heckman 1995, 2000). Specifically, newborn health and the reproductive behavior of mothers are essential variables that explain children's and adult's well-being (Heckman, 2007; Di Cesare y Sabates, 2017; Cunha et al., 2008; Heckman et al. 2013). Poor health conditions among newborns are associated with less adequate cognitive skills, health problems, poor educational performance, and lower income later in life (Boardman et al., 2002; Behrman and Rosenzweig, 2004; Black et al. 2007; Currie and Moretti, 2007; Royer, 2009). Health deficiencies among newborns lead to large healthcare burdens throughout life, which translate to higher costs for society. Health programs can generate direct improvements in pregnant women's health and can have an important positive effect on the health of newborns (Gertler, 2014). Guaranteeing universal access to quality maternal and child health services has thus become a fundamental tool for countries worldwide.

The costs and benefits associated with improving maternal and newborn health in developing countries have led governments to try a range of policies to achieve better outcomes. Pay-for-performance (P4P) schemes are one such attempt. Governments in developed countries have introduced them as incentives to increase the quality of healthcare service. The main rationale behind P4P is that allocating resources to the health sector without any incentive mechanism to improve performance will not result in better health outcomes (Toonen et al. 2009). The prospect of inducing healthcare providers to improve thus makes these types of programs an attractive tool for governments.

Several studies have shown that inducing healthcare providers to adopt new practices depends on a variety of determinants, including the characteristics of the institutions, their management, the costs of adopting new practices, and psychological factors (Berwick, 2003; Cabana et al., 1999). Celhay et al. (2019) discuss the role of incentives resulting from P4P programs. They state that if the institutions perceive the new practice will be beneficial in the future they are more willing to use the financial incentives to pay for the cost of implementing such new practices and vice versa. In terms of the principal-agent model, P4P designs offer incentives for achieving results and thus reduce the need to monitor the agents' behavior (Eldrige and Palmer, 2008).

Creating financial incentives for improving maternal and newborn health leads healthcare institutions to develop strategies to encourage maternal care. The literature on the topic has shown that the introduction of different incentive schemes rewarding performance has had positive effects. Celhay et al. (2019) found that a P4P implemented in Argentina (*Plan Nacer*), which aimed to increase early initiation of care during pregnancy, increased the number of women who came for checkups during the first trimester of pregnancy and encouraged healthcare institutions to adopt new strategies to attract the targeted population. Celhay et al. (2019) found, that the program had no effect on newborn health. However, increased checkups for mothers persisted for at least 24 months after the financial incentives were removed. Messen et al. (2007) highlights the important role of P4P systems in improving the poor service provided by healthcare facilities in Rwanda. Olken et al. (2013) find that a performance-based financing system provided by the local government in Indonesia had positive effects on its intended outcomes, which were to improve health services and achieve better educational results. Regalia and Castro (2007) analyze a program in Nicaragua that grants an incentive to the managers of healthcare institutions to promote neonatal health services and transfers money in a conditional manner to the children who are beneficiaries of the program; they find increased use of health services and higher vaccination rates. Experimental

evidence for Rwanda shows that public healthcare providers reacted positively to a P4P program created to improve the quality of the services provided for neonatal and infant health (Basinga et al., 2011; Gertler and Vermeersch, 2013; Walque, 2013). Although the evidence in these particular cases is strong, evidence for developing countries remains scarce and is mostly limited to small-scale experiments (Gertler, 2014; Gertler, 2015; Hullieri and Seban, 2014).

This article seeks to analyze the effects of a larger, national P4P scheme called *Metas Asistenciales*, which was implemented in the health sector in Uruguay in 2008. The program's main objective was to improve the quality and accessibility of the health services provided to mothers and children. The government offered an economic incentive to healthcare providers who complied with certain care goals. This program is similar to *Plan Nacer* in Argentina, but the latter was for uninsured pregnant women and children under six. To the best of our knowledge, the program is the first countrywide P4P to be introduced in the health sector in a developing country.

Even though we do not observe experimental variation in the introduction of the P4P scheme, its design allows us to estimate the causal effects and the lower bounds for its effects due to certain administrative peculiarities of the intervention that we explain later in the article. The identification strategy exploits the gap in the implementation of the P4P program within a range of public and private healthcare providers and the fact that within the private sector there was a disparity in the degree of adherence.

We compare different health indicators among treated and untreated mothers and newborns, before and after implementation, using data from the Perinatal Information System (SIP), a rich administrative dataset which accounts for all births in the country between 2002 and 2013. Our data allows us to control by mothers' characteristics,

institutional characteristics, regional features, and any other fixed effects between treated and untreated groups.

We explore the channels that can affect newborn health outcomes and suggest that increasing the number of prenatal checkups, improving early detection of pregnancy and the gestational age are fundamental factors that explain the changes in newborns' indicators. An extensive body of medical and economics literature identifies early detection of pregnancy and maternal care as key determinants of birth outcomes. These variables are important determinants to detect diverse risk factors (e.g., diseases, infections, and anemia) that may affect gestational length, intrauterine life conditions, and birth weight (Carroli et al., 2001; Evans and Lien, 2005; Kramer, 1996; Rosenzweig and Schultz, 1983; Slattery and Morrison, 2002). Furthermore, according to Almond et al. (2008) there could be different channels that could generate negative effects on newborn health. On one hand, if the improvements in the intrauterine conditions lead to fewer fetal deaths, there could be a composition effect that generates a negative impact on the birth weight due to the survival of the fetal margin. On the other hand, if the program affects the fertility rates of women from more disadvantaged backgrounds, that could imply a negative impact on neonatal health.

Overall, our results show that Uruguay's P4P program increased, on average, the number of prenatal checkups and the number of times pregnancy was detected early. We also observe modest improvements in newborn health outcomes, particularly in birth weight and neonatal mortality. We observe a 1.2 percentage points reduction in the probability of low birth weight, a 0.36 percentage points reduction in premature births, and a reduction in stillbirth's rates of 2.4 per 1,000 live births. The average weight of newborns rose by 10.92 grams. The observed effects are lower than other papers (see Amarante et al., 2016 and Harris et al., 2015).

The article contributes to the small amount of research that has so far been published on P4Ps and their effects in developing countries. To produce our results, we use a rich source of information that includes data on the institutions affected by the program and information on the pregnant women and their children. The use of a copious nationwide administrative dataset allows us to estimate the causal effect of the program for nearly all of births nationwide.

The article is organized as follows: section 2 provides a brief description of the Uruguayan health system and the P4P program. Section 3 describes the methodology and the data. Section 4 presents the results as well as robustness checks. Finally, section 5 and 6 discusses the results and concludes.

2. The health system and the P4P scheme in Uruguay

Uruguay is a small country with a population of 3.4 million located in South America. In the early 2000s, its poverty level was over 30% and its inequality level measured by the Gini Index was 0.471, according to the data provided by National Institute of Statistics. Before 2005, approximately 9% of Uruguayan newborns were born underweight. This figure increased during the severe economic crisis the country suffered in 2002. A similar trend can be observed for average newborn weights and for the percentage of infants born prematurely.

In 2005, the left-wing party took office and implemented several anti-poverty policies. First, between 2005 to 2007, the government created a cash transfer program (*Plan de Atención Nacional a la Emergencia Social*, or PANES) targeting the poorest 10% of Uruguayan households. In 2008, the PANES program was replaced by a conditional cash transfer program (CCT) called the *Asignaciones Familiares Plan de Equidad* and an unconditional cash transfer program. In 2008, the government implemented a

healthcare reform with the goal of increasing health service coverage and improving the quality of the sector.

The organization of the health sector is rather complex and has undergone several changes in recent decades. In 2008, the Uruguayan government implemented one of the most important healthcare reforms to date. Prior to this reform, the system included coexisting public and private healthcare systems. The public sector provided coverage to the low-income population, while the private sector, dominated by the country's Institutions of Collective Medical Assistance (IAMCs) and health insurance companies, catered to privately paying customers and workers in the formal sector (MSP, 2010). Before 2008, workers registered with the social security system had individual coverage, but this did not cover their families.¹

The healthcare reform implemented in 2008 created the Integrated National Health System (SNIS). Under the principle of universal care, the SNIS sought to guarantee coverage to the entire population and to improve access to quality health service. Hence, all public and private providers became part of Uruguay's Health System. This system was financed through the contributions of workers in the formal sector and extended insurance access to the spouses of formal-sector workers, their children under 18 years old, and retired workers (Bergolo and Cruces, 2011). Funds were transferred to the National Health Fund (FONASA) and, from there, the healthcare institutions where the users registered received a per capita variable payment if they met certain requirements. In this way, the funds financed a minimum set of integral benefits exhaustively defined by the Ministry of Public Health (MSP, 2010).

¹ The healthcare reform is composed of three major laws: Law N° 18.121 which created the National Health Fund (FONASA); Law N° 18.211 which created the SNIS; Law N° 18.161 which decentralized State Administration of Health Service (ASSE) (Ríos, 2013).

The implementation of the SNIS generated an overall increase in healthcare coverage. Today, close to 100% of Uruguayans are registered with one of the comprehensive healthcare providers, either through social security, individual membership in the private sector, or through the public sector. Uruguay's health insurance system is thus now divided into these three sectors (public, private, and health insurance companies). Table 1 shows the relative percentages of affiliation with each insurance sector, including people registered through the social security system along with the rest of the users.

Table 1. Percentage of affiliation to each health subsector

	Affiliation through Social Security	%	Individual affiliation	%	
Public sector	191785	17	947640	83	1139425
Private sector	1628995	81	387078	19	2016073
Health insurance	45079	53	39535	47	84614

Source: Ministry of Public Health, 2012.

The 2008 reform also generated a change in Uruguay's model of care by seeking to strengthen the country's Primary Healthcare (APS) strategy. In more recent years, new benefits have been added and coverage has been extended to mental health, sexual and reproductive health, and oral health services. To reinforce the model-of-care reform and improve service quality, the state implemented a series of health policies, the P4P program being the most important. P4P grants monetary transfers to healthcare institutions when they comply with prioritized services for specific sub-populations of interest. Among the sub-populations prioritized are maternal and neonatal, young adolescents, and the elderly (MSP, 2010).

The pay-for-performance program and maternal and neonatal health services in Uruguay

Both the public and private sectors provide maternal and child healthcare coverage. Since 2008, prenatal care has been provided to patients for free. The program's novelty

lies in its P4P strategy, which grants healthcare providers funds if they comply with specific goals determined by the Ministry of Public Health (MSP). The services prioritized seek to improve coverage and health outcomes (National Health Board -JUNASA, 2007), which were established in a series of guidelines. The maternal care goals the healthcare providers have had to perform are: increasing the number of prenatal visits (establishing a minimum of six prenatal care visits per pregnant woman), increasing the number of pregnancies detected during their first trimester (early detection of pregnancy), improving perinatal medical history records, increasing syphilis and HIV testing during the first and third trimesters, and improving dental care (MSP, 2010).

Initially, the program was going to be implemented across all public and private care providers altogether, but ended up targeting only private healthcare providers. This difference in the implementation stemmed primarily from administrative problems that required public sector institutions to receive a reduced budget credit if they received extra-budgetary resources.² The public sector was thus unable to benefit from the economic incentive established by this P4P (Law 18.211). This created a gap in implementation that serves as a source of exogenous variation.³ Furthermore, within the private sector there was a disparity in the degree of adherence, which we also use as a source of identification.

The financing of the program is structured as follows. For each SNIS member registered with a given healthcare provider, USD\$2.50 is allocated to that institution to the extent to which it complies with the established goals. This represented approximately 6% of FONASA's health quota expenditures (MSP, 2010). The amount of the incentives given

² Article 59 of Law 18.211.

³ The implementation scheme thus turned out to be a bit counter-intuitive, since it first favored private centers instead of public institutions. Public institutions, which serve poorer individuals, would normally be prioritized in this kind of effort.

to the institutions thus varies with the number of affiliated people registered with them through the social security system. The program seeks to use this mechanism to improve health outcomes for mothers and newborns nationwide and to improve the quality of the services provided.

The Ministry of Public Health measured compliance with the goals using healthcare providers' reports. The threshold was determined by the Ministry in agreement with the healthcare providers that are part of the SNIS. Performance pay is divided equally between the various objectives to be met. An institution would need to comply with 70% of the established components to meet the Ministry-defined goals and thus receive the full payment; otherwise, a proportional payment would be transferred in accordance with the proportion of the goal achieved. The program randomly audits healthcare providers to verify the information they report (MSP, 2010).

An additional issue to consider is that the economic incentive is temporary, with an average duration of two years. After that period, some components of the P4P scheme change and the incentives are allocated in other areas for which improvements are sought. Table 2 shows the different components that have been the target of financial incentives through the years.

Table 2. Components of the P4P regarding the maternal and neonatal population since 2008.

(1) Number of women with 6 or more prenatal checkups	(2) Complete medical history in SIP	(3) HIV and VDRL tests	(4) Early detection of pregnancy (first trimester)	(5) Pregnant women with dental check-up	(6) Information regarding domestic violence.
------------------------------------------------------	-------------------------------------	------------------------	----------------------------------------------------	-----------------------------------------	----------------------------------------------

This incentive structure is consistent with the economic literature on P4P schemes. They seek to induce beneficiaries to adopt practices that may be costly to implement, but that

are economical in the long term (Celhay et al., 2019; Gertler, 2014). The program is producing a large institutional adaptation in terms of the availability of human resources, the incorporation of certain practices that did not exist until that moment, and institutional reorganization. It is important to clarify that these incentives can be used to finance the expenses of implementing the program, but cannot be used to finance salaries (MSP, 2010).

To carry out the goals of the P4P, the healthcare providers incorporated new strategies with the aim of enhancing the quality of maternal and children's care services. On one hand, there was an important increase in the number of health workers. From 2007 to 2008, the number of administrative workers increased around 15% in the private sector and in the institutions with a higher level of compliance with the program. The number of hours worked by gynecologists climbed from an average of 25 hours per week to 30 hours. Other strategies implemented by the providers were the diffusion of maternal health services, advertisement campaigns, and telephone calls to recruit pregnant women, especially for the early detection of pregnancy. The organizations' managers also monitored the execution of the guidelines more intensely. Not all the institutions reacted in the same way to the incentives. Institutions with a high number of users registered through the social security system, for whom the economic incentive in absolute terms is greater, were more likely to adopt new strategies to fulfill the goals established by the program (MSP, 2012). In sum, the factors that might have led to changes in the analyzed outcomes included (1) the amount of the incentives offered to the healthcare providers, (2) the use of the incentives to pay for the costs of the program (human resources, advertisement campaigns, telephone calls, etc.), and (3) stricter monitoring of the provider organizations.

2. Methodology and data

Our analysis is focused on all births in both public and private sectors, nationwide. We attempt to identify the causal effect of Uruguay's P4P on children's health outcomes by exploiting the difference in the implementation of the P4P in the public and private sectors. Due to the heterogeneity of the individuals treated in both sectors, we engage in a comprehensive exploratory analysis to balance the groups in terms of the observable variables of interest, during the pre-implementation period.

3.1. Data Sources

The sources of information used are the aforementioned SIP, which provides individualized information on births and data on mothers, as reported by health centers. We also used data from the National System of Information (SINADI) and System of Control and Analysis of Human Resources (SCARH) from the Ministry of Public Health to obtain information on the health centers. Finally, we used data provided by the Ministry's Division of Health Economics to measure the degree to which each health center adhered to policy guidelines. Data on prenatal care during pregnancy and newborn health outcomes is extracted from the SIP. When compared with birth certificates, this covers, on average, 80% of all births nationwide. Healthcare providers record several key indicators at the time of birth. Our analysis was performed for the period from 2002 to 2013, in order to include several years before and after the implementation of the policy. We are able to link individual information to each health center. We identified the nature of the sector (public or private), the province where the mother received prenatal services, and where she gave birth.

Several variables are used to measure maternal outcomes and newborn health in our analysis. For prenatal care, we look at whether mothers met the standards set by the Ministry of Public Health or the World Health Organization (WHO) (6 and 9 prenatal checkups during pregnancy, respectively), and whether pregnant women were seen by their healthcare provider during the first trimester of pregnancy.

Among newborn health outcomes, we include birth weight, number of weeks of gestation, and stillbirths. In order to account for birth weight, we used the percentage of children born with low birth weight (defined as less weighing than 2,500 grams) and the average birth weight. The variables related to premature birth include the number of weeks of gestation and the percentage of premature births (a birth is considered premature when the mother had fewer than 37 weeks of gestation). Finally, regarding fetal mortality, the number of stillbirths per 1,000 live births is used.

It is important to note is that many policy reforms were implemented at the same time as this one in Uruguay, and previous impact evaluations have shown that they affected newborn outcomes. For example⁴, Amarante et al. (2016) estimate, based on program and social security administrative microdata matched to longitudinal vital statistics, that participation in the PANES program led to a sizable reduction in the incidence of low birthweight in the same period. Along the same lines, Harris et al. (2015) consider the impact of different types of tobacco control policies that were implemented in the same period and find that quitting smoking during pregnancy increased birth weight significantly. Because of the possible effects of programs run simultaneously with the P4P, and to account for these issues in our model, we are including several control variables. First, we incorporated the percentage of households receiving the PANES and the CCT program by health sector (private and public), by province, and by year. Second, we included a dummy that reflected whether the pregnant women were smokers or not. Table 2 displays the details about the outcome and control variables used in each specification.

Table 3. Variables used within the analysis

Data	Period	Source
Outcome Variables		

⁴ We one of the anonymous reviewer for this suggestion.

adequate prenatal check-ups (WHO guidelines) adequate prenatal check-ups (MSP guidelines) Prenatal check-ups in the first trimester of pregnancy New born weight Low birth weight Weeks of gestation Premature children Stillbirths	2002 – 2013	Perinatal Information System – MSP
Control Variables	2002 – 2013	Perinatal Information System – MSP
Mothers' age Education level (dummies) Marital status (dummies: married, single, common law marriage) Quantity of births and prior miscarriages or abortions Pathologies (hypertension, diabetes and syphilis) Habits (including alcohol and cigarette use, etc.) Time of the year within pregnancy (Winter) Type of birth (natural birth or C-section) Trimester of birth Multiple pregnancy Newborn gender		
PANES program and CCT	2006-2013	Household Survey
Data of adherence to the program	2008-2010	Division of Health Economics (MSP)

Before turning to the identification strategy and results, we analyze the variables for the period 2002–2013 in the private and public sectors. Table 4 presents the average and standard deviation for the outcome variables for the period under consideration, for all pregnant women.

Table 4 - Descriptive statistics for the outcome variables. Period: 2002-2013

	Average	SD	Observations
Adequate prenatal check-ups (WHO guidelines)	0,412	0,492	468806
Adequate prenatal check-ups (MSP guidelines)	0,545	0,498	463734

Early detection of pregnancy	0,579	0,494	468806
Newborns' weight	3222,539	585,439	490652
Low birth weight	0,084	0,278	497551
Stillbirth	0,020	0,158	412267
Weeks of gestation	38,452	2,065	486777
Premature	0,093	0,290	497551

Source: Based on data from the Perinatal Information System, 2002-2013

3.2. Identification strategy

As mentioned above, according to the literature, prenatal care is closely related to newborn health outcomes. Mothers with fewer checkups during pregnancy are more likely to have premature babies, and their children are more likely to experience intrauterine growth delays and low birth rates. Nevertheless, a degree of endogeneity in this relationship is caused by omitted variables or selection bias: mothers with certain socioeconomic characteristics have a higher (lower) propensity to access and use health services, which affects the health of their newborns. This means that the ordinary least squares estimation is not consistent.

To address this problem, we used an identification strategy that relies on a source of exogenous variation in the use of health services by pregnant mothers. As explained above, we used the Ministry of Public Health's P4P program, which targeted mothers and children in 2008, as a source of plausible exogenous variation in prenatal care. Specifically, there are two sources of variation in the implementation of the P4P. On the one hand, there was a difference between the policy's implementation in the private and public health sectors, which was caused by issues unrelated to the outcome variables of interest. At first, the P4P was supposed to start nationwide, but there were administrative problems in the public sector, which established that the public sector would receive a reduced budget credit if they received extra-budgetary resources. Thus, the public sector could not benefit from the economic incentive established by the P4P (Law 18.211). In order to extend the P4P to the public sector, it is necessary to change existing laws. This

created a gap in implementation that serves as a source of exogenous variation. On the other hand, within the private sector, there was a disparity in the degree of policy adherence, wherein some institutions reached the thresholds set by the Ministry of Public Health faster, and others did not. We also used this second source of variation to evaluate the effect of the treatment.

To sum up, the program design and its implementation created several degrees of exposure to treatment among treated units (health facilities belonging to SNIS). We exploited the heterogeneity observed in implementation for our main analysis and for the robustness controls. The estimations we performed allow us to recover causal estimates. Counterintuitively, the first group affected were low-risk mothers. Since the program affected maternal care directly, and this variable could impact newborn health indicators, we estimated a difference-in-difference (DiD) model to evaluate maternal and child outcomes.

For the first analysis, since utilization by private and public sectors was very different, we performed an exploratory analysis to balance the groups before engaging in further empirical evaluation. Following a series of tests, we excluded teenage mothers (those under 18 years of age) from our analysis, for several reasons. First, they are the highest risk group of mothers among pregnant women. Second, they are predominantly in the public sector. Third, one of the main features of the 2008 healthcare reform was the inclusion of children (under 18 years of age) of formal-sector workers into the health system. The latter would affect our identification strategy by way of a compositional effect generated by a public-private crowding out.

Before the policy implementation, 74% of teenage mothers were in the public health sector, while this number went down 65% after the policy was implemented. Table 5

displays the percentage of teenage mothers in the public and private health sectors before and after the policy.

Table 5 - Age distribution of mothers according to subsector. Pre and post policy period.

	Pre policy 2002-2007		Post policy 2008-2013	
	Public sector	Private sector	Public sector	Private sector
Adolescents	0.240	0.089	0.273	0.113
Age 20-34	0.652	0.744	0.628	0.713
Age 35-39	0.080	0.134	0.075	0.143
More than 40	0.040	0.049	0.027	0.033

Source: Based on data from the Perinatal Information System, 2002-2013

Tables 6 and 7 show the average of the covariates for the treatment and control groups before and after the policy, as well as the results of the DiD estimator for each of these variables, first including all mothers and then excluding teenage mothers. These results reveal the existence of a composition effect when teenage mothers are included. Table 6 shows that, when all mothers are considered, there is a change in the groups' composition, before and after the policy. When teenage mothers are excluded, the effect vanishes, with no significant differences in most of the covariates between the groups for pre- and post-policy periods.

Table 6 – Descriptive statistics of the covariates for the control group and treatment for pre and post policy and effect composition considering all pregnant women.

	Control group				Treatment group				Dif in Dif
	Pre treatment		Post treatment		Pre treatment		Post treatment		
	Average	SD	Average	SD	Average	SD	Average	SD	
Number of observations									
N(total=493047)	142096		95968		132063		122920		
Share	0.288		0.195		0.268		0.249		
Individual features									
Age	25.069	6.673	24.575	6.661	28.362	6.189	28.050	6.462	0.0700*
Age 2	672.963	364.387	648.314	360.093	841.402	354.912	831.327	365.172	8.539***
Ages 20 to 34	0.652	0.476	0.628	0.483	0.744	0.437	0.713	0.452	-0.00836***
Ages 35 to 39	0.080	0.271	0.075	0.264	0.134	0.341	0.143	0.350	0.0106***
Ages 40+	0.040	0.196	0.027	0.161	0.049	0.215	0.033	0.178	0.000762
Primary School	0.424	0.494	0.378	0.485	0.159	0.365	0.111	0.314	0.00529**
High School	0.500	0.500	0.559	0.497	0.517	0.500	0.605	0.489	0.0332***
University	0.022	0.145	0.026	0.159	0.204	0.403	0.265	0.441	0.0471***
Married	0.229	0.420	0.141	0.348	0.482	0.500	0.343	0.475	-0.0627***
Common law marriage	0.474	0.499	0.557	0.497	0.323	0.468	0.506	0.500	0.107***
Single	0.226	0.418	0.234	0.424	0.098	0.297	0.122	0.327	0.0167***
Abortions	0.161	0.368	0.176	0.381	0.167	0.373	0.187	0.390	0.000149
Pregnant women smoke	0.255	0.436	0.283	0.450	0.094	0.292	0.119	0.323	-0.00975***
Hypertension	0.022	0.147	0.020	0.139	0.021	0.143	0.019	0.137	-9.21e-05
Drink alcohol	0.000	0.020	0.012	0.107	0.000	0.021	0.003	0.058	-0.00785***
C-section	0.226	0.419	0.237	0.425	0.387	0.487	0.363	0.481	-0.0290***
Child's gender	0.503	0.500	0.505	0.500	0.489	0.500	0.498	0.500	0.00520*
Multiple pregnancies	0.023	0.149	0.018	0.134	0.022	0.146	0.020	0.140	0.00291***
CCT program	0.038	0.046	0.114	0.043	0.027	0.039	0.053	0.030	-0.0402***

Table 7 - Descriptive statistics of the covariates for the control group and treatment for pre and post policy and effect composition, excluding teenage mothers

	Control group				Treatment group				Dif in Dif
	Pre treatment		Post treatment		Pre treatment		Post treatment		
	Average	SD	Average	SD	Average	SD	Average	SD	
Number of observations									
N(total=430851)	117171		76892		124035		112753		
Share	0.272		0.178		0.288		0.262		
Individual features									
Age	26.805	5.987	26.487	6.001	29.082	5.610	29.047	5.710	-0.00465
Age 2	754.370	346.799	737.576	344.858	875.852	334.226	879.386	337.958	0.678
Ages 20 to 34	0.791	0.407	0.784	0.411	0.792	0.406	0.778	0.416	0.00573**
Ages 35 to 39	0.097	0.296	0.094	0.292	0.143	0.350	0.156	0.363	0.00189
Ages 40+	0.033	0.178	0.029	0.167	0.035	0.184	0.032	0.177	-0.00154
Primary School	0.426	0.494	0.381	0.486	0.155	0.362	0.104	0.305	0.00248
High School	0.503	0.500	0.550	0.498	0.513	0.500	0.590	0.492	0.0430***
University	0.026	0.158	0.031	0.174	0.219	0.414	0.288	0.453	0.0190***
Married	0.266	0.442	0.168	0.374	0.512	0.500	0.370	0.483	-0.0357***
Common law marriage	0.487	0.500	0.572	0.495	0.319	0.466	0.507	0.500	0.0831***
Single	0.184	0.387	0.191	0.393	0.081	0.273	0.095	0.293	-0.000584
Abortions	0.186	0.389	0.203	0.402	0.177	0.381	0.198	0.399	0.00247
Pregnant women smoke	0.257	0.437	0.295	0.456	0.092	0.290	0.117	0.322	0.00163
Hypertension	0.025	0.158	0.023	0.150	0.022	0.147	0.020	0.141	0.000205
Drink alcohol	0.000	0.019	0.011	0.104	0.000	0.020	0.003	0.056	-0.00262***
C-section	0.235	0.424	0.244	0.430	0.395	0.489	0.371	0.483	-0.0453***
Child's gender	0.506	0.500	0.505	0.500	0.492	0.500	0.498	0.500	0.00418
Multiple pregnancies	0.025	0.155	0.020	0.141	0.022	0.148	0.021	0.143	0.00167
CCT program	0.038	0.046	0.113	0.044	0.027	0.039	0.053	0.030	-0.0419***

It is worth pointing out that excluding teenage mothers, one of the groups most at risk during pregnancy, may result in an underestimation of the possible effects of the policy (Williamson, 2013). Given this possibility, it is important to be careful when extrapolating our results to all pregnant women and births, as it should be taken into consideration that the excluded population has specific characteristics, and that a policy like this P4P could have differential effects on the excluded sector.

We thus pursue two different strategies. In our first strategy, our treatment group consists of pregnant women in the private health sector, without teenage mothers, while our comparison group is made up of pregnant women in the public sector, also excluding teenage mothers. In our second strategy, we exploit the variability in the degree to which the private health sector was exposed to the policy. To try to correct the endogeneity problem that results from the fact that institutions adhered to the treatment according to the mothers' characteristics, to construct the treatment group we narrowed our sample to pregnant women living in provinces that, on average, exceeded the compliance threshold set by the Ministry of Public Health by a deviation of 0.05. It is important to note that private healthcare providers are distributed unequally across national territory. At the provincial level, we can say the institutions are distributed randomly: some institutions exceeded the threshold and others did not. Analyzing results geographically is therefore able to reduce selection bias. Table 8 reflects the number of public and private institutions by province that exceeded the threshold and the number that did not. Hence, we used this second source of variation to evaluate the effect of the treatment and its intensity.

Table 8 – Number of private institutions adhered to the policy per province, 2008

	Compliance within the private sector	No compliance within the private sector	Total institutions
Artigas	0	1	1
Canelones	2	1	3
Cerro Largo	0	2	2
Colonia	1	2	3

Durazno	1	0	1
Flores	0	1	1
Florida	1	0	1
Lavalleja	1	0	1
Maldonado	0	2	2
Montevideo	9	10	19
Paysandú	0	1	1
Río Negro	2	0	2
Rivera	2	0	2
Rocha	1	0	1
Salto	0	1	1
San José	0	1	1
Soriano	1	0	1
Tacuarembó	0	1	1
Treinta y Tres	1	1	2

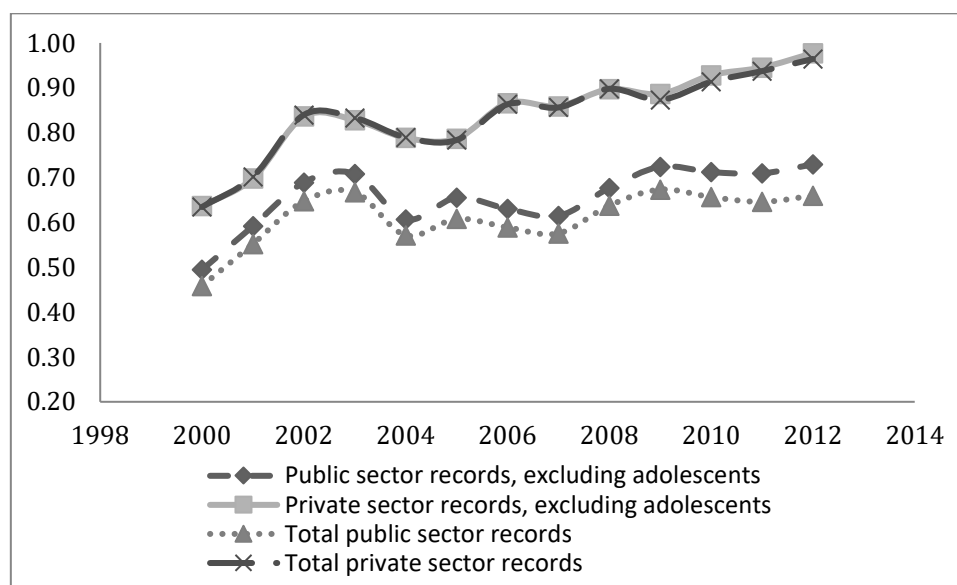
In that context, the treatment group in our specification is composed of pregnant women in the private sector who lived in provinces where private healthcare providers exceeded a deviation of 0.05 from the threshold imposed by the Ministry of Public Health at the beginning of the implementation. The control group is comprised of pregnant women in the private sector who lived in provinces where the providers did not exceed a deviation of 0.05 from the threshold imposed by the Ministry at the beginning of the implementation.

We took advantage of microdata made available from the perinatal information system. We performed an analysis of the 2002–2013 period, which allows us to consider the implementation of the program for private healthcare providers in 2008 and analyze the effect some years after the policy implementation. Our sample records 430,851 births nationwide during this period, excluding children born to teenage mothers. We thus analyzed 83% of national births recorded in the SIP during the period under review.

It is important to highlight some of the limitations of SIP data: total births are underreported, and some reports are incomplete. To test the consistency of our data, we compared SIP records with data from the Certificate of Live Births (CNV) for the period

under consideration. Figure 1 shows how records changed over time throughout the period for both sectors, for all pregnant women and for non-teen mothers.

Figure 1 – Evolution of administrative records. Comparison between the Perinatal Information System (SIP) and the Certificate of Live Birth (CNV).



Source: Data from the Perinatal Information System, 2000-2013

Table 9 presents descriptive statistics of the variables of interest for the treatment and control groups, taking account of both strategies. In panel A of Table 9, we observe that, on average, the treatment group (women in the private sector) presented better newborn health outcomes, a higher percentage of women with an adequate number of checkups, and early detection of pregnancy according to the epidemiological literature, as compared to the control group (women in the public sector). We can also observe that both maternal care indicators and newborn health outcomes improve during this period. In panel B, when the groups are compared according to level of adherence within the private sector, we observe that for the variables related to maternal care, the control group has better results on average than the treatment group in both periods. And when newborn health outcomes are considered, the pattern is the opposite.

Table 9 - Descriptive statistics of the outcome variables for treatment and control groups for the different strategies. Pre and post-policy period

	Pre policy				Post policy			
	Average	SD	Average	SD	Average	SD	Average	SD
	Control group		Treatment group		Control group		Treatment group	
<i>Panel A: Groups considering private and public sector</i>								
Adequate check-ups (WHO)	0.231	0.422	0.445	0.497	0.354	0.478	0.678	0.467
Adequate check-ups (MSP)	0.344	0.475	0.599	0.490	0.492	0.500	0.806	0.395
Early detection of pregnancy	0.385	0.487	0.628	0.483	0.540	0.498	0.826	0.379
Newborns' weight	3151.539	614.579	3247.440	583.782	3225.058	595.140	3327.086	521.245
Stillbirths	0.029	0.188	0.017	0.147	0.029	0.191	0.015	0.132
Low birth weight	0.108	0.310	0.075	0.264	0.093	0.290	0.051	0.220
Premature	0.101	0.301	0.080	0.272	0.103	0.304	0.079	0.270
Weeks of gestation	38.405	2.167	38.586	1.935	38.435	2.013	38.478	1.953
<i>Panel B: Groups considering level of compliance within private sector</i>								
Adequate check-ups (WHO)	0.464	0.499	0.411	0.492	0.691	0.462	0.652	0.476
Adequate check-ups (MSP)	0.626	0.484	0.552	0.497	0.823	0.382	0.773	0.419
Early detection of pregnancy	0.654	0.476	0.582	0.493	0.844	0.363	0.793	0.406
Newborns' weight	3246.316	573.373	3249.488	602.295	3311.427	527.561	3357.711	507.276
Stillbirths	0.015	0.139	0.019	0.163	0.014	0.132	0.015	0.133
Low birth weight	0.075	0.263	0.075	0.264	0.053	0.224	0.047	0.212
Premature	0.084	0.277	0.073	0.260	0.092	0.288	0.055	0.229
Weeks of gestation	38.558	2.007	38.639	1.794	38.438	2.031	38.555	1.792

Source: Calculations based on data from the Perinatal Information System, 2002-2013

As previously mentioned, the validity of the identification strategy rests on the assumption that in the absence of the policy, the treatment and control groups would have evolved similarly. We could assume that, if both groups evolved similarly and there was then a change at the point in time when the policy was implemented, this change must have been generated by the policy implementation. Graphic analysis supports this assumption by demonstrating that the groups had similar trajectories prior to treatment. In Charts A.1 to A.4 in the Annex, we observe that, although the sectors show distinct levels, the variables related to newborn health and the use of health services for both groups seem to present a similar trajectory prior to the implementation of the program.⁵

As mentioned, the degree to which health facilities adhered to the policy varied somewhat, as did the behavior of the mothers at those facilities. We therefore distinguish between individuals affected by the program and untreated individuals and consider the differences in the degree of compliance of private healthcare providers where the mothers received medical care. This strategy is known as an intention-to-treat estimate; it considers the average effect of the program on those who participated.

As a natural experiment, we exploited the differences in healthcare providers' implementation and degree of exposure to the P4P program. The period that is considered for this estimation spans from 2002 to 2013. Because we expect a delay in timing between treatment and outcomes, we evaluate the results in $t+1$.

We estimate the following equation:

$$Y_{ijk,t+1} = \alpha + \beta PMI_{jkt} + \gamma X_{ijkt} + \tau S_{kt} + \theta Z_{jkt} + \mu_j + \rho_k + \delta_t + \varepsilon_{ijkt} \quad (1)$$

⁵ The graphs show the analysis for the first strategy. The graphs for the second strategy are available upon request.

where, $Y_{ijk_{t+1}}$ are the outcome variables related to maternal care for pregnant women and newborn health i in sector j in province k at point in time $t+1$; PMI_{jkt} , is the variable that indicates whether the sector j of province k implemented the policy at point in time t ; X_{ijk_t} is the vector of control variables that capture mother's characteristics i , of the different sectors j , in the province k at point in time t ; S_{kt} is a dummy that indicates the health sector in province k at the point in time t ; Z_{jkt} is the vector that captures the percentage of households who receive the CCT program of sector j , of province k at point in time t ; μ_j controls the unobservable heterogeneity between sectors; ρ_k are binary variables that indicate the province (or geographical unit) where the mothers receive services; δ_t are binary variables indicating the year and ε_{ijk_t} is the error term. Standard errors are clustered at the individual level.

We also estimate the equations for the second strategy, in which we exploited the variability in the degree to which the private health sector was exposed to the policy. In this strategy, our treatment group is composed of pregnant women living in provinces that, on average, exceeded the compliance threshold set by the Ministry of Public Health by a deviation of 0.05. Our comparison group is made up of pregnant women living in provinces that, on average, did not exceed, by the same deviation, the compliance threshold set by the Ministry.

The parameter β captures the program's causal effect. It is assumed that there are no omitted variables that are specific to sectors that vary over time, correlated with the policy. The inclusion of fixed effects by sector allows us to control by characteristics that are constant over time. In turn, fixed effects by province and year allow us to control by characteristics shared by the treatment group and control group within each province and each year.

The specifications conducted include the mothers and the P4P's co-existence with the transfer programs by sector and provinces per year. It is important to control for certain variables, as suggested by the literature, that may have an impact on children's health outcomes (Balsa and Triunfo, 2011; Permutt and Hebel, 1989; Veloso da Veiga and Wilder, 2008;. Reichman et al., 2009). Variables related to the mothers' characteristics include: their age and quadratic specification, educational level (primary, secondary, or tertiary), marital status (married, single, divorced, or widowed), number of previous births, number of abortions and miscarriages, certain diseases (hypertension, diabetes, syphilis), drinking and smoking habits, pregnancy during winter, trimester of birth, multiple pregnancies, and the gender of the child.

It is important to emphasize that this program may be a channel for improving newborn health, through changes in the number of adequate prenatal checkups and through early detection of pregnancy.

Our results are mainly short-term results, since the program is too recent to assess its medium- and long-term economic effects. It would be useful to conduct future studies to assess whether the program generated medium and long-term benefits, and whether it has contributed to reducing healthcare costs.

As mentioned above, one issue to considered is the co-existence of the P4P with other programs that could affect our outcomes. The P4P was implemented at the same time as the healthcare reform, the PANES program, the CCT program, and a tobacco control policy. This could make it difficult to isolate the effect of the P4P. To control for this, we analyze different sets of treatment and control groups, exploiting the different sources of variation the program allows. We also exclude teenage pregnant women, because they were the group most affected by the healthcare reform. This allows us to minimize the

composition-effect problems. Last, we incorporate the percentage of households receiving the PANES and the CCT program by health sector (private and public), by province and by year, as well as a dummy that reflects whether or not the pregnant women are smokers, as control variables.

Another shortcoming of our analysis is that even if the different healthcare institutions received similar incentives related to the fulfilment of the goals, there are differences in the institutional behavior employed to achieve those goals. Several institutions conducted phone campaigns, reaching out to mothers to prompt them to get an adequate number of checkups. Others advertised inside the institutions, while still others may not have conducted any campaign because they had already achieved the goals (MSP, 2012).

4. Results

Table 10 presents the results of the estimation when we consider women of the private sector as a treatment group and women of the public sector as a comparison group. We evaluate the effect of the policy in terms of (1) the utilization of services by pregnant women excluding teenagers, and (2) newborn health outcomes, considering fixed effects by sector, region, year, and various other specifications. The variables that relate to maternal care are adequate prenatal checkups (according to WHO guidelines), adequate prenatal checkups (according to Ministry of Public Health guidelines), and early detection of pregnancy. The outcomes related to newborn health are newborn weight, low birth weight, stillbirths, number of weeks of gestation, and prematurity. The second row shows the impact of the P4P on the outcome variables we consider. As noted, in excluding teenage mothers, the analysis excludes from consideration a group of mothers that is among the most at-risk.

FE per year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.376*** (0.0229)	-0.211*** (0.0228)	-0.132*** (0.0226)	2,831*** (27.82)	0.149*** (0.0129)	-0.0459*** (0.00904)	37.99*** (0.0981)	0.148*** (0.0138)
Observations	407,361	403,270	407,361	425,037	430,663	359,822	422,231	430,663
R-squared	0.165	0.174	0.161	0.086	0.084	0.008	0.073	0.075

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

*Covariates: Age, educational level, marital status, amounts of previous births, amount of previous abortion, diseases (hypertension, diabetes), habits (alcohol and cigarettes), trimester of birth, C-section, newborns' gender, multiple pregnancies, percentage of households' CCT in each health sector, region and year.

Source: Calculations based on data from the Perinatal Information System, 2002-2013

The results in panel A show a significant increase in the number of prenatal checkups and rates of early detection of pregnancy. We can presume that this increase in prenatal care due to the P4P translates into an improvement in newborn health outcomes, with significant decreases, on average, in measures of low birth weight, premature births, and stillbirths⁶. We can evaluate the effectiveness of the policy in terms of the levels of use and access to health services, observing that the program increased the number of women who received an adequate number of prenatal checkups according to WHO guidelines (nine checkups) by 10 percentage points. According to Ministry of Public Health guidelines, the number rose around 6 percentage points. The number of women checked during their first trimester of pregnancy rose by around 4 percentage points.

Regarding newborn health outcomes, we observe that the average newborn weight increased by 11 grams. Furthermore, the percentage of babies born with low birth weight fell by 1 percentage points. We observe a decrease in the number of weeks of gestation, although we also see a decline in premature infants by almost 0.35 percentage points. Neonatal mortality fell by 2 stillbirths per 1,000 live births.

Panel B shows the estimates produced by adding different covariates. When different control covariates are used, we see that the results are quite similar in terms of sign and magnitude of the coefficients.

The following table shows the results for the case in which the treatment group are pregnant women living in provinces that, on average, exceeded the compliance threshold set by the Ministry of Public Health by a deviation of 0.05 and as a comparison group

⁶ In order to test for this link between improved pre-natal care and newborn outcomes – as suggested by one of the anonymous reviewers- we conducted some additional estimations. We conducted several Instrumental Variables (IV) estimations, by instrumenting maternal care and early detection with several specifications of our treatment. The results are positive and statistically significant for all the specifications. We did not include them here due to space limitations, but they are available upon request.

pregnant women living in provinces that, on average, not exceeded the compliance threshold set by the Ministry by the same deviation.

FE per year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.874*** (0.0391)	-0.638*** (0.0374)	-0.568*** (0.0366)	2,927*** (45.08)	0.102*** (0.0200)	-0.0165 (0.0134)	38.47*** (0.160)	0.115*** (0.0223)
Observations	223,539	221,550	223,539	232,660	236,600	186,009	230,282	236,600
R-squared	0.122	0.132	0.129	0.074	0.078	0.006	0.076	0.076

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

*Covariates: Age, educational level, marital status, amounts of previous births, amount of previous abortion, diseases (hypertension, diabetes), habits (alcohol and cigarettes), trimester of birth, C-section, newborns' gender, multiple pregnancies, percentage of households' CCT in each health sector, region and year.

Source: Calculations based on data from the Perinatal Information System, 2002-2013

In this case we observe a similar pattern as before, the results in panel A and B of Table 11, show a significant increase in the number of prenatal checkups and rates of early detection of pregnancy. The number of women receiving an adequate number of prenatal checkups according to WHO guidelines (nine checkups) rose by 3 percentage points. According to Ministry of Public Health guidelines, the number increased around 4 percentage points. The number of women checked during their first trimester of pregnancy rose by around 4 percentage points.

Regarding newborn health outcomes, we observe that the average newborn's weight increased by 39 grams. Furthermore, the percentage of babies born with low birth weight fell by 0.5 percentage points. Although the number of weeks of gestation is not statistically significant, we observed a decrease of almost 2.5 percentage points in the number premature births. Neonatal mortality fell by 2.8 stillbirths per 1,000 live births. In panel B, when we control for different covariates, we observe that the number of weeks of gestation increased on average and that the rest of the results are quite similar.

The next table shows the intensity of the treatment in relation to the percentage of compliance of the different healthcare providers. In this case, again, the results are along the same lines as the previous analysis.

Control Variables	No	No	No	No	No	No	No	No
Constant	-1.328*** (0.0477)	-0.792*** (0.0467)	-0.697*** (0.0457)	2,964*** (55.99)	0.0933*** (0.0247)	-0.0310* (0.0159)	38.58*** (0.195)	0.0865*** (0.0278)
Observations	223,539	221,550	223,539	232,660	236,600	186,009	230,282	236,600
R-squared	0.122	0.132	0.129	0.074	0.078	0.006	0.076	0.076

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

Source: Calculations based on data from the Perinatal Information System, 2002-2013

4.1. Robustness checks

The DiD method requires certain assumptions to be satisfied in order to claim that our estimates are causal. The model uses changes in the results of the control group to estimate the counterfactual that shows how the treatment group would have evolved if the program had not been implemented. The most important identification assumption, therefore, is that changes in the control group's results serve as a valid counterfactual of what would have happened to the treatment group if the policy had not been applied.

We need to address several sources of confoundedness in our results. To ensure the validity of the outcomes, we run several robustness checks. First, we compared the differences between the private and public health sectors outside the main capital, Montevideo. We selected sectors outside Montevideo for two reasons: (1) the beneficiaries of the health system outside the capital more closely resemble each other than do the beneficiaries in Montevideo; and (2) deciding whether going to public vis a vis private health centers is less associated with population characteristics in the countryside due to the higher quality of the public sector outside of Montevideo. Second, as we mentioned before, at the beginning of the program (the first trimester) the institutions implemented a range of measures to reach the aims. In general, the institutions in this first step established goals close to the percentage of the pregnant women with proper care. When the program began, therefore, some institutions in the treatment group became part of the control group when the ministry established the goals and some in the comparison group that became part of the treatment group. We thus exploited this first variation in policy adherence when the program began. Third, we consider the intensity of the treatment in terms of the percentage of private clinics into the provinces. And last, we consider only the private healthcare providers, taking as comparison group those institutions that in the pre-P4P period had better health outcome

results (assuming that the program did not affect them) and as treatment group those institutions with poorer results.

In our first robustness check, in which we consider the private and public healthcare providers outside the main capital, we observe positive and significant effects on the outcome variables considered, with a magnitude that is comparable to that found for the different specifications (Table A.1 of the statistical annex). We can safely assume that Montevideo does not exhibit any differential pattern in terms of visiting health facilities. Our results are not driven from the results observed in Montevideo. In our second check, in which different healthcare providers implemented different measures to reach the program's objectives, we observe a lower effect than the one found in our main results (Table A.2, panel A). When we control for individual characteristics, we do not find any effect on maternal care and on the variables related to newborn weight (Table A.2, panel B). A lower level of effort to reach the objectives may generate fewer changes in the utilization rates of pregnant women.

One of the main threats to our identification strategy is that the people who use the private sector are different from the people who use public hospitals. There are two possible effects to consider in this regard. The first is that the first group might be more responsive to the policy, since they are better educated and more affluent, which would give our estimates an upward bias. The second effect is that the first group in the pre-P4P period had better health outcomes than the control group, which would lead us to expect lower levels of response to the policy. In that case, our estimates would be a downward biased. As previously mentioned, Table shows 9 that the percentage of women of the treatment group with an adequate number of checkups, according Ministry requirements, was 44% and 68% in the pre- and post-policy period. The percentages in the control group was 23% and 35% respectively.

In this context, the third exercise thus reinforces our estimations, enabling us to better understand the effect generated from the heterogeneity of the groups. Table A.3 shows the intensity of the treatment in relation to the percentage of private clinics into the provinces.⁷ The results obtained are quite similar when we compare them with the main outcomes. However, the magnitude of the coefficients of maternal care, when the percentage-of-private-clinics variable is considered, is bigger, which leads us to think that women in the private clinics are more responsive than the women in the public ones.

Finally, Table A.4 shows only the private healthcare providers, taking as comparison group those institutions that in the pre-P4P period had better health outcome results and as treatment group those institutions with poorer results. The results are lower than our main outcomes. This result leads us to expect lower levels of response to the policy, so that our estimates would be a lower bound.

5. Discussion of the results

Table 13 summarizes the results of the program. In the first column, we exhibit the results with pregnant women in the private sector considered as treatment group and women of the public sector as control group. The third column displays the average of the control group for each outcome variable. The fourth column shows the results when compliance within the private sector is considered. Finally, in the last column we observe the average for the control group.

Table 13 – Summary of the P4P results

VARIABLES	1 st . Specification.		2 nd . Specification	
	Results	Average control group	Results	Average control group
Observations		194063		155710

⁷ We interact this variable with the treatment variable.

Adequate Check-ups (WHO)	10 pp	0.354	2.5 pp	0.691
Adequate Check-ups (MSP)	6 pp	0.492	3.5 pp	0.823
Early detection	4.5 pp	0.540	3 pp	0.844
New-borns' weight	11 grams	3225.058	37 grams	3311.427
Low birth weight	-1.2 pp	0.093	-0,5 pp	0.053
Stillbirth	-2.4 per 1000 born live	0.029	3.2 per 1000 born live	0.014
Weeks of gestation	-0,144 weeks	38.435	0.02 weeks	38.438
Prematurity	-0,36 pp	0.103	-2.5 pp	0.092

Source: Calculations based on data from the Perinatal Information System, 2002-2013

For the maternal care outcomes, the first specification implies that adequate prenatal checkups, as laid out in WHO guidelines, rise from 34% to 45%. If we consider prenatal checkups, established in Ministry of Public Health guidelines, it implies an increase of 6 percentage points (rising from 49% to 55%). Finally, the early detection of pregnancy increases from 54% to 58.5%. When we observe the second specification, the outcomes related to maternal care, are smaller. The effect on the average of newborn health is higher for the second specification, but it is lower when we consider the impact on low birth weight. Estimates for the first case thus decrease from 9% to 8% and, for the second specification, go down from 5% to 4.5%. Last, the impact on stillbirth and prematurity is higher in the second case, when differences within the private sector are considered.

Overall, results are positive and encouraging and consistent with what is observed in some other studies. The range of the increase in birth weight for example is important. To put the improvement in perspective and in a complete different context, Bozzoli and Quintana-Domeque (2014) found a decrease in average birth weight of 30g as a consequence of a deep economic crisis in Argentina, where GDP contracted more than 10%.

Our results are, in some cases, in line with previous studies, but they differ in other ways. Balsa and Triunfo (2015) analyzed the relationship between the prenatal care of low-income pregnant women and newborn outcomes in Uruguay. They found that an

adequate number of prenatal checkups (at least nine), along with early detection of pregnancy, reduced the likelihood of low birth weight by 6 percentage points and the probability of births occurring before 37 weeks of gestation by 11 percentage points. They also found an increase in newborn weight by 149 grams. Our results are smaller in magnitude. Celhay et al. (2019) analyzed how temporary economic incentives offered to healthcare providers through Argentina's *Plan Nacer* to encourage the early detection of pregnancy affected this variable and the results related to health outcomes. In line with our results, they showed that the program increased the early detection of pregnancy by 34 percent in the treatment period and persisted 24 months after the program ended. The results for health outcomes are not consistent with our results since they did not find any effects. However, Celhay et al. (2019) point out that they may lack statistical power to find effects on some of the health outcomes. Amarante et al. (2016) estimated the effect of a cash transfer program (PANES) on children's birth outcomes found a decline in low birth weight of 1.9 to 2.4 percentage points. But they did not find evidence of any increase in prenatal care. In sum, our results are smaller than the previous works.

Although the results of the causal effects have internal validity, we cannot confirm their external validity because of the specific characteristics of the program. Given this, we should be cautious when extrapolating conclusions from these results. **The program may have worked well in a small country such as Uruguay, a country that enjoys very low level of corruption compared to other developing countries.**

These types of programs might be an attractive tool for governments, for inducing healthcare providers to improve prenatal care services' quality and to achieve better results on newborns' health outcomes. As mentioned before, there is very limited evidence in Latin America on a program introducing incentives at the country level and for the whole population. Plan Nacer, implemented in Argentina has a different

population coverage and also shower positive results in terms of increasing maternal care indicators.

6. Final considerations

Uruguay's *Metas Asistenciales* is a nationwide P4P program that seeks to improve prenatal care and, through it, newborn health outcomes, by providing monetary incentives to healthcare providers. The program is part of the healthcare reform implemented in 2008, which attempts to move the health services towards a model of primary healthcare coverage. Since its implementation in 2008, the program has continued to reinforce policies that improve access to health services for pregnant women and children. The novelty of the program is the way health centers are financed, which is an innovation in terms of health service delivery in developing countries. The program offers a pay-for-performance (P4P) incentive, which provided funds to health facilities that adhere to compliance standards set by the Ministry of Public Health. Health facilities receive funding for complying with certain indicators measured in the target population. If the standards are met, the centers are paid a certain amount per patient. This incentive may represent a fundamental channel for improving children's health outcomes.

We evaluated the effect of the incentive scheme on the utilization rates of maternal care services and newborn health outcomes. Given the non-experimental nature of the policy, we performed a DiD estimation to compare the outcomes of variables of interest among populations treated in health centers against those that were not treated. We use two main sources of variation to identify possible causal effects. The first is the gap between private and public health sectors in the implementation of incentives. This difference was unintended. It was caused by a standing regulation stating that the financial incentive

granted to public sector institutions had to be discounted from their allocated budgets. Second, we consider the difference in compliance within the private sector as a second source of variation.

We observe a significant improvement in prenatal care, which increased by 10 percentage points, while the level of early detection of pregnancy increased by 4.5 percentage points. Women receiving services in a treated institution had a lower likelihood of having a newborn with low birth weight (1.2 percentage points). Premature births decreased by 0.36 percentage points. In turn, we observed a decrease in neonatal mortality by 2.4 per 1,000 births. Although the results of the causal effects have internal validity, we cannot confirm their external validity because of the specific characteristics of the program. Given this, we should be cautious when extrapolating conclusions from these results. Finally, robustness checks confirmed that the results obtained are robust.

This study thus provides relevant evidence of the potential benefits of a P4P scheme put in place to improve maternal and birth outcomes. Our results are, to the best of our knowledge, the first to use national data and administrative statistics for a nationwide policy implementing P4P schemes in an effort to improve newborn health in a developing country. We must stress, however, that the policy should be fully implemented in the public health sector in order to serve the most vulnerable population. Our estimates represent a lower bound and we can conjecture they will be of greater magnitude when considering at-risk populations such as teenage mothers who are normally served by the public health sector. It will be important to examine whether greater compliance generates a differential effect.

Ultimately, however, we can conclude that this Uruguayan P4P program had a positive and statistically significant impact on prenatal care and on newborn health outcomes.

This finding provides evidence in support of the idea that monetary incentives are a promising model that should be considered for regional and international expansion.

7. References

Alderman H. and Behrman J.R. Reducing the Incidence of Low Birthweight in Low-Income Countries Has Substantial Economic Benefits. *World Bank Research Observer* 2006; **21**: 1 - 25-48.

Almond, D., Hoynes, H., and Schanzenbach, D. Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes. *Review of Economics and Statistics* 93 (2): 387–403. 2011.

Amarante, V, Manacorda, M., Miguel, E., and Vigorito, A. Do Cash Transfers Improve Birth Outcomes? Evidence from Matched Vital Statistics, Program, and Social Security Data. *American Economic Journal: Economic Policy*, 8 (2):1-43. 2016

Balsa A., and Triunfo P. ¿Son los cuidados prenatales efectivo? Un enfoque con datos individuales de panel. *Documento de Trabajo 06/12. Departamento de Economía, Universidad de la República, Uruguay* 2011.

Basinga P., Gertler P., Binagwaho A., Soucat A., Sturdy J., and Vermeersch, C. Paying Primary Health Care Centres for Performance in Rwanda. *Policy Research Working Paper, World Bank*. 2010; **5190**.

Behrman J.R. and Rosenzweig M.R. Returns to Birth weight. *Review of Economics and Statistics* 2004; **86**: 2 - 586–601.

Bergolo M., and Cruces G. Work and tax evasion incentive effects of social insurance programs. Evidence from an employment based benefit extension. *Journal of Public Economics* 2014; **117**: 211-228

Berwick, D.M. Disseminating Innovations in Health Care. *JAMA* 289 (15): 1969–75. 2003.

Black S.E., Devereux P. J. and Salvanes K.G. From the Cradle to the Labour Market? The Effect of Birth Weight on Adult Outcomes. *Quarterly Journal of Economics* 2007; **122**: 1 - 409–39.

Boardman J.D., Powers D.A., Padilla Y.C., Hummer R.A. Low birth weight, social factors, and developmental outcomes among children in the United States. *Demography* 2002; **39**: 2 - 353–363.

Bozzoli, C and C Quintana-Domeque, The weight of the crisis: Evidence from newborns in Argentina *Review of Economics, and Statistics* 2014; **96** (3), 550-562

Cabana, M.D., Rand, C., Powe, N., Wu, A., Wilson, M., Abboud, P., and Rubin, H. Why Don't Physicians Follow Clinical Practice Guidelines? A Framework for Improvement. *JAMA* 282 (15): 1458–65. 1999.

Capurro A., and Recchi V. Bajo peso al nacer. Factores determinantes. *Archivos de Ginecología y Obstetricia* 1994; **33**:1-35-42.

Carroli, G., Rooney, C., and Villar, J. How Effective Is Antenatal Care in Preventing Maternal Mortality and Serious Morbidity? An Overview of the Evidence. *Pediatric and Perinatal Epidemiology* 15 (s1): 1–42. 2001.

Celhay, P, Gertler, P., Giovagnoli, P. and Vermeersch, C. Long-Run Effects of Temporary Incentives on Medical Care Productivity. *American Economic Journal: Applied Economics*. 11(3): 92–127. 2019.

Cunha, F., Heckman, J. J., and Lochner, L. Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education, Elsevier*. 2006

Currie J. Biology as Destiny? Short and Long-Run Determinants of Intergenerational Transmission of Birthweight. *Journal of Labour Economics* 2007; **25**:231-264.

De Walque D., Gertler P., Baitista-Arredondo S., Kwan A., Vermeesch C., de Dieu Bizimina J., Binagwaho A., and Condo J. Using Provider Performance Rewards to Increase HIV testing and Counseling Services in Rwanda. *World Bank Policy Research Working Paper n° 6364, Impact Evaluation Series 84, Washington DC: The World Bank* 2013.

Di Cesare, M., and Sabates, R. Access to antenatal care and children's cognitive development: A comparative analysis in Ethiopia, Peru, Vietnam and India. *International Journal in Public Health*, 58(3), 459-467. 2017.

Eldridge C. and Palmer N. Performance-based payment: some re_ections on the discourse evidence and unanswered questions, *Health Policy and Planning* 2009; **24**:160- 166.

Evans, W.N., Lien, D.S. The Benefits of Prenatal Care: Evidence from the PAT Bus Strike. *Journal of Econometrics* 125 (1–2): 207–39. 2005.

Garriga S. Impacto del Plan Nacer sobre la Mortalidad Infantil en Argentina. *Tesis de Maestría, Maestría en Economía. Universidad Nacional de La Plata* 2012.

Gertler P., and Vermeersch C. Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes. *NBER* 2013.

Gertler P., Giovagnoli P. and Martínez S. Rewarding Provider Performance to enable a healthy start to life. Evidence from Argentina's Plan Nacer. *Policy Research Working Paper. The World Bank. Latin America and the Caribbean Region Health, Nutrition and Population Unit* 2014; **6884**.

Harris J., Balsa A. and Triunfo P. Tobacco control campaign in Uruguay: Impact on smoking cessation during pregnancy and birth weight. *Journal of Health Economics*, Vol.42: 186-196. 2015.

Heckman J. Lessons from the Bell Curve. *Journal of Political Economy* 1995; **103**: 51 - 1051- 1120.

- Heckman J. Policies to Foster Human Capital. *Research in Economics* 2000; **54**:1- 3-56.
- Heckman, J. The Economics, Technology and Neuroscience of Human Capability Formation. *National Bureau of Economic Research*. WP. 13195. 2007
- Heckman, J., Pinto, R., and Savelyev, P. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *The American Economic Review*, 103(6), 2052-2086. 2013
- Huillery E. and Seban. J. Performance based financing, motivation and final output in the health sector: experimental evidence from the democratic republic of Congo. *Discussion paper, Department of Economics. Sciences Po, Paris* 2014; **12**.
- Instituto Nacional de Estadística. *Datos de la Encuesta Continua de Hogares* 2014. Available at <http://www.ine.gub.uy/encuesta-continua-de-hogares1>
- Junta Nacional de Salud (JUNASA). *Rendición de Cuentas, Ejercicio 2009. Montevideo, Uruguay, Ministerio de Salud Pública. Versión preliminar* 2010.
- Kramer, M.S. Nutritional Advice in Pregnancy. *Cochrane Database of Systematic Reviews*. 1996.
- Messen B., Kashala J., and Musango L. Output-based payment to boost staff productivity in public health centres: Contracting in Kaburate district, Rwanda. *Bulletin of World Health Organization*, 2007; Vol.**85**: 2 - 108-115.
- Ministerio de Salud Pública (MSP). División de Economía de la Salud. *Instructivo de metas asistenciales*; 2010.
- Ministerio de Salud Pública and World Health Organization. Avances en la consolidación del Sistema Nacional Integrado de Salud. Available: https://www.paho.org/uru/index.php?option=com_docman&view=download&alias=673-avances-en-la-consolidacion-del-snis-baja&category_slug=publications&Itemid=307. 2019.
- Mora R., and Reggio I. Treatment effect identification using alternative path assumptions. *Departamento de Economía, Universidad Carlos III, Madrid* 2012.
- Olken B, Onishi J., and Wong S. Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia. *Working paper, MIT, Department of Economics, Cambridge, MA* 2013.
- Permutt T., Hebel J.R. Simultaneous-Equation Estimation in a Clinical Trial of the Effect of Smoking on Birthweight. *Biometrics* 1989; **45**:619-622.
- Regalia F., and Castro L. Performance-based incentives for health: demand-and supply-side incentives in the Nicaraguan Red de Protección Social. *Center for Global Development Working Paper* 2007.

Reichman N.E., Corman H., Noonan K., and Dave D. Infant health production functions: What a difference the data make? *Health Economics* 2009; **18**:7 - 761-782.

Rosenzweig, M.R., and Schultz, T.P. Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight. *Journal of Political Economy* 91 (5): 723–46. 1983.

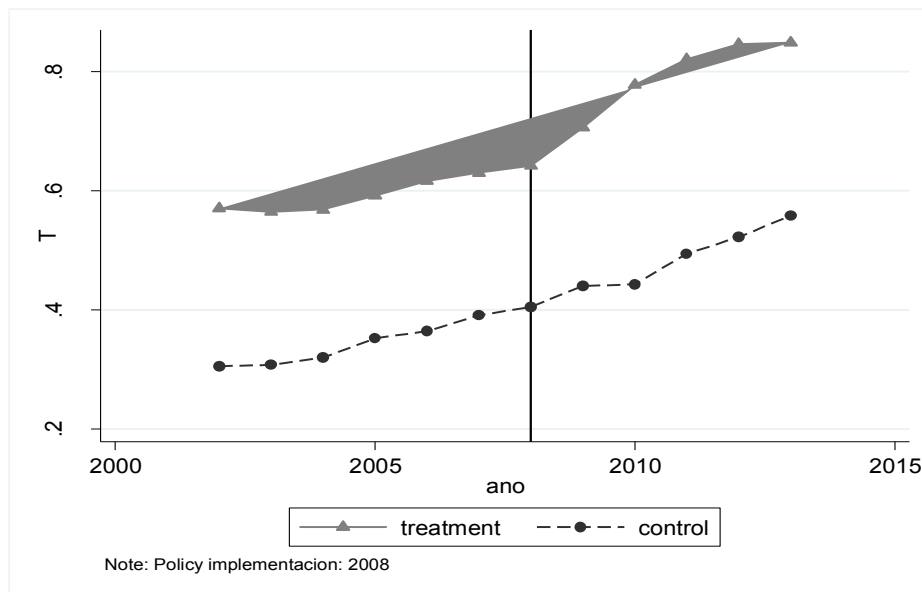
Royer H. Separated at Girth: U.S. Twin Estimates of the Effects of Birthweight. *American Economic Journal: Applied Economics* 2009; **1**:1 - 49-85.

Veloso da Veiga P., and Wilder R. Maternal Smoking During Pregnancy and Birthweight: A Propensity Score Matching Approach. *Maternal Child Health Journal* 2008; **12**:194–203.

Williamson, N. Maternidad en la niñez. Enfrentar el reto del embarazo en adolescentes. *El Estado de la Población Mundial 2013, informe producido por la División de Información y Relaciones Externas del UNFPA, Fondo de Población de las Naciones Unidas*

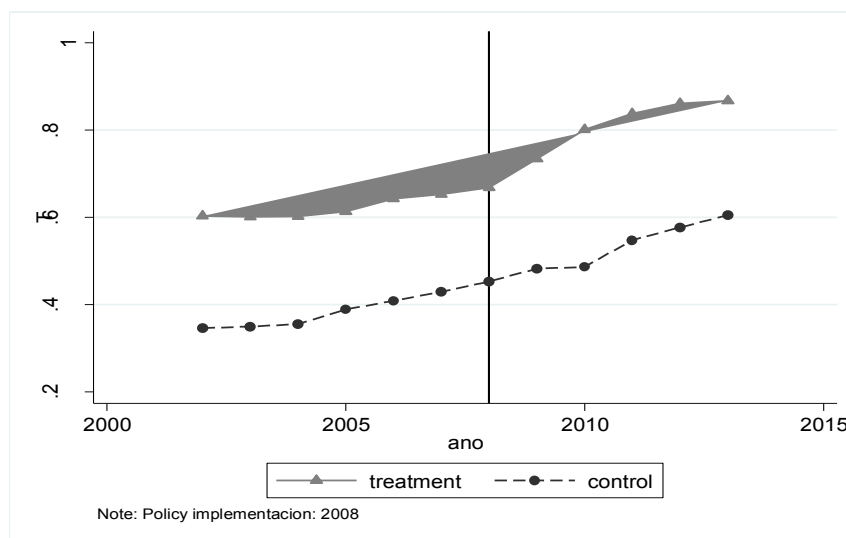
Statistical annex

Figure A.1 – Trend of the number of the adequate prenatal check-ups within the treatment and control groups



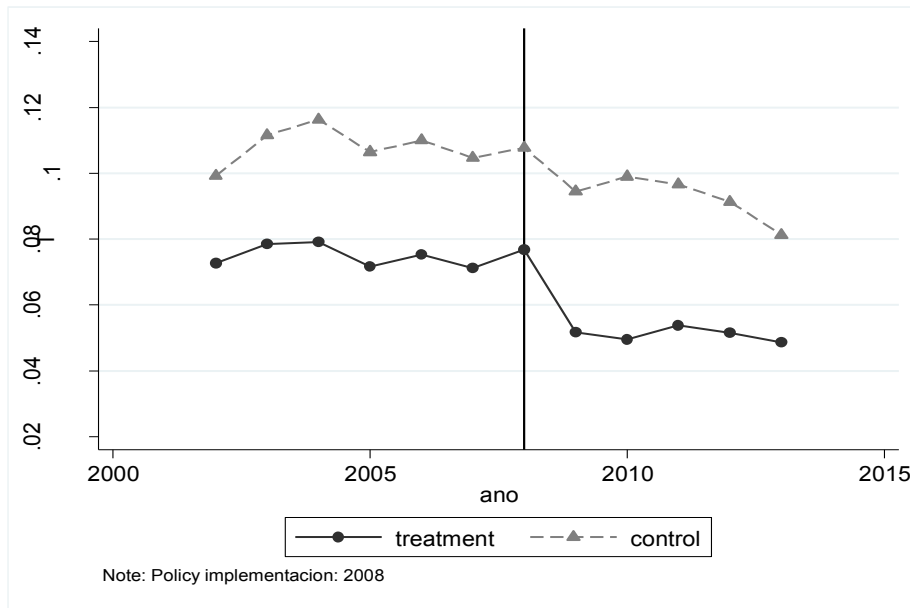
Source: Data from the Perinatal Information System, 2002-2013

Figure A.2 – Trend of early detection of pregnancy



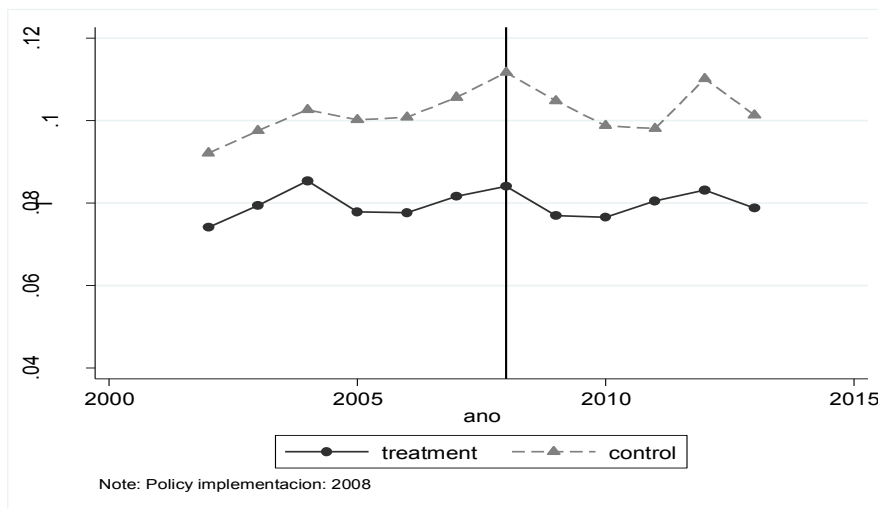
Source: Data from the Perinatal Information System, 2002-2013

Figure A.3 – Trend of the average of low birth weight, within the treatment and control groups



Source: Data from the Perinatal Information System, 2002-2013

Figure A.4 – Trend of the amount of premature babies, within the treatment and control groups



Source: Data from the Perinatal Information System, 2002-2013

Table A.1 – Robustness checks. Private and public health providers outside the main capital

<i>Panel A</i>								
VARIABLES	(1) Adequate prenatal check-ups (WHO)	(2) Adequate prenatal check-ups (MSP)	(3) Early detection of pregnancy	(4) Newborn weight	(5) Low birth weight	(6) Stillbirths	(7) Weeks of gestation	(8) Prematurity
Treatment	0.157*** (0.00304)	0.183*** (0.00332)	0.175*** (0.00334)	35.04*** (3.864)	-0.000117 (0.00173)	- 0.00706*** (0.00133)	-0.0318** (0.0126)	0.00103 (0.00176)
Treatment*Post_Policy	0.109*** (0.00458)	0.0589*** (0.00475)	0.0440*** (0.00471)	54.91*** (5.522)	-0.0249*** (0.00248)	-0.00465** (0.00182)	0.0700*** (0.0184)	-0.0378*** (0.00270)
FE per province	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE per year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	No	No	No	No	No	No	No	No
Constant	0.134*** (0.00512)	0.285*** (0.00567)	0.334*** (0.00573)	3,182*** (7.448)	0.0932*** (0.00338)	0.0272*** (0.00223)	38.73*** (0.0251)	0.0850*** (0.00339)
Observations	182,127	180,174	182,127	187,238	189,129	157,817	186,196	189,129
R-squared	0.124	0.115	0.106	0.011	0.004	0.001	0.006	0.005
<i>Panel B</i>								
Treatment	0.0800*** (0.00326)	0.103*** (0.00354)	0.0992*** (0.00356)	9.076** (4.035)	-0.00108 (0.00179)	- 0.00706*** (0.00142)	0.00823 (0.0132)	-0.00210 (0.00186)
Treatment*Post_Policy	0.0777*** (0.00595)	0.0550*** (0.00611)	0.0465*** (0.00605)	39.99*** (6.857)	-0.0269*** (0.00306)	-0.00335 (0.00234)	0.0414* (0.0232)	-0.0296*** (0.00343)

FE per year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	No	No	No	No	No	No	No	No	No
Constant	0.220*** (0.0100)	0.453*** (0.0103)	0.486*** (0.0101)	3,259*** (11.99)	0.0727*** (0.00522)	0.0141*** (0.00276)	38.99*** (0.0405)	0.0746*** (0.00604)	
Observations	223,539	221,550	223,539	232,660	236,600	186,009	230,282	236,600	
R-squared	0.080	0.081	0.079	0.009	0.003	0.001	0.004	0.003	

Panel B

Treatment	0.262*** (0.0121)	0.101*** (0.0122)	0.0839*** (0.0120)	-1.558 (14.69)	0.00215 (0.00649)	0.00796* (0.00462)	-0.0674 (0.0490)	0.00764 (0.00729)	
Treatment*Post_Policy	0.00759 (0.00517)	0.00434 (0.00482)	0.00612 (0.00473)	0.0151 (5.855)	-0.000437 (0.00249)	-0.00327* (0.00188)	0.0404** (0.0195)	-0.0106*** (0.00275)	
FE per province	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE per year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.508*** (0.0311)	-0.289*** (0.0294)	-0.231*** (0.0287)	3,013*** (34.74)	0.0843*** (0.0152)	-0.0106 (0.0105)	38.68*** (0.122)	0.102*** (0.0171)	
Observations	223,539	221,550	223,539	232,660	236,600	186,009	230,282	236,600	
R-squared	0.122	0.131	0.128	0.073	0.078	0.006	0.076	0.076	

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Calculations based on data from the Perinatal Information System, 2002-2013

Table A.3 – Robustness controls. Intensity of the treatment considering the percentage of private clinics into the provinces

<i>Panel A</i>								
VARIABLES	(1) Adequate prenatal check- ups (WHO)	(2) Adequate prenatal check- ups (MSP)	(3) Early detection of pregnancy	(4) Newborn weight	(5) Low birth weight	(6) Stillbirths	(7) Weeks of gestation	(8) Premat
Treatment	0.217*** (0.00199)	0.260*** (0.00208)	0.247*** (0.00207)	89.21*** (2.532)	-0.0292*** (0.00119)	-0.0125*** (0.000762)	0.159*** (0.00873)	-0.018*** (0.00119)
Intensity	1.640*** (0.0746)	0.432*** (0.0782)	0.230*** (0.0783)	69.84 (97.32)	-0.0636 (0.0446)	0.0300 (0.0321)	1.859*** (0.327)	-0.079*** (0.046)
Intensity*treatment	0.143*** (0.00405)	0.0839*** (0.00404)	0.0614*** (0.00398)	22.03*** (5.006)	-0.0206*** (0.00236)	-0.00280* (0.00154)	-0.132*** (0.0179)	-0.001*** (0.002)
FE per province	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE per year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	No	No	No	No	No	No	No	No
Constant	-0.695*** (0.0401)	0.0689 (0.0424)	0.217*** (0.0425)	3,134*** (53.07)	0.130*** (0.0243)	0.0128 (0.0172)	37.71*** (0.179)	0.131*** (0.025)
Observations	407,361	403,270	407,361	425,037	430,663	359,822	422,231	430,663
R-squared	0.131	0.137	0.127	0.018	0.009	0.002	0.007	0.007
<i>Panel B</i>								
Treatment	0.123*** (0.00223)	0.163*** (0.00234)	0.158*** (0.00234)	54.25*** (2.764)	-0.0265*** (0.00128)	-0.0129*** (0.000885)	0.191*** (0.00964)	-0.0213*** (0.00119)
Intensity	1.934***	0.753***	0.527***	169.1*	-0.0550	0.0248	1.521***	-0.03***

	(0.0728)	(0.0763)	(0.0765)	(94.66)	(0.0435)	(0.0320)	(0.320)	(0.045)
Intensity*treatment	0.136*** (0.00439)	0.0894*** (0.00441)	0.0711*** (0.00437)	20.07*** (5.244)	-0.0230*** (0.00249)	-0.00394** (0.00178)	-0.145*** (0.0189)	-0.0007*** (0.002)
FE per province	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE per year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-1.340*** (0.0450)	-0.586*** (0.0469)	-0.394*** (0.0469)	2,746*** (58.06)	0.176*** (0.0267)	-0.0583*** (0.0191)	37.22*** (0.198)	0.165*** (0.027)
Observations	407,361	403,270	407,361	425,037	430,663	359,822	422,231	430,663
R-squared	0.165	0.174	0.162	0.086	0.084	0.008	0.073	0.073
Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1								

Source: Calculations based on data from the Perinatal Information System, 2002-2013

Table A.4 – Robustness controls. Private health providers: comparison group institutions on the pre period policy with better health outcomes and treatment group institutions with worst results.

<i>Panel A</i>								
VARIABLES	(1) Adequate prenatal check-ups (WHO)	(2) Adequate prenatal check-ups (MSP)	(3) Early detection of pregnancy	(4) Newborn weight	(5) Low birth weight	(6) Stillbirths	(7) Weeks of gestation	(8) Prematurity
Treatment	0.000930	0.00138	0.000780	-8.617***	0.00249**	0.000358	-0.0173*	0.00401***

	(0.00219)	(0.00204)	(0.00199)	(2.529)	(0.00111)	(0.000714)	(0.00890)	(0.00124)
Treatment*Post_Policy	0.00849*	0.0114***	0.0102***	23.51***	-0.00248	-0.00249*	0.00592	-0.0108***
	(0.00437)	(0.00404)	(0.00393)	(4.906)	(0.00212)	(0.00137)	(0.0169)	(0.00227)
FE per province	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE per year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	No	No	No	No	No	No	No	No
Constant	0.218***	0.452***	0.485***	3,265***	0.0714***	0.0140***	39.00***	0.0725***
	(0.0101)	(0.0104)	(0.0102)	(12.07)	(0.00526)	(0.00278)	(0.0408)	(0.00609)
Observations	223,539	221,550	223,539	232,660	236,600	186,009	230,282	236,600
R-squared	0.080	0.081	0.079	0.009	0.003	0.001	0.004	0.003

Panel B

Treatment	0.000762	0.00135	0.000662	-8.001***	0.00242**	0.000328	-0.0181**	0.00406***
	(0.00214)	(0.00199)	(0.00193)	(2.440)	(0.00107)	(0.000712)	(0.00857)	(0.00119)
Treatment*Post_Policy	0.00608	0.00817**	0.00708*	22.93***	-0.00266	-0.00274**	0.0129	-0.0112***
	(0.00428)	(0.00393)	(0.00383)	(4.768)	(0.00206)	(0.00137)	(0.0164)	(0.00219)
FE per province	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE per year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.508***	-0.289***	-0.231***	3,019***	0.0829***	-0.0108	38.69***	0.0997***
	(0.0311)	(0.0294)	(0.0287)	(34.77)	(0.0153)	(0.0105)	(0.122)	(0.0171)
Observations	223,539	221,550	223,539	232,660	236,600	186,009	230,282	236,600
R-squared	0.122	0.131	0.128	0.073	0.078	0.006	0.076	0.076

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Source: Calculations based on data from the Perinatal Information System, 2002-2013