

Long-Run Effects of Temporary Incentives on Medical Care Productivity

Pablo Celhay
Paul Gertler
Paula Giovagnoli
Christel Vermeersch



WORLD BANK GROUP

Health Nutrition and Population Global Practice Group

June 2015

Abstract

The adoption of new clinical practice patterns by medical care providers is often challenging, even when the patterns are believed to be efficacious and profitable. This paper uses a randomized field experiment to examine the effects of temporary financial incentives paid to medical care clinics for the initiation of prenatal care in the first trimester of pregnancy. The rate of early initiation of prenatal care was 34 percent higher in the treatment group than in the control

group while the incentives were being paid, and this effect persisted at least 15 months and likely 24 months or more after the incentives ended. These results are consistent with a model where the incentives enable providers to address the fixed costs of overcoming organizational inertia in innovation, and suggest that temporary incentives may be effective at motivating improvements in long-run provider performance at a substantially lower cost than permanent incentives.

This paper is a product of the Health Nutrition and Population Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at cvermeersch@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Long-Run Effects of Temporary Incentives on Medical Care Productivity

Pablo Celhay
Paul Gertler
Paula Giovagnoli
Christel Vermeersch

JEL Classification: I12, I13, I15, I18

Keywords: Keywords: Pay-for-performance, results-based financing, provider performance, birth outcomes, impact evaluation, maternal and child health, organizational inertia, temporary incentives

Author Affiliation: Pablo Celhay (pcelhay@uchicago.edu) is a Ph.D. candidate at the University of Chicago. Paul Gertler (Gertler@haas.berkeley.edu) is the Li Ka Shing Professor at the University of California, Berkeley. Paula Giovagnoli (pgiovagnoli@worldbank.org) is an Economist with the World Bank. Christel Vermeersch (cvermeersch@worldbank.org) is a Senior Economist at the World Bank.

Acknowledgements: The experiment described in this paper was developed under the leadership of Martin Sabignoso, National Coordinator of Plan Nacer and Humberto Silva, National Head of Strategic Planning of Plan Nacer, Ministry of Health, Argentina. Together with the national team, Luis Lopez Torres and Bettina Petrella from the Misiones Office of Plan Nacer oversaw the implementation of the pilot, facilitated access to provincial data, supported the authors in interpreting datasets and the provincial legal framework and in carrying out the in-depth interviews. Fernando Bazán Torres, Ramiro Flores Cruz, Santiago Garriga, Alfredo Palacios, Rafael Ramirez, Silvestre Rios Centeno, Gabriela Moreno, and Adam Ross provided excellent assistance and project management support. Alvaro S. Ocariz, Javier Minsky and the staff of the Information Technology unit at Central Implementation Unit (UEC) at the Ministry of Health provided valuable support in identifying sources of data. The authors acknowledge the contributions of Sebastian Martinez, Luis Perez Campoy, Vanina Camporeale and Daniela Romero in the initial design of the pilot. The authors also thank Ned Augenblick, Dan Black, Nick Bloom, Megan Busse, Stefano DellaVigna, Damien de Walque, Emanuela Galasso, Jeff Grogger, Petra Vergeer, as well as participants in seminars at UC Berkeley, Northwestern University and Chicago University for helpful comments. The authors gratefully acknowledge financial support from the Health Results Innovation Trust Fund (HRITF) and the Strategic Impact Evaluation Fund (SIEF) of the World Bank. The authors declare that they have no financial or material interests in the results of this paper.

1 INTRODUCTION

Successful organizations are able to efficiently and reliably produce high quality products through the use of reproducible and stable routines.¹ Routines shape the production process by defining each person's role and their patterns of action, and by coordinating the tasks performed by the different team members.² They can be thought of as organizational habits that reduce the complexity of decision-making, facilitate coordination across team members, and speed production. However, once established, routines are costly to change. The cost of adjustment includes the time and money needed to retool routines, an adjustment period in which production is less reliable while the new routines are being learned, and possibly psychological resistance to change. As a result, organizations tend to be resistant to adopting structural changes that are thought to be productive and profitable (Hannan and Freeman 1984; Carroll and Hannan 2000). While organizational routines are necessary for efficient and reliable production, they can result in organizational inertia to innovation.

Nowhere are organizational routines more important than in the production of medical care services (Hoff 2014). Medical care entails coordinating a large, complex set of tasks such as deciding what information to collect from the patient, assessing social and medical risks, deciding what diagnostic tests to prescribe, interpreting symptoms and test results, and prescribing and implementing treatments.³ Typically, a team coordinated by a physician implements these tasks. Nurses often take medical and social histories, conduct preliminary physical exams, and administer injections. Laboratory technicians analyze blood and urine. Pharmacists dispense drugs and monitor negative drug interactions. Physical and occupational therapists provide rehabilitation services. Community health-workers provide outreach, promotion and preventive services, and follow-up care to patients. Clinics establish practice routines that are consistent with their training and experience to standardize and coordinate care.

There is substantial evidence of organizational inertia in medical care as indicated by the remarkably low level of compliance with Clinical Practice Guidelines (CPGs) worldwide (Figure 1).

¹ Organizational routine has been studied extensively since popularized by Nelson and Winter (1982). In a review of the literature Becker (2004) defines routines as “recurrent interaction patterns” within an organization, or as “established rules, or standard operating procedures”.

² Often relationships between team members and management are enforced by informal relational contracts (Gibbons and Henderson 2012 and 2013).

³ Complex production technologies with sophisticated routines such as medical care require strong management to be efficient and productive. Bloom et al. (2014) provide evidence that better management increases public hospital productivity.

CPGs define medical care production possibility frontiers in that they prescribe the clinical content of care that maximizes the likelihood of successful health outcomes based on medical science, clinical trials, and practitioner consensus. Local CPGs are regularly updated and serve as the basis of training in medical schools and practitioner refresher courses. While the lack of compliance with CPGs may in part reflect a lack of knowledge, evidence shows that practitioners often provide a standard of care well below their level of knowledge of CPGs.⁴ In a systematic review of the literature on reasons for non-compliance of CPGs, Cabana et al. (1999) report that resistance to changing existing practice patterns is one of the most important barriers to CPG adherence. For example, Grol and Grimshaw (2003) surveyed nurses and doctors in the UK about the adoption of new hand hygiene guidelines. Forty-nine percent responded that resistance to changing old routines was an obstacle to complying with new guidelines.⁵

Changing deep-rooted habits is hard and even small costs of adjustment may inhibit changes in favor of maintaining the status quo, (DellaVigna 2009; Thaler and Sunstein 2009).⁶ In these circumstances, temporary incentives may speed adoption by helping to compensate providers for the initial fixed costs of changing their practice pattern routines. This amounts to paying providers a time-limited per unit incentive for the provision of a component of the CPGs for a specific condition.⁷

The use of temporary incentives to overcome organizational inertia in firms is similar in spirit to the use of temporary incentives to change individual and consumer behavior. Firms often use temporary price discounts, such as sales and coupons, to market their products (Blattberg and Neslin 1990; Kirmani and Rao 2000; and Dupas 2014). Discounts encourage individuals to purchase goods that they are not in the habit of buying which in turn allow them to update their beliefs about the product's benefits. Similarly, temporary incentives have been used to try to help individuals develop better health habits such as exercise and quitting smoking.⁸ Recently, temporary incentives have

⁴ See Das and Hammer (2005); Das and Gertler (2007); Das, Hammer and Leonard (2008); Barber and Gertler (2009); Leonard and Masatu (2010); Gertler and Vermeersch (2012); and Monahan, M. et al. (2015).

⁵ For more evidence of organizational inertia serving as a barrier to CPG compliance see Grol (1990); Hudak, O'Donnell and Mazyrka (1995); Main, Cohen and DiClemente (1995); and Pathman et al. (1996).

⁶ We use a different definition of habits than the behavioral economics literature where habits are based on the addiction of models of Becker and Murphy (1988). Instead, we rely on the notions of fast and slow thinking discussed in Kahneman (2012) where tasks performed based on fast thinking become habits.

⁷ Paying an upfront lump sum amount is another option. However, it may be harder to ensure and verify the actual change in practice patterns. By paying based on actual performance the incentives also include a commitment device for compliance.

⁸ See for example Volpp et al. (2008); Volpp et al. (2009); Charness and Gneezy (2009); John et al. (2011); Royer et al. (2012); Cawley and Price (2013); and Acland and Levy (2015).

been used to stimulate long-term savings in the form of initially high interest rates and price-linked savings or lotteries (Gertler et al. 2015, and Schaner 2015). To our knowledge, our study is the first to use a field experiment to examine the effects of temporary incentives on long-run firm performance.

We test the effects of temporary incentives paid to clinics for early initiation of prenatal care using a field experiment conducted with *Plan Nacer*, an Argentine government program that provides health insurance to otherwise uninsured pregnant women and children.⁹ Prenatal care by skilled health professionals beginning in the first trimester of pregnancy is essential for good maternal and newborn health outcomes, and is part of standard medical training throughout the world (WHO 2006). Through early initiation of care, providers are able to detect and correct important health conditions such as infections or anemia before they jeopardize maternal or newborn outcomes as well as advise mothers on proper prenatal nutrition and prevention activities (Schwarcz et al. 2001; Carroli et al. 2001a and 2001b; Campbell and Graham 2006). Despite these recommendations and the scientific evidence, take-up of early initiation of prenatal care remains low worldwide (WHO 2014).

The field experiment randomized temporary financial incentives to health care clinics in which treatment clinics were paid a 200% premium for early initiation of prenatal care, i.e. before week 13. We find that the rate of early initiation of prenatal care was 34% higher in the treatment group than in the control group (0.42 versus 0.31) while the incentives were being paid, and that the higher levels of early initiation of prenatal care in the treatment group persisted at least 15 months and likely more than 24 months after the incentives ended. We document that clinics changed their routines by developing strategies to identify likely pregnant women and expanding the role of community health workers to find pregnant women and encourage them to start care early, and that these changes in routines also persisted at least 15 months after the incentives ended. Despite the large effect of the incentives on early initiation of care, we find no evidence of an effect on birth outcomes.

Our results may explain the mechanism behind recent evidence that permanent performance incentives do indeed improve both quality and quantity of care.¹⁰ The standard neoclassical

⁹ In 2013, Plan Nacer was expanded to other populations and renamed Programa Sumar.

¹⁰ See for example Basinga et al. (2011); Flores et al. (2013); Bonfrer et al. (2013); De Walque et al. (2015); Gertler and Vermeersch (2013); Gertler et al. (2014); and Huillery and Seban (2014). Miller and Babiartz (2013) provide a review.

explanation is that providers are reallocating their effort across services in response to the increased profit opportunities.¹¹ However, previous studies have been unable to distinguish between this mechanism and organizational inertia. One way to distinguish between the two mechanisms is to observe what happens when incentives are removed. While the incentives are in play both models predict a positive response. However, once the incentives are removed, practice patterns should revert to prior levels in the standard models but continue at the higher levels under organizational inertia.

Understanding the mechanism by which financial incentives work is not only scientifically interesting, but also policy relevant. If temporary financial incentives are able to induce providers to adopt permanent changes to their clinical practice patterns, then temporary incentives can achieve a boost in performance at a substantially cheaper cost than permanent incentives. Our results suggest that the mechanism behind positive provider responses to price increases is more related to adjustment costs than to responding to higher profit margins. In this case, long-term increases in productivity can be achieved more cheaply than through a permanent increase in fees.

2 CONCEPTUAL FRAMEWORK

We develop a stylized model of clinical practice patterns where clinics incur a fixed cost to change clinical practice routines. We assume that patients are identical, that clinics provide the same services to all patients, and that demand is exogenously determined.

Objective Function: Clinics have a pay-off function $R = \pi + \alpha HN$, where π is profits, H is health of the representative patient, N is the number of patients, and $\alpha \in [0,1]$ is the provider's intrinsic value of a unit of patient health.¹² As α rises the clinic is willing to sacrifice more income for patient health. When α takes on value 0, the clinic is purely extrinsically motivated, and when α is 1 the clinic is purely intrinsically motivated. While we allow for both extrinsic and intrinsic motivation in the model, all of the results follow even with pure extrinsic motivation. Allowing for intrinsic motivation does not change the direction of the predictions just the magnitude. Moreover, pure intrinsic

¹¹ See Baker et al. (1988); Holmstrom and Milgrom (1991); Gibbons (1997); and Lazear (2000).

¹² There is evidence to support intrinsic motivation as at least partially motivating medical care providers. See for example Leonard and Masatu (2010); Kolstad (2013); and Clemenens and Gotlieb (2014).

motivation by itself does not predict that temporary incentives would have long terms effects on productivity.¹³

Health Production Function: Treatment technology, as defined by CPGs, involves two services, S_1 and S_2 where $S_i = 1$ if the clinic provides the service and 0 if not. If the clinic provides both services, then it is operating at the production possibilities frontier. The health production function for the representative patient is $H = \lambda_1 S_1 + \lambda_2 S_2 + \varepsilon$, where ε is a mean zero random shock.

Clinical Practice Routine: Consider a clinic whose current clinical practice pattern routine is to provide S_1 to all patients. In this case, S_1 is the clinic's existing clinical practice pattern routine, and S_2 is an additional service that the clinic could choose to add to its practice routine. If the clinic wants to integrate the provision of S_2 into its practice pattern routine then it must incur an upfront fixed cost F . The fixed cost includes the cost of retooling to be able to provide S_2 , the cost of less reliable service provision while the new routine is being learned, and the cost of overcoming psychological resistance to change.

Profits: Clinics are paid p_i for S_i and the marginal cost of providing S_i to a patient is c_i . Clinic profits can then be expressed as:

$$\pi = \sum_{t=1}^{\infty} \beta^t [(p_1 - c_1) + (p_2 - c_2)S_2]N - FS_2 \quad , \quad (1)$$

where β is the clinic's discount rate.

Adoption: The clinic adopts S_2 if

$$R(S_2 = 1) - R(S_2 = 0) \geq 0 \quad . \quad (2)$$

Substitution of (1) and (2) into the pay-off function and rearranging terms allows us to write the condition in (3) as:

$$\sum_{t=0}^{\infty} \beta^t [p_2 - c_2 + \alpha \lambda_2]N \geq F \quad . \quad (3)$$

¹³ Without some sort of fixed costs of adjustment, both intrinsically and extrinsically motivated providers would still operate at the efficient frontier. Moreover, the intrinsic motivation literature suggests that incentives can negatively impact performance. The psychology literature in particular has long argued that performance-contingent incentives can be demotivating for intrinsically motivated workers. For example see Deci (1971); Pittman and Heller (1987); Deci et al. (1999); Deci (2001); Eccles and Wigfiel (2002); Deci and Ryan (2010). Benabou and Tirole (2003) embed these ideas in principle-agent models that they use to demonstrate the mechanisms through which financial incentives can "crowd-out" intrinsic motivation and thereby negatively affect performance. Recent laboratory experimental evidence on performance-contingent contracts confirms that incentives in the presence of intrinsic motivation can result in worse performance. For example see Fehr and Falk (1999); Fehr and Schmidt (2000); Gneezy and Ruitichini (2000a and 2000b); and Ariely et al. (2009).

Clinics are more likely to adopt S_2 if the profit margin from S_2 is higher, they are more intrinsically motivated, the effect of S_2 on patient health is higher, they have higher patient volumes, and they have lower discount rates.

Organizational inertia: Inertia is defined as when the present value of the fixed costs of changing organizational routine prevents the clinic from adopting a valuable improvement to production. The conditions are $(p_2 - c_2 + \alpha\lambda_2) \geq 0$ and $\sum_{t=1}^{\infty} \beta^t [p_2 - c_2 + \alpha\lambda_2]N < F$, i.e. S_2 is valuable but not adopted because of the fixed cost of adjusting organizational routine to be able to provide S_2 . Clinics who are more intrinsically motivated (i.e. higher α) are less likely to be frozen by organizational inertia and maybe even willing to lose money in order to adopt S_2 , especially if S_2 is very productive (i.e. higher λ_2).

Temporary Incentives: Organizational inertia can be overcome with a temporary increase in p_2 , the price of S_2 .¹⁴ Consider an increase to the price paid in period 1 that disappears in subsequent periods. Without loss of generality we can simplify the model to 2 periods with β as the discount rate. In this case, the increase of θ in p_2 in period 1 necessary to induce the provider to adopt S_2 is:

$$\theta \geq \frac{F}{N} - (1 + \beta)[p_2 - c_2 + \alpha\lambda_2]. \quad (4)$$

The temporary incentive, θ , at minimum covers the remainder of the fixed cost of adjustment that is not paid for the discounted present value of the future stream of surplus generated from the provision of S_2 . The incentive goes down with scale N , the profit margin $(p_2 - c_2)$, the extent to which clinics are extrinsically motivated times the marginal product of S_2 in the health production function $(\alpha\lambda_2)$, and the discount rate.

Cross-Price Effects: One concern voiced in the literature is that price increases for some services might lead to a reallocation of effort from other services that remain unchanged leading to negative cross-price effects. The implicit underlying model in these papers is an individual physician allocating time between activities with a time budget constraint. In our model of a medical care organization that can hire more staff, cross-price effects are generated based on the nature of economies of scope in either the health care production function or cost function. If both the production and cost functions are additively separable, then there are no cross-price effects. If the

¹⁴ The alternative is a lump sum payment that is vulnerable to the possibility of noncompliance and maybe difficult to verify. However, a temporary increase in p_2 requires the clinic to change routines and actually adopt S_2 in order to get paid. In this sense the temporary price increase also includes a commitment device and hence is ex ante preferable.

functions are not separable, then it is possible to have either negative or positive cross-price effects depending the nature of substitutability in the production and cost functions.

3 EXPERIMENTAL DESIGN

The field experiment was conducted by Plan Nacer, a public insurance program that began in 2005 to improve access to quality health care for otherwise uninsured pregnant women and children less than 6 years old (Musgrove 2010; Gertler et al. 2014). Like Medicaid in the U.S. and Seguro Popular in Mexico, the national Plan Nacer program transfers funds to local governments, in this case Provinces, who are then responsible for enrolling beneficiaries, organizing the provision of services, and paying medical care providers. An innovative feature of the Argentine program is that it uses financial incentives to ensure that beneficiaries receive high-quality care. Financing from the National level to Provinces is based for 60% on program enrollment and for 40% on performance. Provinces then use those funds to pay public health care facilities on a fee-for-service basis for health care provided to program beneficiaries. The national government determines the content of the benefits package, which is uniform across provinces, while provincial governments set the price they will pay to providers for each service in that package. Health facilities are free to choose how to use realized revenues within relatively broad guidelines. Some, though not all, provinces allow health facilities to pay bonuses to personnel.

Plan Nacer scaled up by first recruiting and training clinics in the operations of its program, including fee structure, billing, and other rules. The program regularly retrains the clinics to keep them up to date on any changes and reinforce areas that are perceived to be weak. After clinics are enrolled, clinic community outreach staff identify eligible women and children in the clinics' catchment areas in order to enroll them into the program. Clinic outreach staff also regularly contact beneficiaries to encourage them to take advantage of program benefits.

The field experiment was conducted with primary health care clinics in the Province of Misiones, one of the poorest in the country and with high rates of maternal and child mortality. In Misiones, the clinic is allowed to use up to 50% of revenue from Plan Nacer fees to pay bonuses to facility personnel at the discretion of the facility director. The rollout of Plan Nacer in Misiones was completed in 2008 long before the pilot study. As such, both providers and beneficiaries were knowledgeable of the operation of Plan Nacer before the experiment began.

The experimental intervention was designed to encourage early initiation of prenatal care for Plan Nacer beneficiaries, thereby aligning the incentives in Plan Nacer with official Argentine clinical

practice guidelines, medical school training, and international scientific evidence. Before the experiment, only one-third of Plan Nacer beneficiaries were initiating care in the first trimester (National Ministry of Health, 2009 and 2010). The experiment randomized temporary financial incentives to primary health care clinics in which treatment clinics were paid a 200% premium for early initiation of prenatal care, i.e. before week 13.

Table 1 presents the payment schedule for the periods before, during and after the intervention. Prior to the intervention period, the province paid facilities \$40 ARS for each prenatal visit regardless of when it occurred or whether it was the first or a subsequent visit.¹⁵ During the intervention period the fee was increased to \$120 ARS for 1st visits that occurred before week 13 but remained at \$40 ARS for subsequent visits. After that, the intervention period fees reverted to the original payment of \$40 ARS for all visits. The modification amounted to a 3-fold increase in the fee for 1st visits before week 13. The modified fee structure was implemented for 8 months - from May 2010 to December 2010. Facilities selected to receive the modified fee structure were invited to participate and notified of the time-limited implementation on April 14, 2010. Facility directors were required to sign a formal modification of their existing contract with Plan Nacer in order to receive the modified fee structure.

The study design included 37 clinics out of 262 primary care facilities of the province, of which 18 were randomly assigned to the treatment group and were offered the modified fee schedule. The other 19 formed the control group. Table 2 shows that compliance with treatment assignment was not perfect: out of 18 facilities assigned to the treatment group, 14 were actually treated as three refused to sign the agreement and a fourth closed before the intervention started. In addition, one of the facilities originally assigned to the control group was mistakenly offered the treatment and agreed to the modified fee structure. In the end, there were 36 facilities in the study excluding the one that closed.

4 DATA

The Province of Misiones maintains a well-developed and long-established automated medical record information system managed by the provincial authorities. Personnel at public primary health clinics and hospitals digitize a record of each service provided to each patient. The data are of unusually high quality in that key outcomes such as dates of visits, services delivered, and birth weight are recorded at the time of care by the provider; therefore we do not need to rely on maternal

¹⁵ The exchange rate for \$1 ARS was around \$0.25 USD between 2009 through 2011.

recall of these variables collected in surveys long after the visit. The data used in the analysis are extracted from these clinic records and contain information on the universe of patients for the 36 clinics in the study. The records also include the individual's national identity number, which is used to link the individual clinic medical records from primary health facilities with the registry of health insurance coverage, the registry of Plan Nacer beneficiaries, and hospital medical records. In all, 97% of the primary clinic medical records were merged with the data on insurance status and program beneficiary status. In addition, 75% of these were successfully merged with medical records data from hospitals. Therefore our analysis is able to evaluate the impact of the intervention for those women who initiated their prenatal care in one of the primary care clinics of the sample.

4.1 ANALYSIS SAMPLE

Figure 2 depicts the timeline of the study and the availability of data divided into 4 different sub-periods: (i) a 16-months pre-intervention period from January 2009 to April 2010, (ii) an 8-month intervention period from May 2010 to December 2010, (iii) a 15-month "post-intervention period I" from January 2011 to March 2012 and (iv) a 9-month "post-intervention period II" from April 2012 to December 2012.

Prenatal care data was consistently collected for the first 3 periods from January 2009 through March 2012. Starting in April 2012, however, Misiones adopted a new information system and as a result data from post-intervention period II cannot easily be compared to data from the earlier periods. In particular, the new system changed the codes used to classify the reason for visits in order to facilitate billing. If in the first visit the attending physician requested an ultrasound to confirm a pregnancy, this first visit was labeled as a "care visit" while the subsequent (second) visit, was labeled as the first prenatal visit, if indeed the ultrasound confirmed the pregnancy. On average, this would lead to a reduction in the share of women who had a visit labeled as "first prenatal visit" before week 13 and an increase in the weeks pregnant at the time of this visit. If the new coding system affected the treatment and control groups in the same way, the differences between the treatment and control groups would still capture the impact of the incentives, albeit possibly with some measurement error. Therefore, we analyze the data from post-intervention period II separately, and interpret the results with caution.

The analysis sample includes pregnant women who were beneficiaries of Plan Nacer at the time of their first prenatal visit.¹⁶ While information on prenatal care utilization is available for the full sample period, information related to birth outcomes is only available for women who gave birth in a public hospital through 2011, i.e. women who became pregnant before May 2011.

4.2 MEASUREMENT OF WEEKS PREGNANT AT 1ST PRENATAL VISIT

We construct the number of weeks of pregnancy at the time of the first prenatal visit as the difference between the date of the first visit and the last menstrual date (LMD). The LMD is routinely collected at the time of the visit to calculate the estimated date of delivery (EDD) and both are routinely recorded in the patient's medical record at the clinic.¹⁷

One potential problem is that medical personnel in treatment facilities might misreport the date of late first visit as occurring before week 13 so that they could bill to the program. We think this is unlikely for the following reasons. First, the week of visit is constructed from the date of the first prenatal visit and the LMD, both of which along with the EDD are recorded in real time in the medical record. In order to falsely report that a first visit occurred in the first 12 weeks, the provider would have to alter the date of the first visit relative to either the LMD or the EDD in the medical record. This would require some effort if done in real time and would be noticeable by auditors if altered ex post. Second, Plan Nacer uses external auditors to verify the accuracy of clinic billing. The auditors compare the detailed clinical records to the billing requests to find inconsistencies and the latter can lead to substantial financial penalties for the provinces. Finally, clinical records are legal documents in Argentina and practitioners could lose their medical license if caught systematically misreporting for financial gain.

To corroborate our belief that false reporting in the clinic records is unlikely, we empirically test whether there is any evidence of systematic misreporting using data from an alternative source. Specifically, we use gestational age at birth measured by physical examination obtained from hospital records to construct a second estimate of the LMD and weeks pregnant at the time of the first prenatal visit. The hospital personnel that attend the birth do not have any incentive to misreport hospital records. We then compare the estimated week of first visit based on gestational age at birth to the

¹⁶ We excluded non-beneficiaries because most of them have private health insurance and as such are likely to receive some of care and deliver at private facilities. Since we do not have data from private facilities, the outcomes of most of these observations are censored.

¹⁷ For 10% of the sample LDM was not recorded. For those cases, we use the EDD to recover the LMD.

week of first visit reported by the health facilities. The results do not show any evidence of systematic misreporting due to incentives. Appendix A provides a detailed discussion of the analysis and results.

4.3 DESCRIPTIVE STATISTICS AND BASELINE BALANCE

Table 3 reports the descriptive statistics for the key outcomes of interest and demographic characteristics at baseline, i.e. in the 16-month pre-intervention period (Jan 2009 – April 2010). Outcomes are balanced at baseline in that there are no statistically significant differences in the means of variables between the treatment and control groups. On average women had their first prenatal visit about 17.5 weeks into their pregnancy with about one-third of women having that visit before week 13. Women completed about 4.7 prenatal visits over the course of their pregnancy and more than 80% of them received a tetanus vaccine. Newborns weighed approximately 3,300 grams on average, while about 6% of them were born with low birth weight (i.e. less than 2,500 grams), and slightly more than 9% of births were born prematurely.

5 IDENTIFICATION AND ESTIMATION

We estimate both the intent-to-treat (ITT) and local average treatment (LATE) effects of the incentives on outcomes. The ITT is the effect of assigning a clinic to treatment on outcomes, regardless of compliance. It compares the mean outcome of the group assigned to treatment to the mean outcome of the group assigned to control and is estimated by regressing the outcome against an indicator of whether the clinic was assigned to treatment. The LATE is the effect of a clinic actually receiving the incentives and is estimated regressing the outcome against whether the clinic was actually treated, using the clinic's randomized assignment status as an instrumental variable for actual treatment (Imbens and Angrist 1994). In both cases, the treatment effect is identified off the variation induced by the randomized assignment status. In the discussion of results in the next section, we report the LATE estimates.¹⁸

Our sample is clustered within 36 health clinics since the random assignment of treatment occurred at the clinic level. As such, there may be intra-cluster correlation that must be considered for statistical inference. Standard methods of correcting standard errors rely on large sample theory both in the number of observations and in the number of clusters. Given the small number of clusters in our sample, we instead use statistical inference methods that are robust to randomized assignment of treatment among a small number of clusters. Specifically, we use the Wild bootstrap method to

¹⁸ The ITT results are almost identical to the LATE estimates, which is expected given the relatively high compliance rates to the original assignment. The ITT results are presented in Appendix C.

generate p -values for hypothesis testing in ITT models (Cameron et al. 2008) and an analogous method for hypothesis testing in the LATE models (Gelbach et al. 2009). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals, and uses 999 replications (Davidson and Flachaire 2008).

6 TIMING OF FIRST PRENATAL VISIT

In this section we report the results of analyses of the effects of the temporary incentives on the timing of the first prenatal visit and mechanisms by which clinics achieved those results.

6.1 DENSITIES

Figure 3 compares the densities of weeks pregnant at the time of the first prenatal visits for the clinics assigned to the treatment and control groups. Panel A shows that there is no difference between the densities of the treatment and control groups in the pre-intervention period. Panel B shows that the treatment group density is to the left of the control group density during the intervention period. Finally, Panel C and D show that the treatment group density is placed to the left of the control group density during post-intervention periods I and II. Kolmogorov-Smirnov tests for equality of the distributions cannot be rejected for the pre-intervention analysis, but are rejected for the intervention and both post-intervention periods with p -values of 0.031, 0.004, and 0.009 respectively. These results imply that the temporary incentives led to earlier initiation of care in the treatment group compared to the control group in the intervention period and that these higher levels of care persisted for at least 15 months and likely for 24 months and more after the higher fees were removed.

6.2 SHORT-RUN EFFECTS

Table 4 reports the estimates of the effects of the temporary fees on the early initiation of care. Panel A reports the results for weeks pregnant at the time of the first prenatal visit and Panel B reports the results for whether the first visit occurred before week 13. The first column reports the results for the intervention period and the second and third columns report the results for the post-intervention periods. During the intervention period, on average women in the treatment group had their 1st visit about 1.5 weeks earlier in their pregnancy than women in the control group. The share of women in the treatment group who had their 1st visit before week 13 is 11 percentage points higher than the control group; approximately 35% higher than the control group. Both estimates are significantly different from zero at conventional p -values.

6.3 LONG-RUN EFFECTS

Our model of behavioral inertia provided clear predictions about provider behavior once temporary incentives disappear: i.e. if the fee increase is enough to overcome the fixed costs of adapting a new practice, clinics should maintain higher levels of prenatal care after incentives are removed. Column 2 of Table 4 reports estimated impact of the temporary fee increase on early initiation of care in the 15-month period after the fees were removed. On average, women in the treatment group started their care 1.6 weeks earlier than those in the control group. The difference between the treatment and control groups in the share of women who had their 1st visit before week 13 was 8 percentage points. Both estimates are statistically different from zero at conventional levels. Further, we cannot reject the null hypothesis that the impact is different in the intervention and post-intervention periods. These results are consistent with the hypothesis that temporary incentives help overcome behavioral inertia and motivate long-run changes in performance.

While there is no significant difference between the effect during the intervention and the post-intervention periods, one concern may be that the effect of treatment slowly trended towards zero after the incentives ended. To test this hypothesis, we plot the mean number of weeks pregnant at the time of first prenatal visit for treatment and control groups, before, during and after the intervention (Figure 4).¹⁹ We split the pre-intervention period into two sub-periods of 6-months each and the post-intervention period into 3 sub-periods: the first two are 6 months and the third is 3 months. The treatment effect is the difference between the two lines. While the treatment and control groups have similar trends before the intervention, the treatment group appears to receive earlier care during the intervention, and the change persists after the end of the intervention. Notice that there is little if any fall off over the post-intervention period. Rather the treatment effects remain fairly constant over the 15-month post-intervention period I. Figure 5 depicts the same relationship for the share of women who receive care before week 13 of pregnancy.²⁰ Again, the effects of the intervention appear to continue at a steady rate after it is discontinued.

6.4 LONGER-RUN EFFECTS

The period of analysis in our main results is restricted to January 2009 to March 2012. Recall that starting in April 2012, the visit coding system changed. Hence starting in April 2012 what is reported as first visits in the data is actually a mix of first and second visits. As a result the average of weeks pregnant at first visit increases and the share of pregnant women whose first visit was before

¹⁹ As discussed above, the information from post-intervention period II (April-December 2012) uses a different metric and is therefore not included in this figure.

²⁰ Ibidem.

week 13 falls relative to previous periods. Column 3 in Table 4 shows the results for this last period. The mean average of weeks pregnant at the time of the first visit for the control group is substantially higher for this period than for previous periods and the mean share that had their first visit before week 13 is substantially lower, suggesting that there is measurement error in our main outcome in this period. However, this difference in coding should have a similar effect in treatment and control clinics given the randomized assignment of the treatment. Therefore the difference between treatment and control clinics should cancel out the measurement error and provide us with unbiased estimates of the impact.

The results in Table 4 show a statistically significant reduction in the number of weeks pregnant at the time of the first visit and a statistically significant increase in the share of pregnant women who had their first visit before week 13. These results suggest that improved productivity from the temporary fee increase persisted at least 24 months after the fees were removed.

6.5 ROBUSTNESS

We implement three robustness checks. First, the main sample may include pregnancies that start in one period and end in another, which could cloud the effect of the incentives on timing of the first visit. For example, a woman who is 6 months pregnant and has not had a prenatal visit when the intervention starts and subsequently receives her first prenatal checkup during the intervention, would be counted as a third trimester first visit during the intervention period, even though the intervention cannot affect whether she receives prenatal care before week 13. Hence, in this robustness test we re-estimate the models on a restricted sample where women are no more than one month pregnant in the first month of the period and no less than 3-months pregnant in the last month of the period. The results, reported in Panels B of Appendix Tables B1 and B2, are very close in magnitude and statistical significance to the main results in Table 4.

Second, even though there were no statistical differences in baseline means, it is possible that randomization was not able to fully balance the treatment and control groups on unobservable characteristics given the small number of clinics. In order to test for this possibility, we estimate the models using difference-in-differences with clinic and month fixed effects. The results, reported in Panels C of Appendix Table B1 and B2, are very close in magnitude and statistical significance to the main results in Table 4.

Finally, in studies involving a small sample of clusters there is a concern that a few outliers may drive the average effect found in the previous sections. We explore this possibility by estimating clinic-specific treatment effects whereby we compare each treated clinic individually to the control

clinics as a group. Appendix Figures B1 and B2 plot these individual clinic treatment effects for the outcomes of weeks pregnant at the time of the first prenatal visit (B1) and for the probability of that the first visit occurred before week 13 (B2), respectively. The results are sorted along the x-axis from the lowest to the highest estimated effect, while the dashed blue line is the intent-to-treat effect calculated by pooling the intervention and the first post intervention period. The solid black line represents a zero treatment effect. The vertical lines are 95% confidence intervals constructed using standard errors obtained from the Wild bootstrap procedure. The figures show that the hypothesis of no treatment effect is rejected for 11 out of 17 clinics in Figure B1 and 12 out of 17 clinics in Figure B2. In addition, the treatment effects have the expected sign in 15 out 17 clinics in Figure B1 and 14 out of clinics in Figure B2. This provides evidence that our results are not driven by a few large-effect clinics.

6.6 MECHANISMS

In order to better understand how clinics were able to achieve such large increases in the share of women who initiated prenatal care before week 13, we conducted a series of in-depth interviews with professionals in a sub-sample of 5 treatment clinics and 3 comparison clinics.²¹ We find that treatment clinics adopted new practices and changed routines in order to increase early initiation of prenatal care. After the initial invitation to participate in the pilot, all 5 interviewed treatment clinics organized a team meeting with the staff in order to discuss strategies to respond to the new incentive scheme. Various treatment clinics adopted different strategies, but all of them involved expanding the scope of work of community health workers to identify and encourage newly pregnant Plan Nacer to initiate their prenatal care early. In some clinics, the director supported the change in strategies by changing the way the financial incentives were distributed between staff members.²² In particular, some of them started allocating the incentives conditional on the number

²¹ The clinics interviewed are located in Posadas, the capital of Misiones Province. Each interview took approximately 45 minutes. The interviews were carried out in May 2015.

²² Up to 2013, any health facility participating in *Plan Nacer* in Misiones was able to use up to 10% of their Plan Nacer funds to pay incentives to personnel. If the facility achieved a set of health targets measuring using performance indicators (tracers) set by the province, that facility was able to use up to 50% of funds for monetary incentives to health professionals. The bonuses could be assigned to any person working at the health facility, including the health workers, administrative personnel, volunteers, and even to personnel affiliated with other programs as long as they were not absent for more than 10 working days in a month, they did not participate in a strike organized by the union, and they were not subjected to a disciplinary sanction (suspension without pay or dismissal). In all cases, the final decision regarding assignment of incentives to personnel was the prerogative of the clinic director.

of pregnant women that each team member brought to the clinic in a month. This allocation further incentivized health workers to test new practices.

The in-depth interviews uncovered several innovative strategies that treatment clinics developed to identify pregnancies early. For instance, health workers started to follow up women who used birth control pills.²³ Specifically community health workers prioritized home visits to women who had not picked up their pills. Second, health workers started targeting women at high risk of not coming in for an early checkup. According to the interviewed doctors, mothers who already have children are less likely to initiate their prenatal visits early in a new pregnancy. However, many of these women are also eligible for weekly free milk distribution for their older children. Health workers met these mothers at the time of the milk distribution, enquired about their last menstruation date, and offered an instant-read pregnancy test to those women whose menstruation was overdue. Third, health workers identified difficulties in providing early prenatal care to adolescents, as they might be unwilling to reveal a pregnancy, especially to their parents. Community health workers therefore decided to change the timing of home visits, so as to increase the chance of finding adolescents by themselves. In one of the interviewed clinics, the work flow was modified so as to ensure predictable availability of a gynecologist on certain days of the week. This in turn provided an easy way for community health workers and administrative staff to schedule patient appointments. Other clinics introduced new ways of keeping track of “at risk” patients, such as a notebook that kept track of any visits to the homes of women that were at risk, or a map that identified catchment areas of community health workers with corresponding (potential) pregnancies.

We are able to substantiate the claims of increased outreach using clinic administrative records on the number of community outreach activities that resulted in actual maternal-child service at the clinic.²⁴ Figure 6 displays the average and median number of outreach activities that resulted in actual maternal-child services for the pre-intervention, intervention, and post-intervention I periods.²⁵ The results show that there is little difference in outreach activities between treatment and control clinics in the pre-intervention period. In the intervention period the treatment

²³ Birth control pills are dispensed free of charge by each health facility’s pharmacy unit, though women cannot collect more than a monthly supply at any one time. The pharmacy unit keeps records of all birth control pill collections.

²⁴ Plan Nacer finances clinic outreach activities on a fee-for-service basis and employs an external independent auditor to audit clinic activity reports. Treatment and comparison clinics were paid the same fee for these activities before, during and after the experiment.

²⁵ The medians are better measures of central tendency as the densities of both activities are asymmetric heavily skewed to the right.

group evidenced substantially more activities than the control group, and this difference is sustained through the post-intervention period.

We use the data to estimate the differences in log number of activities between the treatment and control groups. The results show no differences in activities in the pre-intervention period and positive and statistically significant higher levels of activities in the treatment clinics in the intervention and post-intervention I periods (Table 5). Again, we cannot reject that the hypothesis that the effects are different in the intervention and post-intervention periods implying that the increase in successful outreach activities persisted after the temporary incentives were removed.

6.7 PSYCHOLOGICAL BARRIERS

In the previous subsection we documented tangible costs of adjustment to increase early initiation of prenatal care. An additional potential cost of adjustment is psychological barriers to change. One way to overcome psychological resistance is to make the guideline or task more salient in the minds of the clinic staff.²⁶ The issue is not one of lack of knowledge or information as initiating care in the first trimester has been in CPGs since the 1970s and has been a long-standing part of standard medical education. Rather the issue is the importance or priority that staff place on the task.

The temporary incentives might have increased the importance of early initiation of care in the staff's minds, thereby making it a higher priority for action. The higher the priority of a task, the less likely psychological barriers would stand in the way of adoption. Kahneman (2012, pp 8) states that "...frequently mentioned topics populate the mind..." more than others and "...people tend to assess the relative importance of issues by the ease with which they are retrieved from memory". As such, salience "...is enhanced by mere mention of an event" (Kahneman 2012, pp 331). If incomplete or non-adoption of a task is a matter of salience then the observed treatment effects may be explained by the fact that temporary incentives help to overcome this type of psychological barrier to change.

While we do not have information on the salience of early initiation of care during or shortly after the experiment, we explore whether the temporary fee increase made early initiation of care more important in the minds of the clinic staff after the end of the experiment, using an online survey administered to the chief medical officer of each clinic about the absolute and relative importance of

²⁶ Taylor and Thompson (1982) define salience as, "...the phenomenon that when one's attention is differentially directed to one portion of the environment rather than to others, the information contained in that portion will receive disproportionate weighting in subsequent judgments". See Bordalo et al. (2012, 2013) for a more recent discussion of salience and choice theory. See De Mel et al. (2013), and Karlan et al. (2015) for empirical analysis of salience effects through informational reminders.

seven different prenatal care procedures including initiating prenatal care prior to week 13 of pregnancy (see Appendix D).

Figures 8 and 9 compare the absolute score and relative ranking of the procedures in terms of importance for prenatal care. The absolute scores ranges from 0 to 5, with 5 being the highest while the relative ranking sorts the seven practices from 1 to 7, with 1 being the highest ranking. Our outcomes of interest are the absolute score and relative ranking assigned to early initiation of prenatal care. Figure 8 shows that the absolute score assigned to early prenatal care is on average 4.8 in the treatment group and 4.7 in the control group. Figure 9 shows that on average the relative ranking for this practice is also similar between the two groups, 2.0 for the treatment group and 1.9 for the control group. Moreover, these differences are not statistically significant at conventional levels (see Appendix D). These results suggest that the early initiation of prenatal care is of similar high absolute and relative importance and that temporary fees did not have a lasting effect on either the absolute nor relative importance.

6.8 ALTERNATIVE EXPLANATIONS

One alternative explanation for the short-term treatment effects is that the incentives are causing treatment clinics to try to attract pregnant women who otherwise would have used other clinics. This is unlikely to be true as beneficiary women are assigned to specific clinics when enrolled in Plan Nacer. Moreover, the number of patients per month and the share that initiate care before week 13 are the same in the pre- and post-intervention periods for control clinics, and the average monthly number of patients is also the same in the pre- and post-intervention periods for the treatment clinics.

An alternative explanation for long-run results is that after the temporary incentives ended, women who were pregnant during the intervention periods passed the message of the importance of early initiation of care onto other beneficiary women who became pregnant during the post-intervention period. Hence, the persistence of the effect of the incentives after the incentives might be caused by an informational spillover. However, the higher amount of the community outreach activities in treatment clinics, the mechanism used to generate higher early initiation of care, continued into the post-experimental period at the same level as in the intervention period. Hence, if there were information spillovers in the post-intervention period, then one would expect to see higher treatment effects in the post-intervention period than in the intervention period.

Finally, one might argue that the clinics continued the new routines after the temporary fees were eliminated because they faced a large fixed cost of reverting to the old routines and not because

the new routines added net value. However, in this case, we think that the fixed costs of reversing the routines were small, because the community health workers could simply have returned to their old patterns of activities.

7 CROSS-PRICE EFFECTS (SPILLOVER)

While the modified fee schedule was designed to affect the timing of the first prenatal visit, we might expect providers to reduce effort supplied to other services, resulting in a lower provision of such services to patients. We test for this by estimating the effect of the incentives on the probability of pregnant women having a valid tetanus vaccine, and the number of prenatal visits. The results presented in Table 6 report no evidence of cross-price effects, positive or negative, in either the intervention period or in post-intervention period I. In fact, the levels of these services appear to be constant over time. While the concern about crowding-out is typically for a context of individual providers facing time and effort constraints, our results are consistent with a firm setting where there are no overall effort or time constraints.

8 BIRTH OUTCOMES

Next we address the question of whether the effect of the incentives for early initiation of prenatal care translated into improved birth outcomes as measured by birth weight, low birth weight, and premature birth. As shown in Figure 7 and reported in Table 7 we find no effect of the incentives on birth outcomes in either the intervention period or in the post-intervention period.

There are a number of possible reasons for this. First, the sample could be too small to be able to detect a statistically significant effect on outcomes. However, the point estimates are very small, half of them are negative and they are of similar magnitude to differences between treatment and control groups in the pre-intervention period. Second, given that the results on birth outcomes are obtained from an analysis of a subsample of beneficiaries for whom we were able to merge prenatal care records with hospital medical records, it is possible that the results in Table 4 do not hold for this subsample. We therefore replicate the prenatal care analysis using only the subsample of women for whom hospital medical records are available. Overall, we obtain similar results to those obtained with the full sample.²⁷ Third, despite the medical literature and CPG recommendation, it is possible that early initiation of care matters only a small amount for the general population of pregnant

²⁷ Results of this analysis are available upon request.

women, even if early initiation of care matters a great deal for high-risk patients. High risk patients include, among others, smokers, substance abusers, those with poor medical and pregnancy histories, and those who start prenatal care very late in their third trimester or only when a problem occurs. It may be that the increase in early initiation of care comes from primarily low-risk mothers who are less likely to benefit from early initiation of care. One would think that it would be easier to persuade low-risk mothers to come a littler earlier than to convince high-risk mothers who are reluctant to come for any care at all.

In fact, this is consistent with the small reduction in the average weeks pregnant at the time of the first prenatal visit. On average, women in the treatment group initiated prenatal care about 1.5 weeks earlier than women in the control group. Prenatal care may affect birth outcomes by diagnosing and treating illness such as hypertension and gestational diabetes as well as trying to change maternal behavior through promoting activities such as good nutrition, not smoking and not consuming alcohol. If the intervention had induced high-risk women who otherwise would have had 1st visit much later in the pregnancy, then the incentives may have had a measurable impact on birth outcomes. Hence, while the incentives were effective in increasing early initiation of care, they did not manage to sufficiently affect the group most likely to benefit. The solution might be to condition incentives on attending high-risk women, but risk is difficult and expensive to identify and verify and therefore may not be contractible.

9 DISCUSSION

We examine the effects of temporary financial incentives for medical care providers to increase early initiation of prenatal care for pregnant women using a randomized controlled trial in Argentina. The intervention randomly allocates a three-fold increase in the fee paid to health facilities for each initial prenatal visit that occurs before week 13 of pregnancy. This premium was implemented for a period of 8 months and then ended. Using data on health services and birth outcomes from medical records, we investigate both the short-term effects of the incentive and whether the effects persist once the direct monetary compensation disappears.

Our results suggest that the temporary incentives motivated long-run changes in performance. We find that the incentives led to pregnant women being 35% more likely to initiate prenatal care before week 13 and that the higher levels of early initiation of care persisted for at least 15 months and likely more than 24 months after the incentives ended. These results are consistent with a model of providers who face a fixed cost to changing their clinical practice routines, i.e.

organizational inertia. Temporary incentives induced providers to adopt changes to their clinical practice patterns by helping them to overcome inertia. Once they adopt changes to practice patterns that they believe are beneficial to patients, the changes persist even after the monetary incentives disappear. These results are consistent with the findings from in-depth interviews that evidenced that treatment clinics adopted innovative practices and changed routines in order to increase early initiation of prenatal care.

Our study adds to the growing body of evidence that incentives are effective in improving provider performance. Our results also have a number of important policy implications. First, our results suggest that temporary incentives may be effective in motivating long-term provider performance at a substantially lower cost than permanent incentives. Second, while we find that incentives are able to motivate changes in clinical practice patterns, we did not find improvements in health outcomes. The monetary incentives that were implemented were not able to sufficiently reach those women for whom early initiation of prenatal care would have the largest health impact. Therefore, incentives may be made more effective by defining ex-ante the population most likely to benefit, and tailoring incentives towards this population. However, tailoring incentives to high risk populations or those most likely to benefit from the services may not be contractible as these characteristics are typically not observable. This is maybe a major limitation of using incentive contracts to improve health outcomes.

REFERENCES

- Acland, D., & Levy, M. R. (2015). "Naiveté, projection bias, and habit formation in gym attendance," *Management Science*, 61(1), 146-160.
- Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). "Large stakes and big mistakes," *The Review of Economic Studies*, 76(2), 451-469.
- Baker, G. P., Jensen, M. C., & Murphy, K. J. (1988). "Compensation and incentives: practice vs. theory," *The Journal of Finance*, 43(3), 593-616.
- Barber, S. L., & Gertler, P. J. (2009). "Empowering women to obtain high quality care: evidence from an evaluation of Mexico's conditional cash transfer programme," *Health Policy and Planning*, 24(1), 18-25.
- Basinga, P., Gertler, P. J., Binagwaho, A., Soucat, A. L., Sturdy, J., & Vermeersch, C. M. (2011). "Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation," *The Lancet*, 377(9775), 1421-1428.
- Becker, G. S. & Murphy, K. M. (1988). "A theory of rational addiction," *The Journal of Political Economy*, 96(4), 675-700.
- Becker, M. C. (2004). "Organizational routines: a review of the literature," *Industrial and Corporate Change*, 13(4), 643-678.
- Benabou, R. & Tirole, J. (2003). "Intrinsic and extrinsic motivation," *The Review of Economic Studies*, 70(3), 489-520.
- Blattberg, R. C. & Neslin, S. A. (1990). "Sales promotion: concepts, methods, and strategies," Englewood Cliffs, Prentice Hall, New Jersey.
- Bloom, N., Propper, C., Siler, S., & Van Reenan, J. (2015). "The impact of competition on management quality: Evidence from public hospitals," *The Review of Economic Studies*, 82(2), 457-489.
- Bonfrer, I., Soeters, R., van de Poel, E., Basenya, O., Longin, G., van de Looij, F., & van Doorslaer, E. (2013). "The effects of performance-based financing on the use and quality of health care in Burundi: an impact evaluation," *The Lancet*, 381, S19.
- Bordalo, P., Gennaioli, N. & Shleifer, A. (2012). "Salience theory of choice under risk," *The Quarterly Journal of Economics*, 127 (3): 1243-1285.
- Bordalo, P., Gennaioli, N. & Shleifer, A. (2013). "Salience and consumer choice," *The Journal of Political Economy*, 121(5), 803-843.
- Cabana, M. D., Rand, C. S., Powe, N. R., Wu, A. W., Wilson, M. H., Abboud, P. A. C., & Rubin, H. R. (1999). "Why don't physicians follow clinical practice guidelines?: A framework for improvement," *JAMA*, 282(15), 1458-1465.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). "Bootstrap-based improvements for inference with clustered errors," *The Review of Economics and Statistics*, 90(3), 414-427.
- Campbell, O. M. & Graham, W. J. (2006). "Strategies for reducing maternal mortality: Getting on with what works," *The Lancet*, 368(9543), 1284-1299.
- Campbell, S., Reeves, D., Kontopantelis, E., Middleton, E., Sibbald, B., & Roland, M. (2007). "Quality of primary care in England with the introduction of pay for performance," *The New England Journal of Medicine*, 357(2), 181-190.

- Carroll, G. R., & Hannan, M. T. (2000). "The demography of corporations and industries," Princeton University Press.
- Carroli, G., Villar, J., Piaggio, G., Khan-Neelofur, D., Gülmezoglu, M., Mugford, M., & Bergsjø, P. (2001). "WHO systematic review of randomized controlled trials of routine antenatal care," *The Lancet*, 357(9268), 1565-1570.
- Carroli, G., Rooney, C., & Villar, J. (2001). "How effective is antenatal care in preventing maternal mortality and serious morbidity? An Overview of the Evidence," *Paediatric and Perinatal Epidemiology*, 15(s1), 1-42.
- Cawley, J., & Price, J. A. (2013). "A case study of a workplace wellness program that offers financial incentives for weight loss," *Journal of Health Economics*, 32(5), 794-803.
- Charness, G. & Gneezy, U. (2009). "Incentives to exercise," *Econometrica*, 77(3), 909-931.
- Clemens, J. & Gottlieb, J. D. (2014). "Do physicians' financial incentives affect medical treatment and patient health?" *The American Economic Review*, 104(4), 1320-1349.
- Das, J., & Gertler, P. J. (2007). "Variations in practice quality in five low-income countries: a conceptual overview," *Health Affairs*, 26(3), w296-w309.
- Das, J. & Hammer, J. (2005). "Which Doctor? Combining vignettes and item response to measure clinical competence," *Journal of Development Economics*, 78(2), 348-383.
- Das, J., Hammer, J., & Leonard, K. (2008). "The quality of medical advice in low-income countries," *The Journal of Economic Perspectives*, 22(2), 93-114.
- Davidson, R. & Flachaire, E. (2008). "The Wild bootstrap, tamed at last," *Journal of Econometrics*, 146(1), 162-169.
- de Mel, S., McIntosh, C., & Woodruff, C. (2013). "Deposit collecting: Unbundling the role of frequency, salience, and habit formation in generating savings," *The American Economic Review*, 103(3), 387-92.
- De Walque, D., Gertler, P. J., Bautista-Arredondo, S., Kwan, A., Vermeersch, C., de Dieu Bizimana, J., & Condo, J. (2015). "Using provider performance incentives to increase HIV testing and counseling services in Rwanda," *Journal of Health Economics*, 40(2), 1-9.
- Deci, E. L. (1971). "Effects of externally mediated rewards on intrinsic motivation," *Journal of Personality and Social Psychology*, 18, 105-115.
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation," *Psychological Bulletin*, 125(6), 627.
- Deci, E. L., Koestner, R., & Ryan, R. M. (2001). "Extrinsic rewards and intrinsic motivation in education: Reconsidered once again," *Review of Educational Research*, 71(1), 1-27.
- Deci, E. L. and Ryan, R.M. (2010). "Self-determination," *John Wiley & Sons, Inc.*
- DellaVigna, S. (2009). "Psychology and economics: evidence from the field," *Journal of Economic Literature*, 47(2), 315-372.
- Dupas, P. (2014). "Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment," *Econometrica*, 82(1), 197-28.
- Eccles, J. S. & Wigfield, A. (2002). "Motivational beliefs, values, and goals," *Annual Review of Psychology*, 53(1), 109-132.

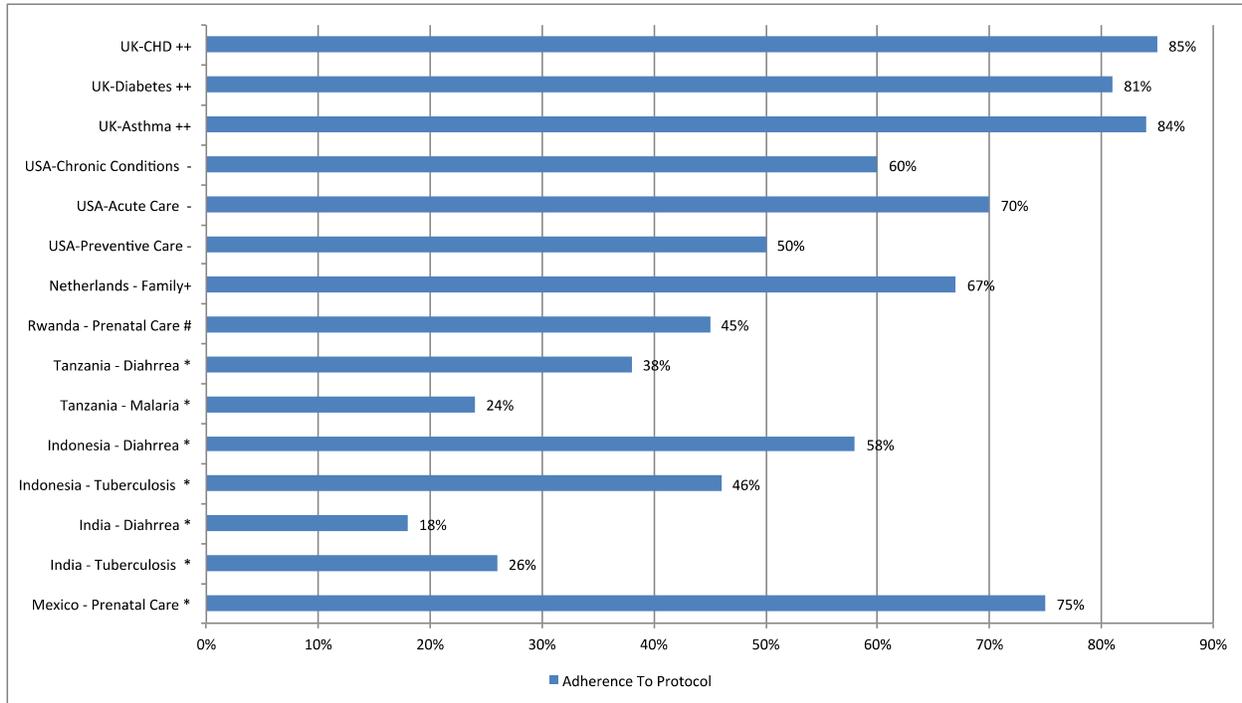
- Fehr, E. & Falk, A. (1999). "Wage rigidity in a competitive incomplete contract market," *Journal of Political Economy*, 107(1), 106-134.
- Fehr, E. & Schmidt, K. M. (2000). "Fairness, incentives, and contractual choices," *European Economic Review*, 44(4), 1057-1068.
- Flores, G., Ir, P., Men, C. R., O'Donnell, O., & van Doorslaer, E. (2013). "Financial protection of patients through compensation of providers: The impact of health equity funds in Cambodia," *Journal of Health Economics*, 32(6), 1180-1193.
- Gelbach, J. B., Klick, J., & Stratmann, T. (2009). "Cheap donuts and expensive broccoli: the effect of relative prices on obesity," *Working Paper*.
- Gertler, P., Giovagnoli, P. I., & Martinez, S. W. (2014). "Rewarding provider performance to enable a healthy start to life: evidence from Argentina's Plan Nacer," *World Bank Policy Research Working Paper*, 6884, World Bank, Washington, DC.
- Gertler, P. Seira E., and Scott A. (2015). "Long-term effects of temporary prize-linked savings lotteries on accounts openings and balances," *UC Berkeley Working Paper*, Berkeley California.
- Gertler, P., & Vermeersch, C. (2012). "Using performance incentives to improve health outcomes," *World Bank Policy Research Working Paper*.
- Gertler, P. & Vermeersch, C. (2013). "Using performance incentives to improve medical care productivity and health outcomes," *NBER Working Papers* 19046, National Bureau of Economic Research, Cambridge, MA.
- Gibbons, R. (1997). "An introduction to applicable game theory," *Journal of Economic Perspectives*, 11(1), 127-149.
- Gibbons, R., & Henderson, R. (2012). "Relational contracts and organizational capabilities," *Organization Science*, 23(5), 1350-1364.
- Gibbons, R., & Henderson, R. (2013). "What do managers do? Exploring persistent performance differences amongst seemingly similar enterprises," *The Handbook of Organizational Economics*, Chapter 17, pages 680-731, Robert Gibbons and John Roberts, Editors, Princeton University Press, Princeton and Oxford.
- Gneezy, U., & Rustichini, A. (2000a). "Pay enough or don't pay at all," *The Quarterly Journal of Economics*, 115(3), 791-810.
- Gneezy, U., & Rustichini, A. (2000b). "A fine price," *The Journal of Legal Studies*, 29(1), 1-17.
- Grol, R. P. T. M. (1990). "National standard setting for quality of care in general practice: attitudes of general practitioners and response to a set of standards," *British Journal of General Practice*, 40(338), 361-364.
- Grol, R. (2001). "Successes and failures in the implementation of evidence-based guidelines for clinical practice," *Medical Care*, 39(8), 11-46.
- Grol, R., & Grimshaw, J. (2003). "From best evidence to best practice: effective implementation of change in patients' care," *The Lancet*, 362(9391), 1225-1230.
- Hannan, M. T., & Freeman, J. (1984). "Structural inertia and organizational change," *American Sociological Review*, 149-164.
- Hoff, T. (2014). "When routines support or stifle innovation: Evidence from primary care practices," *Academy of Management Proceedings*, Vol. 2014, No. 1, p. 11116.

- Holmstrom, B. & Milgrom, P. (1991). "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job Design," *Journal of Law, Economics, & Organization*, 7 (Special Issue), 24-52.
- Hudak, B. B., O'Donnell, J., & Mazyrka, N. (1995). "Infant sleep position: pediatricians' advice to parents," *Pediatrics*, 95(1), 55-58.
- Huillery, E. & Seban, J. (2014). "Pay-for-Performance, motivation and final output in the health sector: Experimental evidence from the Democratic Republic of Congo," *Working Paper*, Department of Economics, Sciences Po, Paris.
- Imbens, G. W. & Angrist, J. D. (1994). "Identification and estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467-475.
- John, L. K., Loewenstein, G., Troxel, A. B., Norton, L., Fassbender, J. E., & Volpp, K. G. (2011). "Financial incentives for extended weight loss: a randomized, controlled trial," *Journal of General Internal Medicine*, 26(6), 621-626.
- Kahneman, D. (2012). "Thinking, fast and slow," Farrar, Straus and Giroux, New York.
- Karlan, D., M. McConnell, S. Mullainathan & Jonathan Zinman (2015). "Getting to the top of mind: How reminders increase savings," *Management Science*, forthcoming.
- Kirman, A. & Rao, A. R. (2000). "No pain, no gain: A critical review of the literature on signaling unobserved product quality," *Journal of Marketing*, 64(2), 66-79.
- Kolstad, J. T. (2013). "Information and quality when motivation is intrinsic: Evidence from surgeon report cards," *The American Economic Review*, 103(7), 2875-2910.
- Lazear, E. P. (2000). "Performance pay and productivity," *The American Economic Review*, 90(5), 1346-1361.
- Leonard, K. L. & Masatu, M. C. (2010), "Professionalism and the know-do gap: Exploring intrinsic motivation among health workers in Tanzania," *Health Economics*, 19(12), 1461-1477.
- Main, D. S., Cohen, S. J., & DiClemente, C. C. (1995). "Measuring physician readiness to change cancer screening: preliminary results," *American Journal of Preventive Medicine*.
- Miller, G. & Babiarz, K. S. (2013). "Pay-for-performance incentives in low- and middle-income country health programs," *NBER Working Papers* 18932, National Bureau of Economic Research, Inc.
- Mohanan, M., Vera-Hernández, M., Das, V., Giardili, S., Goldhaber-Fiebert, J. D., Rabin, T. L., & Seth, A. (2015). "The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural India," *JAMA Pediatrics*.
- Musgrove, P. (2010). "Plan Nacer, Argentina: Provincial maternal and child health insurance using Results-Based Financing (RBF)," Mimeo.
- National Ministry of Health (2009). "Informe de gestión Plan Nacer," Área Técnica, Unidad Ejecutora Central. Buenos Aires, Argentina.
- National Ministry of Health (2010). "Informe de gestión Plan Nacer," Área Técnica, Unidad Ejecutora Central. Revised version March. Buenos Aires, Argentina.
- National Ministry of Health (2010b). "Nomenclador único 2010," Plan Nacer, Buenos Aires, Argentina.
- Nelson, R. & S. Winter (1982). "An evolutionary theory of economic change," Harvard University Press.

- Pathman, D. E., Konrad, T. R., Freed, G. L., Freeman, V. A., & Koch, G. G. (1996). "The awareness-to-adherence model of the steps to clinical guideline compliance: the case of pediatric vaccine recommendations," *Medical Care*, 34(9), 873-889.
- Pittman, T. S. & Heller, J. F. (1987). "Social motivation," *Annual Review of Psychology*, 38(1), 461-490.
- Royer, H. M. Stehr, and J. Sydnor (2012). "Incentives, commitments and habit formation in exercise: evidence from a field experiment with workers at a Fortune-500 company" NBER Working Paper 18580, forthcoming in *American Journal of Economics: Applied Economics*.
- Schaner, S., (2015). "The persistent power of behavioral change: Long run impacts of temporary savings subsidies for the poor." Department of Economics, Dartmouth University, http://www.dartmouth.edu/~sschaner/main_files/Schaner_LongRun.pdf
- Schuster, M. A., McGlynn, E. A., & Brook, R. H. (1998). "How good is the quality of health care in the United States?," *Milbank Quarterly*, 76(4), 517-563.
- Schwarcz, R., Uranga, A., Lomuto, C., Martinez, I., Galimberti, D., García, O. M., Etcheverry, M. E., & Queiruga, M. (2001). "El cuidado prenatal: Guía para la práctica del cuidado preconcepcional y del control prenatal." National Ministry of Health, Argentina.
- Taylor, S. E., & Thompson, S. C. (1982). "Stalking the elusive 'vividness' effect," *Psychological Review*, 89(2), 155.
- Thaler, R. H. & Sunstein C.R. (2009). "Nudge: Improving decisions about health, wealth, and happiness," Penguin Books, New York.
- Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J., & Loewenstein, G. (2008). "Financial incentive-based approaches for weight loss: a randomized trial," *JAMA*, 300(22), 2631-2637.
- Volpp, K. G., Troxel, A. B., Pauly, M. V., Glick, H. A., Puig, A., Asch, D. A., ... & Audrain-McGovern, J. (2009). "A randomized, controlled trial of financial incentives for smoking cessation," *The New England Journal of Medicine*, 360(7), 699-709.
- Wooldridge, J. M. (2007). "Inverse probability weighted estimation for general missing data problems," *Journal of Econometrics*, 141(2), 1281-1301.
- World Health Organization (2006). "Standards for maternal and neonatal care: Provision of effective antenatal care," World Health Organization, Geneva.
- World Health Organization (2014). "World Health Statistics: Health related millennium development goals," World Health Organization, Geneva.

FIGURES AND TABLES

Figure 1: Provider Compliance with Clinical Practice Guidelines

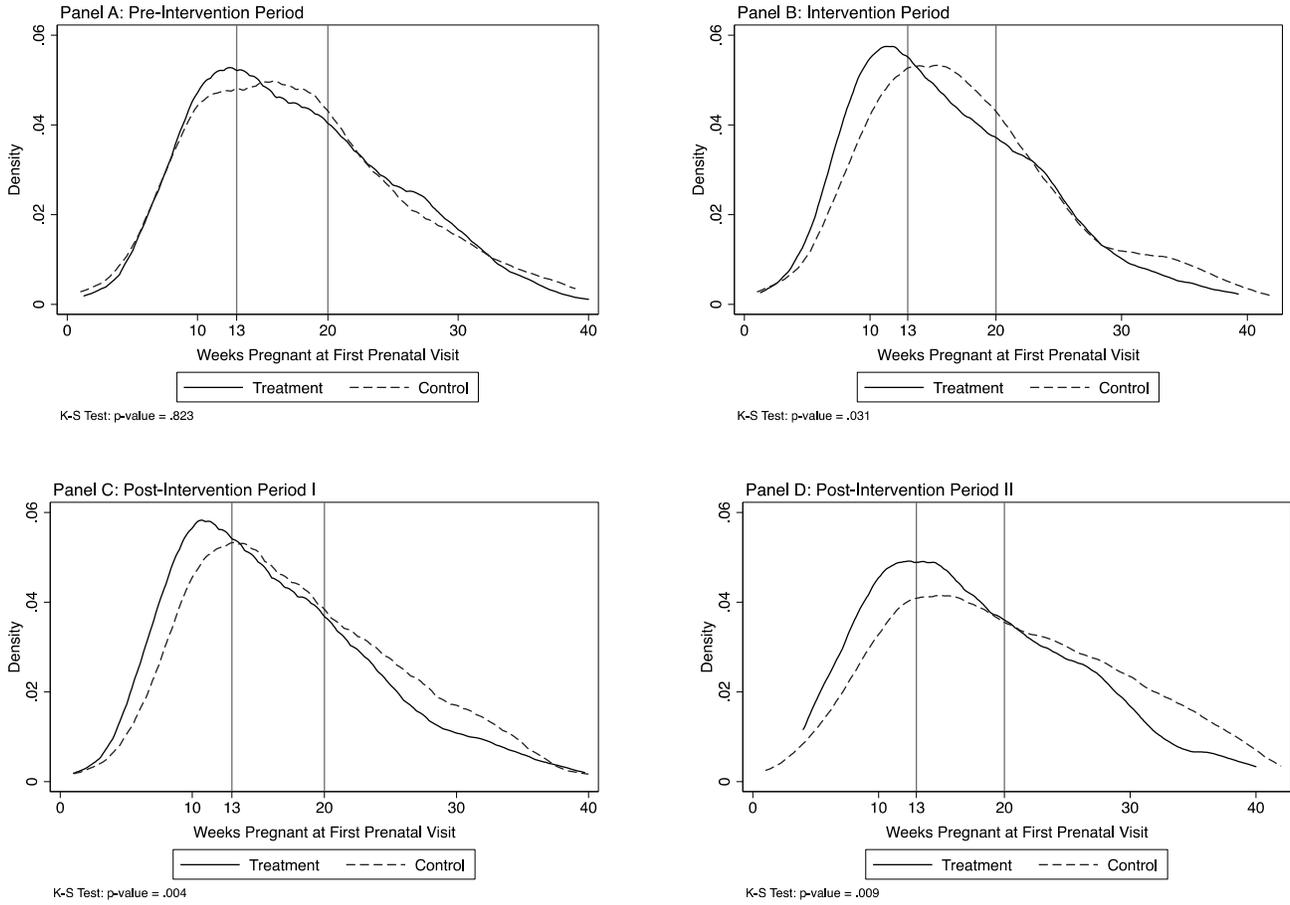


Source: Authors' elaboration based on (-) Schuster et al. (1998); (+) Grol (2001); (++) Campbell et al. (2007); (*) Das and Gertler (2007); and (#) Gertler and Vermeersch (2012).

Figure 2: Timeline and Data Availability

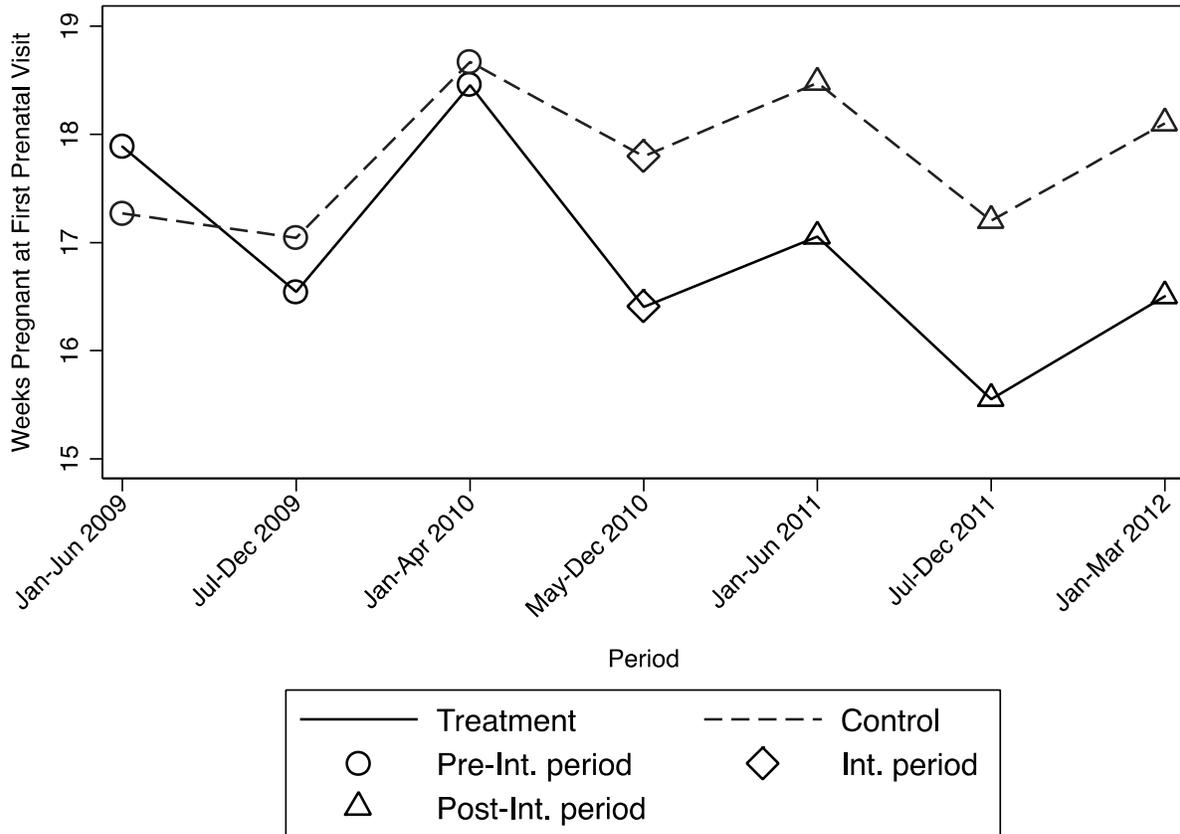
January 2009				December 2012
Pre-Intervention Period January 2009 – April 2010	Intervention Period May – December 2010	Post Intervention Period I Jan 2011 – March 2012	Post Period II April – Dec 2012	
Accurate Prenatal Care Data January 2009 – March 2012			Prenatal Care Data from new coding system	
Birth Outcome Data January 2009 – December 2011				

Figure 3: Densities of Weeks Pregnant at 1st Prenatal Visit



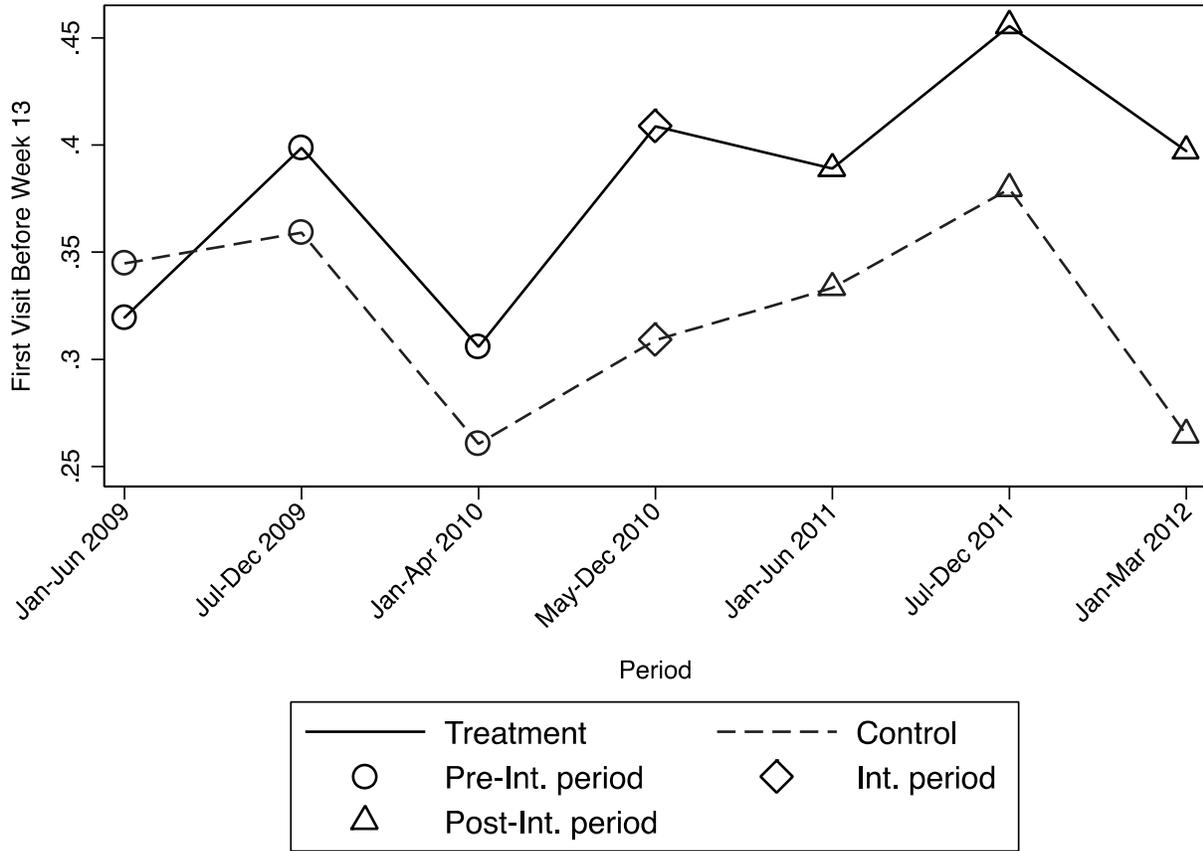
Notes: Densities estimated using an Epanechnikov kernel with optimal bandwidth. P-values of Kolmogorov-Smirnov tests of equality of distributions between groups reported below figure. The two vertical lines indicate weeks 13 and 20 of pregnancy. Source: Authors' own elaboration based on data from the provincial medical record information system.

Figure 4: Mean Number of Weeks Pregnant at 1st Prenatal Visit



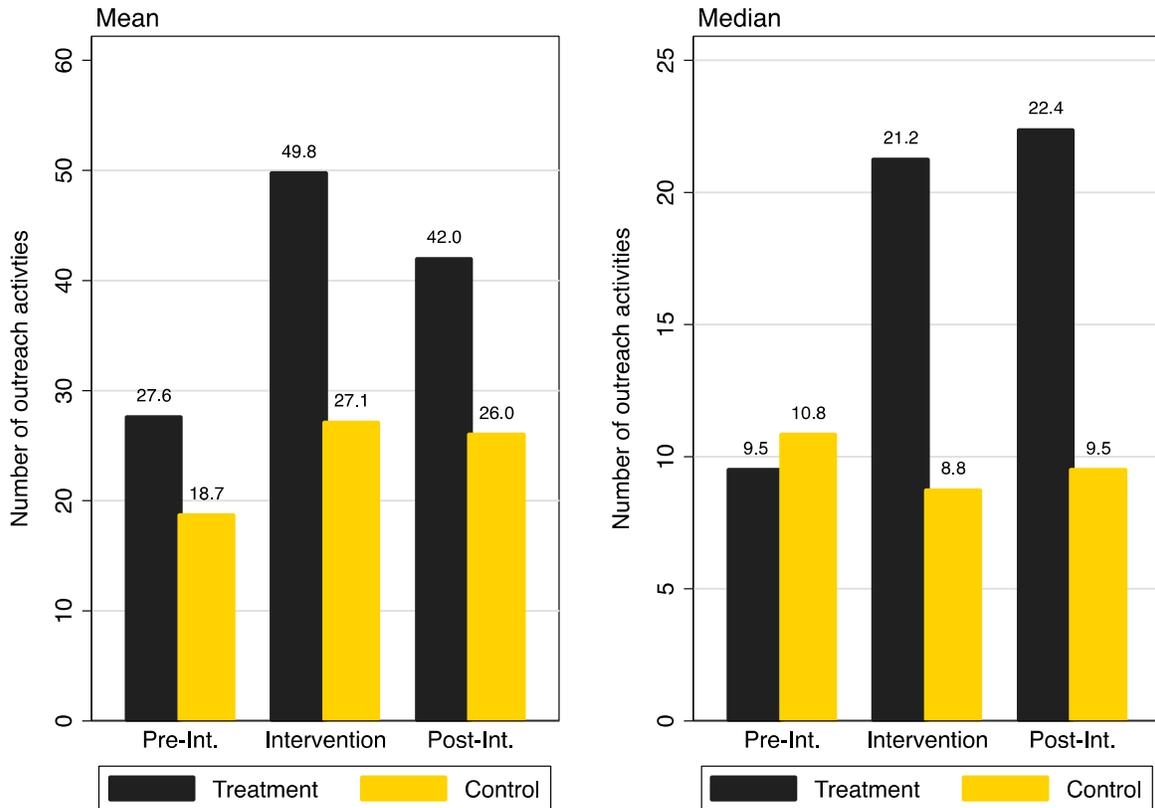
Notes: The first two points (circles) are means for 6-month periods prior to the intervention period. The third point (Diamond) corresponds to the intervention period. The fourth and fifth points (triangles) correspond to 6-months periods after the intervention period, while the last point (triangle) is for a 3-month period.

Figure 5: Proportion of Mothers with 1st Prenatal Visit before Week 13 of Pregnancy



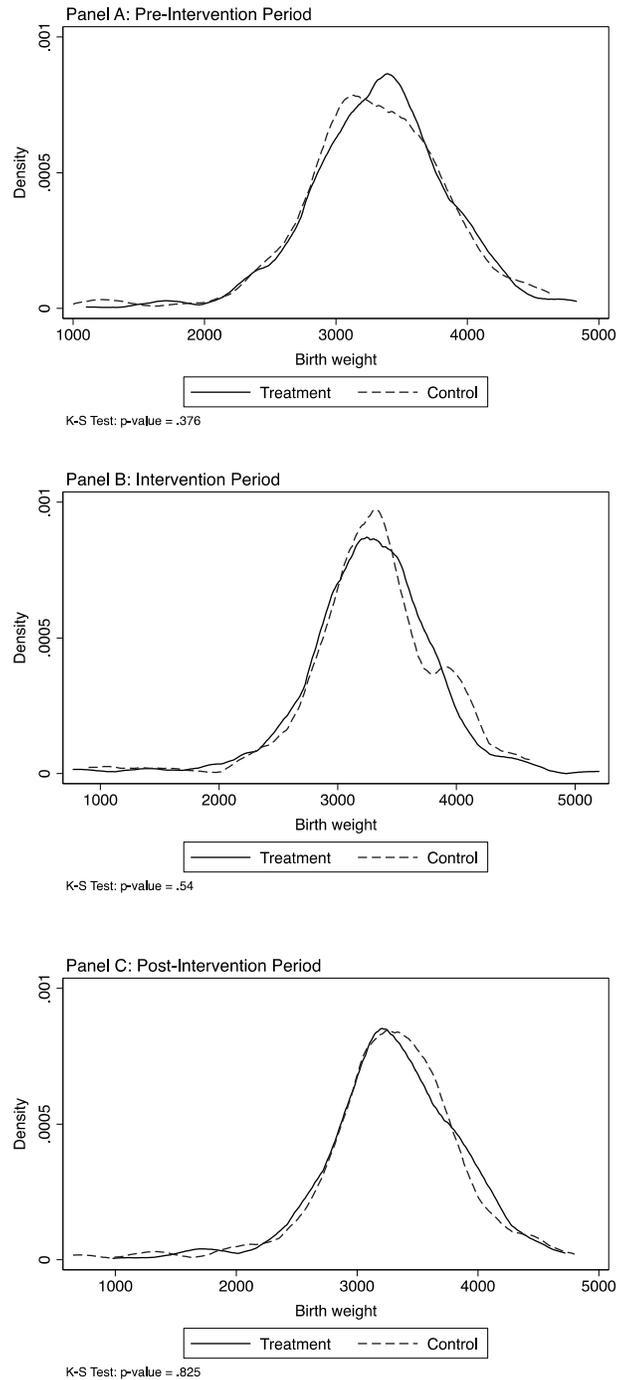
Notes: The first two points (circles) are means for 6-month periods prior to the intervention period. The third point (Diamond) corresponds to the intervention period. The fourth and fifth points (triangles) correspond to 6-months periods after the intervention period, while the last point (triangle) is for a 3-month period.

Figure 6: Number of Clinic Outreach Activities



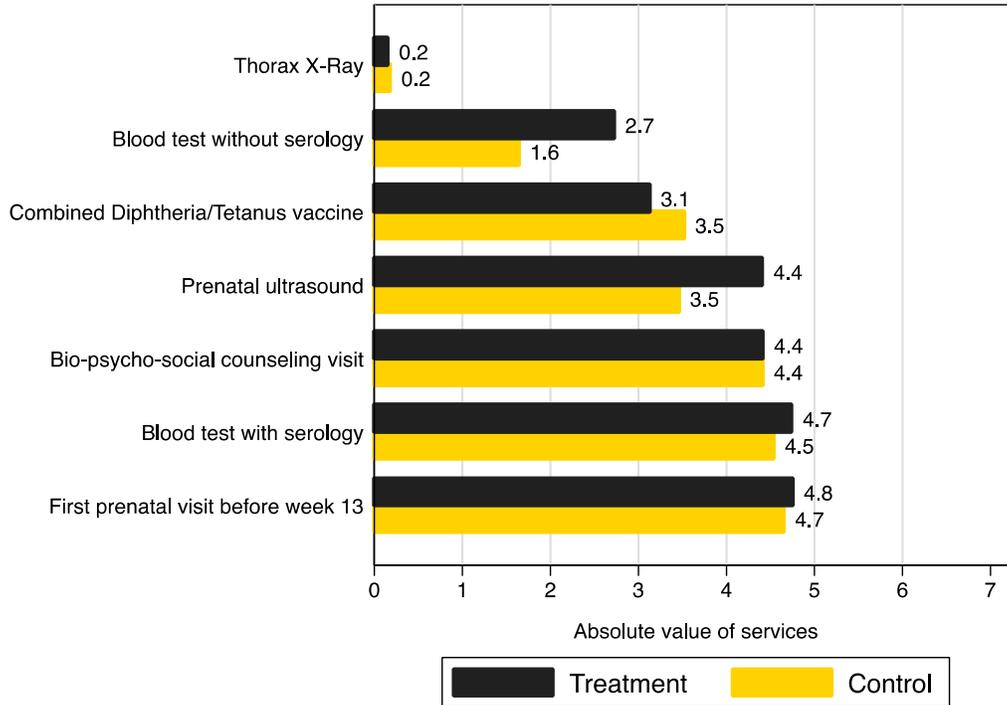
Notes: The height of the bars report the mean and median number of outreach activities that resulted in actual maternal-child service at the clinic, per trimester for the pre-intervention period (January 2009-April 2010), the intervention period (May-December 2010), and post-intervention period I (January 2011-March 2012)

Figure 7: Birth Weight Densities



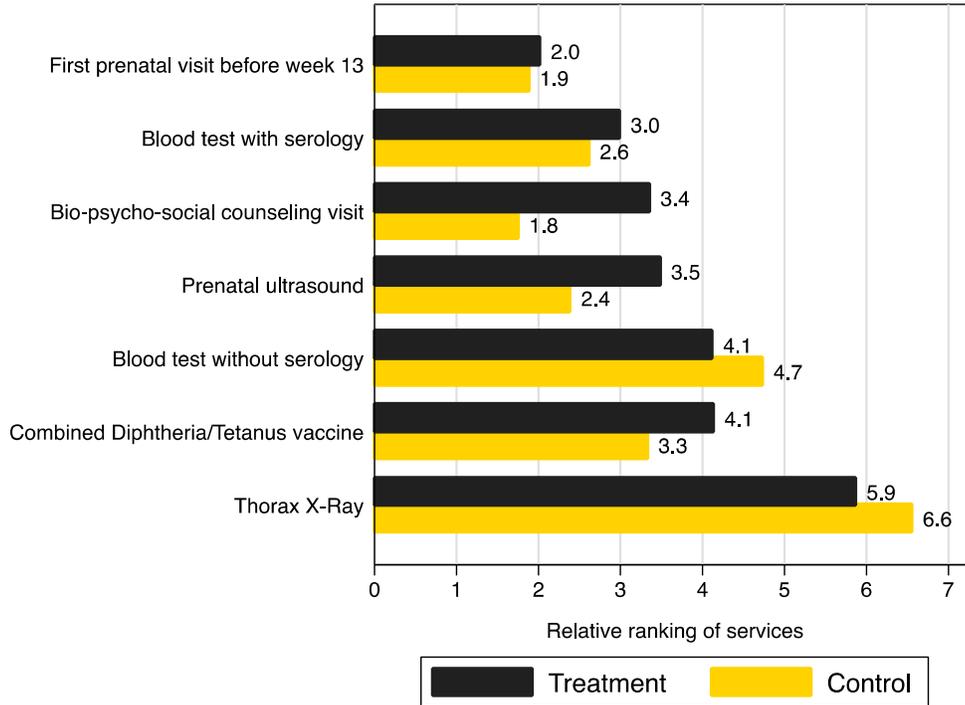
Notes: Densities estimated using an Epanechnikov kernel with optimal bandwidth. *P*-values of Kolmogorov-Smirnov tests of equality of distributions between groups reported below figure. Source: Authors' own elaboration based on medical record information system.

Figure 8: Absolute Score of Importance of Prenatal Care Services



Notes: This graph reports the average of the absolute score that measures the importance given by clinics to seven different prenatal care procedures including initiating prenatal care prior to week 13 of pregnancy. The data were collected using a short online survey conducted in the clinics that participated in the experiment. (see Appendix D) The absolute scores range from 1 to 5, with 5 being the highest score in terms of importance. The respond was coded zero if the respondent reported that this procedure is inappropriate for a pregnant woman.

Figure 9: Relative Ranking of Importance of Prenatal Care Services



Notes: This graph reports the average of the relative ranking that measures the degree of priority given by clinics to seven different prenatal care procedures including initiating prenatal care prior to week 13 of pregnancy. The data were collected using a short online survey conducted in the clinics that participated in the experiment. (see Appendix D) The relative scores aimed to rank the seven practices from 1 to 7, with 1 being the highest ranking. In practice however, the survey instrument allowed the respondent to repeat numbers.

Table 1: Payments for 1st Prenatal Visit

Time Period	Dates		Payment for 1 st Prenatal Visit	
	Begin	End	Before Week 13 of pregnancy	At week 13 of pregnancy or after
Pre-Intervention	January 2009	April 2010	\$ 40 ARS	\$ 40 ARS
Intervention	May 2010	December 2010	\$ 120 ARS	\$ 40 ARS
Post Intervention	January 2011	December 2012	\$ 40 ARS	\$ 40 ARS

Source: National Ministry of Health, Argentina (2010b)

Table 2: Clinic Assignment and Compliance Status

Assigned to Treatment	Actually Treated		Total
	Yes	No	
Yes	14	4	18
No	1	18	19
Total	15	22	37

Source: Authors' elaboration.

Table 3: Baseline Descriptive Statistics

	Assigned Treatment Group		Assigned Control Group		<i>p</i> -Value for test of equality of means	
	Mean (s.d.)	N	Mean (s.d.)	N	Large sample	Wild Boot-Strapped
Weeks Pregnant at 1 st Prenatal Visit	17.5 (7.48)	743	17.6 (7.74)	497	0.89	0.84
1 st Visit before Week 13 of Pregnancy	0.35 (0.48)	743	0.33 (0.47)	497	0.57	0.56
Tetanus Vaccine During Prenatal Visit	0.80 (0.40)	743	0.84 (0.37)	497	0.34	0.41
Number of Prenatal Visits	4.68 (2.94)	743	4.28 (2.77)	497	0.39	0.45
Birth Weight (grams)	3,328 (519)	552	3,291 (558)	379	0.36	0.37
Low Birth Weight (< 2500 grams)	0.06 (0.23)	552	0.06 (0.23)	379	0.96	0.98
Premature (gestational age < 37 weeks)	0.09 (0.29)	319	0.10 0.30	249	0.83	0.82
Maternal Age	25.36 (6.49)	354	25.75 6.10	270	0.47	0.48
Number of Previous Pregnancies	2.31 (2.39)	354	2.10 (2.10)	273	0.29	0.32
First Pregnancy	0.25 (0.43)	354	0.26 (0.44)	273	0.70	0.77

Notes: This table presents means and standard deviations in parentheses for the treatment and control groups during the 16-month pre-intervention period from January 2009 through April 2010. *P*-values for tests equality of treatment and control groups means are presented in the last 2 columns. We present both the *p*-value computed for large samples and a Wild bootstrapped *p*-value that is robust in samples with small numbers of clusters (Cameron et al. 2008). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (Davidson and Flachaire 2008) and uses 999 replications.

Table 4: Effects on Temporary Incentives on Timing of 1st Prenatal Visit

	(1)	(2)	(3)
	Intervention Period	Post-Intervention Period I (Jan 2011 – March 2012)	Post-Intervention Period II (April – Dec 2012)
A. Weeks Pregnant at 1st Prenatal Visit			
Treatment	-1.47** (0.71)	-1.63** (0.75)	-2.47** (1.02)
Large Sample p -value	0.04	0.03	0.02
Wild Bootstrapped p -value	0.08	0.03	0.03
Control Group Mean	17.80	17.90	20.10
Sample Size	769	1,296	710
B. First Prenatal Visit Before Week 13 of Pregnancy			
Treatment	0.11** (0.04)	0.08** (0.04)	0.08** (0.04)
Large Sample p -value	0.01	0.02	0.04
Wild Bootstrapped p -value	0.03	0.05	0.06
Control Group Mean	0.31	0.34	0.27
Sample Size	769	1,296	710

Notes: This table reports LATE estimates of the treatment effect of the modified fee schedule on indicators of the timing of the 1st prenatal visit. The differences are estimated from 2SLS regressions of the dependent variable on actual treatment status instrumented with clinic treatment assignment type. The p -values are for 2-sided hypothesis tests of the null that the difference is equal to zero. We present both the p -value computed for large samples and a Wild bootstrapped p -value that is robust in samples with small numbers of clusters (Cameron et al. 2008). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (Davidson and Flachaire 2008) and uses 999 replications. Column (1) reports the results for the sample observed in an 8-month intervention period (May 2010 – December 2010). Column (2) reports the results for the sample observed in the 15-month period following the end of the intervention (January 2011 – March 2012). Column (3) reports the results for the 9-month period after the change in the coding of the first prenatal visit (April 2012 – December 2012). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Impact on Log Number of Outreach Activities

	(1)	(2)
	Intervention Period	Post-Intervention Period I (Jan 2011 – March 2012)
Treatment	0.47** (0.23)	0.56** (0.22)
Large Sample <i>p</i> -value	0.04	0.01
Wild Bootstrapped <i>p</i> -value	0.04	0.02
Log (Control Group Mean)	1.93	1.93
Sample Size	324	324

Notes: This table reports LATE estimates of the treatment effect of the modified fee schedule. The dependent variable is the log of the number of clinic outreach activities that resulted in actual maternal-child service at the clinic per trimester. The *p*-values are for 2-sided hypothesis tests of the null that the difference is equal to zero. We present both the *p*-value computed for large samples and a Wild bootstrapped *p*-value that is robust in samples with small numbers of clusters (Cameron et al. 2008). Our Wild bootstrap assigns symmetric weights and equal probability after re-sampling residuals (Davidson and Flachaire 2008) and uses 999 replications. These are only computed for the coefficients of treatment interacted with each period. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Cross-Price Effects (Spillover)

	(1)	(2)
	Intervention Period	Post-Intervention Period I (Jan – Dec 2011)
A. Tetanus Vaccine		
Treatment	0.02 (0.08)	-0.02 (0.05)
Large Sample <i>p</i> -value	0.76	0.62
Wild Bootstrapped <i>p</i> -value	0.75	0.67
Control Group Mean	0.79	0.84
Sample Size	769	1,053
A. Number of visits		
Treatment	0.39 (0.33)	0.51 (0.58)
Large Sample <i>p</i> -value	0.24	0.38
Wild Bootstrapped <i>p</i> -value	0.27	0.41
Control Group Mean	4.05	4.40
Sample Size	769	1,053

Notes: This table reports LATE estimates of the treatment effect of the modified fee schedule on indicators of other services. The differences are estimated from 2SLS regressions of the dependent variable on actual treatment status instrumented with clinic treatment assignment type. The *p*-values are for 2-sided hypothesis tests of the null that the difference is equal to zero. We present both the *p*-value computed for large samples and a Wild bootstrapped *p*-value that is robust in samples with small numbers of clusters (Cameron et al. 2008). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (Davidson and Flachaire 2008) and uses 999 replications. Column (1) reports the results for the sample observed in an 8-month intervention period (May 2010 – December 2010). Column (3) reports the results for the sample observed in the 12-month period following the end of the intervention (January 2011 – December 2011). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Impact of Incentives on Birth Outcomes

	(1)	(2)
	Intervention Period	Post-Intervention Period I (Jan - Dec 2011)
A. Birth Weight		
Treatment	-37.34 (48.61)	25.109 (40.67)
Large Sample <i>p</i> -value	0.44	0.54
Wild Bootstrapped <i>p</i> -value	0.49	0.51
Control Group Mean	3,304	3,279
Sample Size	555	802
B. Low Birth Weight		
Treatment	0.01 (0.02)	-0.01 (0.02)
Large Sample <i>p</i> -value	0.63	0.60
Wild Bootstrapped <i>p</i> -value	0.61	0.56
Control Group Mean	0.05	0.06
Sample Size	555	802
B. Premature		
Treatment	0.03 (0.03)	-0.04 (0.02)
Large Sample <i>p</i> -value	0.31	0.08
Wild Bootstrapped <i>p</i> -value	0.28	0.12
Control Group Mean	0.09	0.12
Sample Size	414	708

Notes: This table reports LATE estimates of the treatment effect of the modified fee schedule for on indicators of birth outcomes. The observations include woman for whom we are able to obtain information on birth outcomes provided in public hospital birth records. The differences are estimated from 2SLS regressions of the dependent variable on actual treatment status instrumented with clinic treatment assignment type. The *p*-values are for 2-sided hypothesis tests of the null that the difference is equal to zero. We present both the *p*-value computed for large samples and a Wild bootstrapped *p*-value that is robust in samples with small numbers of clusters (Cameron et al. 2008). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (Davidson and Flachaire 2008) and uses 999 replications. Column (1) reports the results for the sample observed in an 8-month intervention period (May 2010 – December 2010). Column (2) reports the results for the sample observed in the 12-month period following the end of the intervention (January 2011 – December 2011). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

APPENDIX A: TEST OF MISREPORTING WEEKS PREGNANT AT 1ST PRENATAL VISIT

One concern is that the financial incentives may cause clinics to misreport the week of pregnancy at the first visit. In this appendix we report the results of test for this behavior. Recall that in our main analysis we construct the week of pregnancy at the first visit using the date of the first visit and the last menstrual date (LMD) as reported by the women. If the latter is not available we use the estimated date of birth (EDD) as recorded by the physician in the first visit. The EDD is calculated off the LMD as reported by the women during her first visit. While clinic medical records should contain both dates, about 10% of records are missing the LMD.

One possible way of misreporting the week of pregnancy at the first visit is to change the LMD and the EDD in the patient's clinical medical record. For instance, if a woman is in her 21st week of pregnancy at the first visit, the physician could add 7 days to the LMD and EDD so that the visit falls into the 20th week of pregnancy. Both would have to be changed in order to deceive the auditors.

To test for this possibility we use gestational age at birth (GAB) in weeks measured by physical examination at the time of birth, registered in the hospital medical record. We then compare the weeks elapsed from the first prenatal visit to the delivery date based on GAB to weeks elapsed from first visit to the delivery date based on EDD. While EDD is collected by the clinic who has an incentive to misreport, the GAB is collected by the hospital at time of delivery where there is no incentive to misreport.

Figure A1 plots the number of weeks to delivery from the time of the 1st visit based on GAB (y-axis) to the one based on EDD (x-axis). If there is no difference between the two measures, then all of the dates should fall on the 45-degree blue line. There should be some differences as EDD is an estimate that assumes no prematurity at birth, and there could be data entry in GAB and EDD and recall errors in EDD. Figure A1 shows that almost all of the data embrace the blue 45-degree line and most of the observations off the line are situated above it, consistent with prematurity explaining the differences.

If the clinic changes the EDD in order to capture higher payments, we would expect greater differences, for the treatment group, between GAB and EDD below the 12-week thresholds than above it during the intervention period when the incentives are in force, but no differences in the pre-intervention period. In order to test this, we estimate the following difference in difference regression:

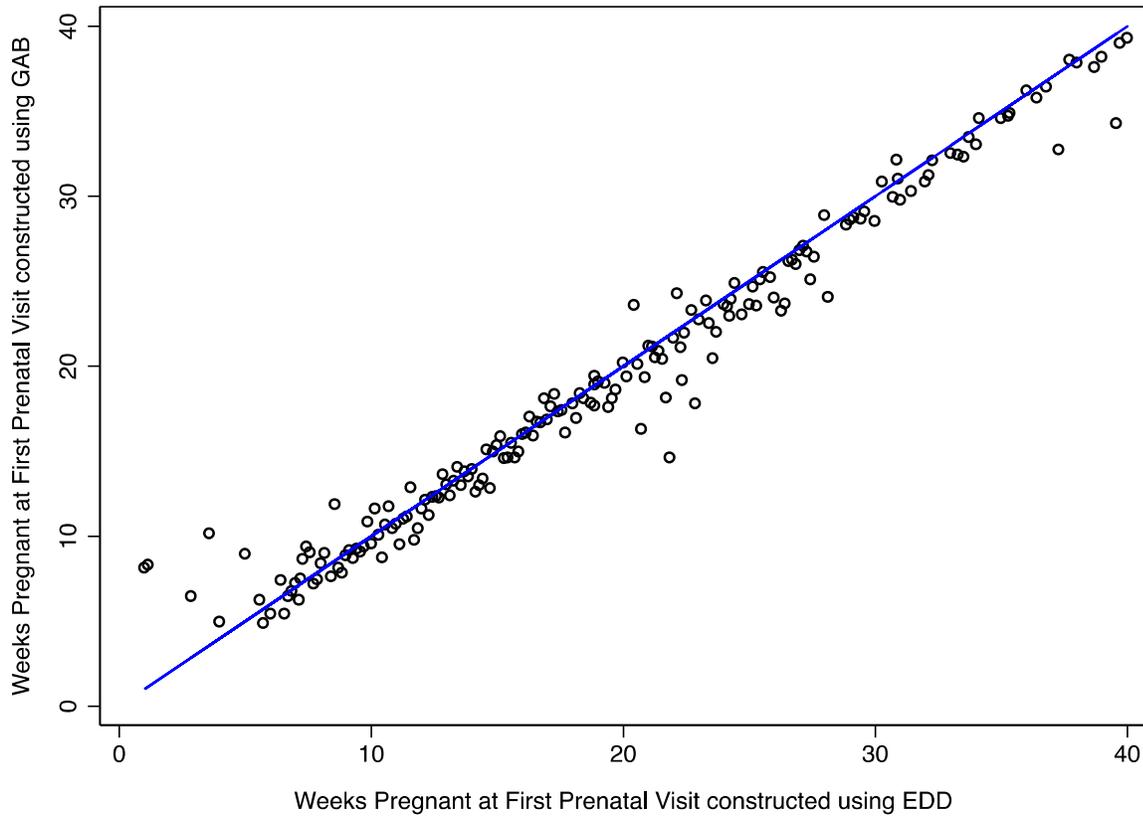
$$W_{ij}^{GAB} = \alpha_j + \beta W_{ij}^{EDD} + \gamma I(W_{ij}^{EDD} < 13) + \delta I(W_{ij}^{EDD} < 13)T_j + \varepsilon_{ij} \quad (A1)$$

where W_{ij}^{EDD} is weeks of pregnant at the first visit based on EDD for individual i getting care in clinic j , W_{ij}^{GAB} is the number of weeks at the first visit based on GAB for individual i getting care in clinic j , α_j is a clinic fixed effect, $I(W_{ij}^{EDD} < 13)$ is an indicator of whether the clinic reported the first visit to be in the first 12 weeks based on EDD, T_j is an indicator of whether the clinic was actually treated, and ε_{ij} is an error term.

In the absence of misreporting and no prematurity there should be no difference between the two measures and β would have a coefficient of 1. However, because premature births occur before EDD, we expect β to be close to but less than one. Then we can interpret the other coefficients as the effect on $W_{ij}^{GAB} - \beta W_{ij}^{EDD}$ accounting for average weeks of prematurity. So the dependent variable is the error in EDD in forecasting actual delivery date. Equation (A1) takes on a difference in difference interpretation in the sense the we are differencing the change in the forecast error between the pre-intervention and intervention periods for the group of pregnant women for which a clinic reports as having their first visit before 13 weeks and the group of pregnant women for which a clinic reports having the first visit in week 13 or later. If there is no difference in the error for the treatment group in the post period then δ , the interaction between treatment and reported having the first period before week 13, will be zero. We find no evidence of misclassification by treated clinics (See Table A1).

Figure A1

Comparison of Weeks Pregnant at 1st Prenatal Visit
Based on Gestational Age at Birth and Based on Date of Last Menstruation



Source: Authors' own elaboration based on data from the provincial medical record information system.

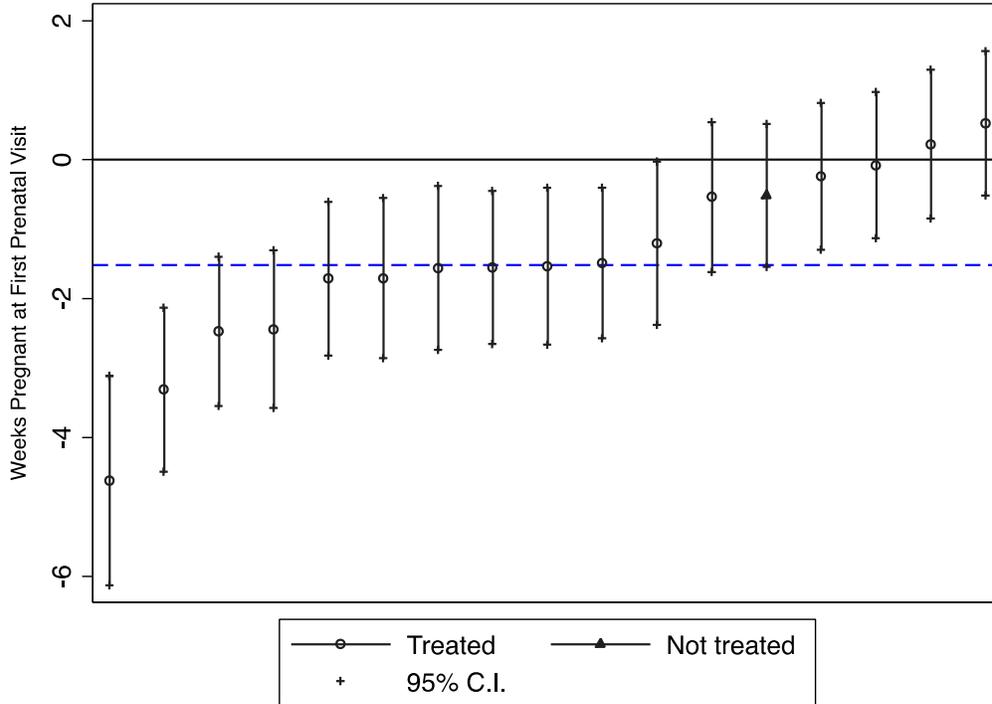
Table A1: Test for Misreporting Weeks Pregnant at 1st Prenatal Visit

Dependent Variable: Weeks Pregnant at 1 st Prenatal Visit, by Gestational Age at Birth	
Weeks Pregnant by EDD	0.90*** (0.02)
1(Weeks Pregnant by EDD<13)	-0.13 (0.31)
1(Weeks Pregnant by EDD<13) x 1(Treated=1)	-0.03 (0.44)
Constant	1.33*** (0.39)
Observations	1730
Adjusted R^2	0.82

The dependent variable is weeks pregnant at the first prenatal visit constructed using gestational age at birth. The independent variable is weeks pregnant at the first visit constructed by using the last day of menstruation or estimated delivery date (EDD). The interaction term interacts a dichotomous indicator for whether the visit was before week 13 and a dichotomous indicator for whether the clinic was actually treated. The regression controls for clinic fixed effects by adding a binary indicator for each clinic in the sample. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

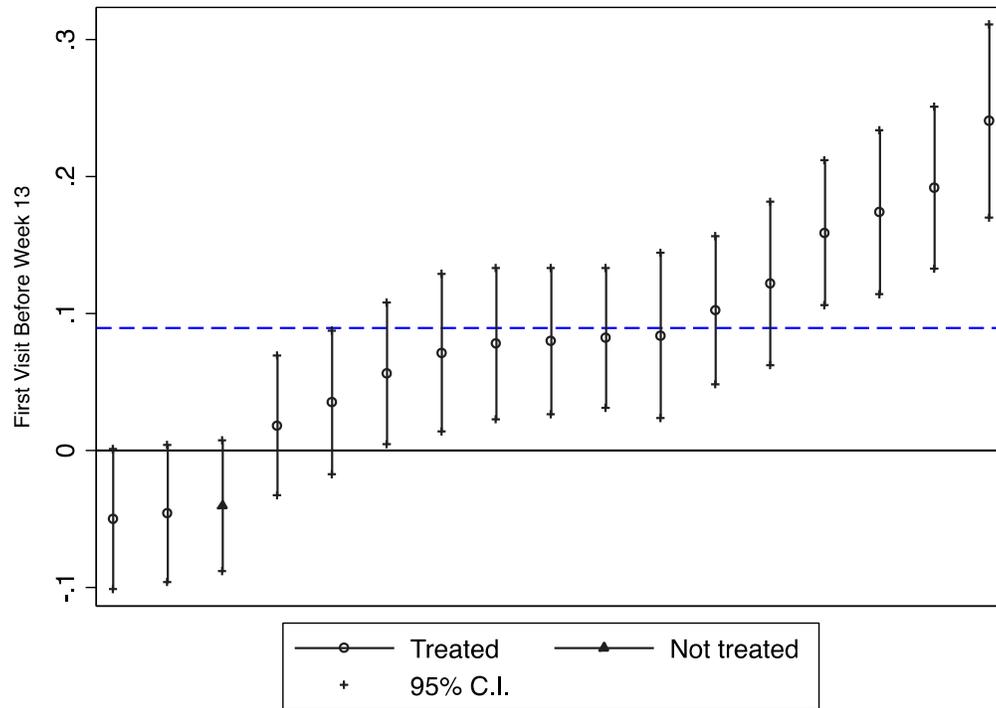
APPENDIX B: ROBUSTNESS TEST RESULTS

Figure B1: Individual Clinic Treatment Effects for Weeks Pregnant at 1st Prenatal Visit



Notes: This figure plots individual clinic treatment effects for the outcome of weeks pregnant at first prenatal visit. We run OLS regression of the outcome comparing each clinic assigned to the treatment group to all clinics assigned to the control group pooling the intervention period and the post-intervention period I (hence May 2010-March 2012). One treatment clinic is not included because of its insufficient sample size. This clinic corresponds to one of the two that did not take up treatment. The triangle symbol refers to the clinic that was assigned to treatment but did not take up the treatment. The x-axis is sorted from the lowest to the highest clinic-specific impact. The dashed blue line is the intent-to-treat effect calculated by pooling the intervention and the first post intervention period. The vertical lines are 95% confidence intervals constructed using standard errors obtained from the Wild bootstrap procedure.

Figure B2: Individual Clinic Treatment Effects for 1st Prenatal Visit before Week 13 of Pregnancy



Notes: This figure plots individual clinic treatment effects for the outcome of first prenatal visit before week 13. We run OLS regression of the outcome comparing each clinic assigned to the treatment group to all clinics assigned to the control group pooling the intervention period and post intervention period I (hence May 2010-March 2012). One treatment clinic is not included because of its insufficient sample size. This clinic corresponds to one of the two that did not take up treatment. The triangle symbol refers to the clinic that was assigned to treatment but did not take up the treatment. The x-axis is sorted from the lowest to the highest clinic-specific impact. The dashed blue line is the intent-to-treat effect calculated by pooling the intervention and the first post intervention period. The vertical lines are 95% confidence intervals constructed using standard errors obtained from the Wild bootstrap procedure.

Table B1: Robustness Tests for Weeks Pregnant at 1st Prenatal Visit

	(1)	(2)	(3)
	Intervention Period	Post-Intervention Period I (Jan 2011 – March 2012)	Post-Intervention Period II (April – Dec 2012)
A. Results from Table 4			
Treatment	-1.47** (0.71)	-1.63** (0.75)	-2.47** (1.02)
Large Sample p -value	0.04	0.03	0.02
Wild Bootstrapped p -value	0.08	0.03	0.03
Control Group Mean	17.80	17.90	20.10
Sample Size	769	1,296	710
B. Estimates Using Restricted Sample			
Treatment	-1.47* (0.77)	-2.01*** (0.70)	-2.01* (1.11)
Large Sample p -value	0.06	0.00	0.07
Wild Bootstrapped p -value	0.09	0.02	0.12
Control Group Mean	17.96	18.32	17.01
Sample Size	760	1,326	425
C. Difference-in-Differences Estimates			
Treatment	-1.35** (0.64)	-1.74*** (0.63)	-2.35* (1.31)
Large Sample p -value	0.036	0.005	0.072
Wild Bootstrapped p -value	0.060	0.014	0.144
Control Group Mean	17.80	17.90	20.10
Sample Size	4,015	4,015	4,015

Notes: This table reports LATE estimates of the treatment effect of the modified fee schedule on weeks pregnant at 1st prenatal visit. The p -values are for 2-sided hypothesis tests of the null that the difference is equal to zero. We present both the p -value computed for large samples and a Wild bootstrapped p -value that is robust in samples with small numbers of clusters (Cameron et al. 2008). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (Davidson and Flachaire 2008) and uses 999 replications. Column (1) reports the results for the sample observed in an 8-month intervention period (May 2010 – December 2010). Column (2) reports the results for the sample observed in the 15-month period following the end of the intervention (January 2011 – March 2012). Column (3) reports the results for the 9-month period after the change in the coding of the first prenatal visit (April 2012 – December 2012). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B2: Robustness Tests for 1st Prenatal Visit before Week 13

	(1)	(2)	(3)
	Intervention Period	Post-Intervention Period I (Jan 2011 – March 2012)	Post-Intervention Period II (April – Dec 2012)
A. Results from Table 4			
Treatment	0.11** (0.04)	0.08** (0.04)	0.08** (0.04)
Large Sample p -value	0.01	0.02	0.04
Wild Bootstrapped p -value	0.03	0.05	0.06
Control Group Mean	0.31	0.34	0.27
Sample Size	769	1,296	710
B. Estimates Using Restricted Sample			
Treatment	0.09** (0.04)	0.10** (0.04)	0.10* (0.06)
Large Sample p -value	0.03	0.01	0.08
Wild Bootstrapped p -value	0.08	0.02	0.11
Control Group Mean	0.31	0.33	0.36
Sample Size	760	1,326	425
C. Difference-in-Differences Estimates			
Treatment	0.09* (0.05)	0.07 (0.05)	0.07 (0.06)
Large Sample p -value	0.08	0.11	0.23
Wild Bootstrapped p -value	0.13	0.17	0.24
Control Group Mean	0.31	0.34	0.27
Sample Size	4,015	4,015	4,015

Notes: This table reports LATE estimates of the treatment effect of the modified fee schedule an indicator of whether the 1st prenatal visit occurred before week 13 of pregnancy. The p -values are for 2-sided hypothesis tests of the null that the difference is equal to zero. We present both the p -value computed for large samples and a Wild bootstrapped p -value that is robust in samples with small numbers of clusters (Cameron et al. 2008). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (Davidson and Flachaire 2008) and uses 999 replications. Column (1) reports the results for the sample observed in an 8-month intervention period (May 2010 – December 2010). Column (2) reports the results for the sample observed in the 15-month period following the end of the intervention (January 2011 – March 2012). Column (3) reports the results for the 9-month period after the change in coding of the first prenatal visit (April 2012 – December 2012). Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

APPENDIX C: ITT RESULTS

Table C1: ITT Estimates of the Effect of Temporary Incentives on Timing of 1st Prenatal Visit

	(1)	(2)	(3)
	Intervention Period	Post-Intervention Period I (Jan 2011 – March 2012)	Post-Intervention Period II (April – Dec 2012)
A. Weeks Pregnant at 1st Prenatal Visit			
Treatment	-1.39** (0.67)	-1.59** (0.73)	-2.47** (1.02)
Large Sample <i>p</i> -value	0.04	0.03	0.02
Wild Bootstrapped <i>p</i> -value	0.09	0.03	0.03
Control Group Mean	17.80	17.90	20.10
Sample Size	769	1,296	710
B. First Prenatal Visit Before Week 13 of Pregnancy			
Treatment	0.10*** (0.04)	0.08** (0.04)	0.08** (0.04)
Large Sample <i>p</i> -value	0.01	0.02	0.04
Wild Bootstrapped <i>p</i> -value	0.03	0.05	0.08
Control Group Mean	0.31	0.34	0.27
Sample Size	769	1,296	710

Notes: This table reports ITT estimates of the treatment effect of the modified fee schedule on indicators of the timing of the 1st prenatal visit. The LATE estimates are reported in Table 4. The differences are estimated from OLS regressions of the dependent variable on an indicator for clinic treatment random assignment. The *p*-values are for 2-sided hypothesis tests of the null that the difference is equal to zero. We present both the *p*-value computed for large samples and a Wild bootstrapped *p*-value that is robust in samples with small numbers of clusters (Cameron et al. 2008). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (Davidson and Flachaire 2008) and uses 999 replications. Column (1) reports the results for the sample observed in an 8-month intervention period (May 2010 – December 2010). Column (2) reports the results for the sample observed in the 15-month period following the end of the intervention (January 2011 – March 2012). Column (3) reports the results for the 9-month period after the change in the coding of the first prenatal visit (April 2012 – December 2012). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C2: ITT of Cross-Price Effects (Spillover)

	(1)	(2)
	Intervention Period	Post-Intervention Period (Jan – Dec 2011)
A. Tetanus Vaccine		
Treatment	0.02 (0.07)	-0.02 (0.05)
Large Sample <i>p</i> -value	0.76	0.62
Wild Bootstrapped <i>p</i> -value	0.80	0.59
Control Group Mean	0.79	0.84
Sample Size	769	1,053
A. Number of visits		
Treatment	0.37 (0.32)	0.50 (0.57)
Large Sample <i>p</i> -value	0.24	0.38
Wild Bootstrapped <i>p</i> -value	0.27	0.40
Control Group Mean	4.05	4.40
Sample Size	769	1,053

Notes: This table reports ITT estimates of the treatment effect of the modified fee schedule on indicators of other services. The LATE estimates are reported in Table 5. The differences are estimated from OLS regressions of the dependent variable on an indicator for clinic treatment random assignment. The *p*-values are for 2-sided hypothesis tests of the null that the difference is equal to zero. We present both the *p*-value computed for large samples and a Wild bootstrapped *p*-value that is robust in samples with small numbers of clusters (Cameron et al. 2008). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (Davidson and Flachaire 2008) and uses 999 replications. Column (1) reports the results for the sample observed in an 8-month intervention period (May 2010 – December 2010). Column (3) reports the results for the sample observed in the 12-month period following the end of the intervention (January 2011 – December 2011). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C3: ITT Effects of Incentives on Birth Outcomes

	(1)	(2)
	Intervention Period	Post-Intervention Period (Jan – Dec 2011)
A. Birth Weight		
Treatment	-34.88 (45.38)	24.48 (39.63)
Large Sample <i>p</i> -value	0.44	0.54
Wild Bootstrapped <i>p</i> -value	0.46	0.57
Control Group Mean	3304.82	3279.13
Sample Size	555	802
B. Low Birth Weight		
Treatment	0.01 (0.02)	-0.01 (0.01)
Large Sample <i>p</i> -value	0.63	0.60
Wild Bootstrapped <i>p</i> -value	0.61	0.63
Control Group Mean	0.05	0.06
Sample Size	555	802
B. Premature		
Treatment	0.03 (0.03)	-0.04* (0.02)
Large Sample <i>p</i> -value	0.31	0.08
Wild Bootstrapped <i>p</i> -value	0.32	0.09
Control Group Mean	0.09	0.12
Sample Size	414	708

Notes: This table reports ITT estimates of the treatment effect of the modified fee schedule for on indicators of birth outcomes. The LATE estimates are reported in Table 6. The observations include woman for whom we are able to obtain information on birth outcomes provided in public hospital birth records. The differences are estimated from OLS regressions of the dependent variable on an indicator for clinic treatment random assignment. The *p*-values are for 2-sided hypothesis tests of the null that the difference is equal to zero. We present both the *p*-value computed for large samples and a Wild bootstrapped *p*-value that is robust in samples with small numbers of clusters (Cameron et al. 2008). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (Davidson and Flachaire 2008) and uses 999 replications. Column (1) reports the results for the sample observed in an 8-month intervention period (May 2010 – December 2010). Column (2) reports the results for the sample observed in the 12-month period following the end of the intervention (January 2011 – December 2011). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

APPENDIX D: ONLINE SURVEY OF CLINICS

In collaboration with the Provincial Management Unit of the program (UGPS), in May 2015 we conducted a short online survey (using Survey Monkey®) in those clinics that participated in the pilot. The survey aims to measure the absolute and relative importance of seven different prenatal care procedures including initiating prenatal care prior to week 13 of pregnancy. The absolute scores range from 1 to 5, with 5 being the highest score in terms of importance, and an additional option of zero indicating that the procedure is not appropriate for a pregnant woman. Hence, the absolute score ranges from 0 to 5 points. The relative ranking aimed to sort the seven practices from 1 to 7, with 1 being the highest ranking. In practice however, the survey instrument allowed the respondent to repeat numbers.

The survey was sent out to by email to clinics directors (or the next person in rank). We were unable to obtain current email addresses for 8 out of the 36 clinics. Another 4 clinics confirmed having received the email but refused to answer it. Out of the 24 clinics that did respond to the survey, 21 fully completed it while 3 only partially completed it. Out of the 21 clinics with complete responses, 13 belong to the treatment group and 8 to the control group. Appendix Table D1 shows that there are no significant differences in baseline characteristics between clinics that responded to the survey and clinics that did not respond. In addition, we account for survey non-response using Inverse Probability Weighting based on the logistic regression reported in Table D2 (Wooldridge 2007). We report results for both IPW and non-IPW regressions.

Figures 8 and 9 do not suggest any difference in the absolute score and relative ranking of the procedures between treatment and control clinics. To test for the significance of the differences between the two groups, we run an OLS regression of the absolute score and the relative ranking against a binary indicator for treatment. To account for the small sample size we also compute the p -value for the differences in means permuting our data and using a random sample of 10,000 permutations. The results are shown in Tables D3 and D4.

Online Survey Questionnaire

We ask for your collaboration in completing a brief survey about prenatal care services provided at your health facility.

Important: When answering the survey, please think of a hypothetical case of a woman with the following characteristics:

- 25 years old
- Living in the same neighborhood where your health facility is located
- Without any apparent sign of disease
- 6 weeks pregnant
- Had a previous low-risk pregnancy

1. Please assign a score between 1 to 5 to each of the following services that could be delivered to the pregnant woman presented in the hypothetical case.

1 corresponds to a service to which you assign the lowest importance

5 corresponds to a service to which you assign the highest importance

	1	2	3	4	5	Not appropriate for a pregnant woman
Prenatal ultrasound	<input type="checkbox"/>					
Thorax X-Ray	<input type="checkbox"/>					
First prenatal visit before week 13 of pregnancy	<input type="checkbox"/>					
Bio-psycho-social pregnancy counseling visit	<input type="checkbox"/>					
Combined Diphtheria/Tetanus vaccine	<input type="checkbox"/>					
Blood test with serology	<input type="checkbox"/>					
Blood test without serology	<input type="checkbox"/>					

Please rank in order of priority (from 1 to 7) the following 7 health services that could be delivered to the pregnant woman of the hypothetical case.

1 corresponds to the service you would prioritize the most
7 corresponds to the service you would prioritize the least

Prenatal ultrasound	<input type="text"/>
Thorax X-Ray	<input type="text"/>
First prenatal visit before week 13 of pregnancy	<input type="text"/>
Bio-psycho-social pregnancy counseling visit	<input type="text"/>
Combined Diphtheria/Tetanus vaccine	<input type="text"/>
Blood test with serology	<input type="text"/>
Blood test without serology	<input type="text"/>

Table D1: Baseline Characteristics of Clinics, by Online Survey Response Status

	Non-respondent	Respondent	<i>P</i> -value	Obs.
Number of Pregnant Women Attended per Year	48.60	64.90	0.33	36
Weeks Pregnant at 1 st Prenatal Visit	17.44	16.77	0.15	36
1 st Visit before Week 13 of Pregnancy	0.34	0.38	0.27	36
% of Pregnant Women who are Plan Nacer Beneficiaries	0.61	0.64	0.59	36
Tetanus Vaccine During Prenatal Visit	0.74	0.81	0.22	36
Number of Prenatal Visits	4.26	4.42	0.72	36
Birth Weight (Grams)	3,283	3,320	0.33	36
Gestational Age (Weeks)	38.65	38.47	0.57	31
Low Birth Weight (< 2500 Grams)	0.06	0.07	0.73	31
Premature (Gestational Age < 37 Weeks)	0.10	0.13	0.60	31

Notes: This table reports the means of baseline characteristics for clinics that responded to the May 2015 online survey and for clinics that did not respond. The characteristics are taken from the medical records information system (2009). The *p*-values for the tests of differences in means are computed using permutation tests that are robust for small sample sizes.

Table D2: Probability of Responding to the Online Survey,
Logit Coefficients and Marginal Effects

	Coeff.	Marg. Eff.
Treatment Group	1.498 (1.111)	0.274 (0.180)
Birth Weight (grams)	0.100 (1.076)	0.018 (0.196)
Weeks Pregnant at 1 st Prenatal Visit	-0.594 (0.648)	-0.109 (0.121)
1 st Visit before Week 13 of Pregnancy	-3.590 (9.026)	-0.657 (1.670)
% of Pregnant Women who are Plan Nacer Beneficiaries	1.620 (4.359)	0.296 (0.774)
Tetanus Vaccine During Prenatal Visit	3.350 (3.817)	0.613 (0.646)
Number of Prenatal Visits	-0.099 (0.559)	-0.018 (0.101)
Constant	7.644 (18.248)	
Observations	36	36

Notes: This table reports the coefficients and marginal effects from a logit regression that estimates the probability that a clinic responded to the May 2015 online survey.

Table D3: Differences in Absolute Score and Relative Ranking of Early Prenatal Care

	Absolute Score		Relative Ranking	
	(1) OLS	(2) OLS-IPW	(3) OLS	(4) OLS-IPW
Difference (Treatment – Control)	0.20 (0.22)	0.13 (0.92)	0.10 (0.21)	0.14 (0.89)
Large Sample <i>p</i> -value	0.38	0.89	0.65	0.88
Permutation <i>p</i> -value	0.35	1.00	0.46	0.99
Observations	20	20	20	20
Control group mean	4.57	1.88	4.66	1.88

Notes: Column (1) shows the differences between treatment and control clinics in the absolute score assigned to the practice of early prenatal care without any adjustment of sample loss. Column (2) adjusts for sample loss by Inverse Probability Weighting. Column (3) shows the differences between treatment and control clinics in the relative ranking assigned to early prenatal care among seven different practices. Column (4) is the same as Column (3) but adjusts for sample loss by Inverse Probability Weighting. (Wooldridge 2007) The coefficients are obtained from an OLS regression of each outcome against a treatment binary indicator. The third row shows the *P*-value obtained from permuting the data using a random sample of 10,000 permutations. Standard errors are in parentheses. We lose one observation in each case because of missing data in each specific question.