NUDGING MEDICAL PROVIDERS TO ADOPT AND SUSTAIN BETTER QUALITY CARE PRACTICES

By PABLO CELHAY, PAUL GERTLER, PAULA GIOVAGNOLI AND CHRISTEL VERMEERSCH*

February 21, 2017

Abstract: *We show that costs of adjustment as opposed to low returns likely explain why better quality care practices diffuse slowly in the medical industry. Using a randomized field experiment conducted in Argentina, we find that temporary financial incentives paid to health clinics for the early initiation of prenatal care 'nudged' providers to test and develop new data driven strategies to locate and encourage likely pregnant women to seek care in the first trimester of pregnancy. These innovations raised the rate of early initiation of prenatal care by 34% while the incentives were being paid in the treatment period. We follow health clinics over time and find that this increase persisted for at least 24 months after the incentives ended. In the absence of incentives, even though it is in the clinics' interest to stimulate early initiation of care, the presence of hard to change habits and cost of experimentation made it too expensive to develop and implement new methods to increase early initiation of care. Despite the large increases in early initiation of prenatal care, we find no effects on health outcomes. (JEL I12, I13, I15, I18)*

I.   Introduction

A well-documented feature of technological change is its remarkably slow diffusion.[1] One reason could be that innovation may have large costs of adjustment above and beyond acquisition expenditures. Firms have to design, test, and learn how to best incorporate an innovation into existing practices.   Firms may also need to purchase complementary technology (Rosenberg 1982, David 1990, Bresnahan and Trajtenberg 1995). Productivity might also be lower during a period of adjustment while the firm learns how best to use the innovation and the marginal productivity of the new technology. Management may have to overcome costly informational deficits (Bloom et al. 2012) and behavioral barriers such as present bias (Duflo et al. 2011).   Innovation may confront worker's resistance if part of their wage varies with performance (Lazonick 1979, Atkin et al. 2015) or if they have developed strong work habits. Though it may seem that such costs of adjustment are easy to overcome, they can be large enough to prevent productive and profitable innovations.[2] Change is hard, and even small costs of adjustment may inhibit changes in favor of maintaining the status quo (DellaVigna 2009; Thaler and Sunstein 2009).

Slow diffusion of better quality medical care practices is a global issue as evidenced by the remarkably low level of compliance with state of the art Clinical Practice Guidelines (CPGs) (Figure 1).[3] While low CPG compliance may in part reflect a lack of knowledge or slow diffusion of information (Phelps 2000), 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, Cabana et al. (1999) report that psychological costs such as resistance to changing

---

[1] Slow adoption of new technologies by firms has been extensively documented in agriculture, manufacturing and medicine. Surveys and studies of slow diffusion include Ryan and Gross (1943), Griliches (1957), Mansfield (1961), Coleman and Menzel (1966), Rosenberg (1972), Parente and Prescott (1994), Foster and Rosenzweig (1995), Geroski (2000), Hall and Khan (2003), Hall (2005), Comin and Hobijn (2010), Conley and Udry (2010).

[2] The organizational literature refers to the phenomenon of high fixed costs of preventing the adopting profitable innovations as organizational inertia (Hannan and Freeman 1984; Carroll and Hannan 2000).

[3] 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.

[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 et al. (2015).

existing practice patterns are one of the most important barriers to CPG adherence.[5]   For example, Grol and Grimshaw (2003) report that 49% of UK nurses and doctors said that resistance to changing old habits was a major obstacle to complying with new hand hygiene guidelines. In a recent study of U.S. hospitals, Skinner and Staiger (2015) state that the large differences observed in the adoption of aspirins and beta-blockers are in part due to resistance to changing old habits of physicians.[6]

In this paper we examine the short-term and long-term effects of paying temporary financial incentives to medical care providers in Argentina to increase the share of women who initiated prenatal care in their first trimester of pregnancy. Before the intervention, the share of women who initiated care early was low and medical care clinics had no specific activities or practices devoted to identifying and encouraging pregnant women to start care early. Providers began prenatal care whenever women chose to first show up at the clinic.

We show that the temporary incentives motivated clinics to invest time and effort to develop and test new data driven methods of how to best identify newly pregnant women and encourage them to seek care early. In other words, clinics innovated in the sense that they had to experiment with new outreach strategies in order to learn the productivity of alternative strategies and what worked best. The temporary incentives nudged providers to change their old medical care practice pattern habits. The temporary fee increase compensated providers for the costs of testing new outreach strategies as well as the costs of overcoming psychological barriers of changing practice patterns and uncertainty about the new service.

Given the well-known market failures in health care, a major policy issue facing both public and private health care systems is how to improve access to higher quality medical care (Chaudhury et al. 2006; Das et al. 2008). Both public (government) and private (insurance plans) payers are turning to pay for performance incentive mechanisms to encourage better quality care (Eldridge et al. 2009; Miller and Babiarz 2014). How best to use these incentives, however, depends on whether slow adoption of better quality care practices is due to low perceived value or to high costs of adjustment. If providers value a new clinical service but are

---

[5] For more evidence of resistance to change 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] For other references on technology adoption in the medical care industry see Baker (2001), Baker and Phibbs (2002), Berwick (2003), Cutler and Huckman (2003), Cutler (2007), and Bech et al. (2009).

reluctant to adopt because of high costs of adjustment, then a temporary incentive large enough to cover these costs should nudge providers to adopt the service and continue it after the incentive is removed.[7] If on the other hand, costs of adjustment are low but providers have a low perceived value of the service, then the price increase should also lead to adoption while the increase is active, but the service will be dropped once the incentive disappears. Hence, studying provider behavior both while incentives are active and after the incentives are removed provides a test of whether costs of adjustment versus low returns are inhibiting innovation.

That early initiation of prenatal care is an important priority is common wisdom and widely disseminated in the medical professional. Early initiation of prenatal care has long been part of the Argentine CPGs for prenatal care, and is part of standard training in Argentine medical and nursing schools as well as throughout the world (WHO 2006). Providers are taught that prenatal care by skilled health professionals beginning in the first trimester of pregnancy is essential for good maternal and newborn health outcomes as it is argued in the medical literature (Schwarcz et al. 2001; Carroli et al. 2001a; Campbell and Graham 2006). Through early initiation of care, providers are able to detect and correct important medical conditions such as maternal infections or anemia in the period in which the fetus is most at risk and before these conditions can jeopardize maternal or newborn outcomes (Carroli et al. 2001b; Hawkes et al. 2013). Early prenatal care also allows providers to advise mothers on proper prenatal nutrition and prevention activities in the period in which the fetus is developing most rapidly.

There is also strong empirical support for the relationship between early initiation of prenatal care and birth outcomes. Rosenzweig and Schultz (1983) and Grossman and Joyce (1988) show that delaying the start of medical care while pregnant significantly reduces birth weight. Evans and Stech-Lien (2005) further show that missing medical visits in the early stages of the pregnancy significantly reduces birth weight, while missing visits at later periods has no effect.

---

[7] In practice, this amounts to paying providers a time-limited per unit incentive for the new service. 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.

We use data from a field experiment conducted with *Plan Nacer,* an Argentine government program similar to Medicaid in the U.S. and Seguro Popular in Mexico that provides health insurance to otherwise uninsured pregnant women and children.[8] The field experiment randomized temporary financial incentives to health care clinics in which treatment clinics were paid a 200% premium for the first prenatal care visit if the visit occurred before the 13th week of pregnancy. The fee increase was paid for 8 months and then removed. Clinics were explicitly made aware that the fee was temporary.

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) and that the average weeks pregnant at the time of the first prenatal care visit fell by about 1.5 weeks while the incentives were being paid. We then show that that the higher levels of early initiation of prenatal care in the treatment group persisted for at least 24 months after the incentives ended.

We also document that clinics developed specific data driven strategies to find newly pregnant women and encourage them to start care early. Clinics designed and tested new beneficiary outreach strategies such as (i) coordinating with local pharmacies to keep track of women who stopped using birth control pills, (ii) meeting with teenagers while parents were less likely to be at home so that they would be more prone to reveal pregnancies, (iii) talking with mothers when they come to pick free milk for children and (iv) modifying gynecologist schedules to be able to more easily make appointments. These new strategies took time to develop and test, and involved opportunity costs to clinical staff beyond marginal costs of actual implementation. We show that all outreach activities doubled in the treatment group relative to the control group during the intervention period, and that this increase persisted at least 15 months after the incentives ended.

We also provide evidence that it was in the clinics' interest to have provided these outreach services absent the costs of adjustment. First, in a survey discussed later, clinic medical directors reported that early initiation of care was ranked as one of the highest of health priorities among all prenatal care services. Second, *Plan Nacer* reimbursed clinics for beneficiary outreach activities at a rate higher than the cost of delivering those activities.

---

[8] In 2013, *Plan Nacer* was expanded to other populations and renamed Programa Sumar.

Finally, prenatal care visits by *Plan Nacer* beneficiaries were profitable to clinic staff since 50% of the fees obtained from prenatal visits were used to pay wage bonuses.

Despite the fact that the incentives succeeded in inducing clinics to innovate and develop new outreach activities that were effective in increasing early initiation of prenatal care, the incentives had no ultimate effect on birth weight. This is likely due to the fact that early initiation of care may have come from primarily low-risk mothers who are less likely to benefit from early initiation of care. Indeed, one would think that it would be easier to persuade low-risk mothers to come a little 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. If the intervention had induced high-risk women who otherwise would have had their 1st visit much later in the pregnancy, then the incentives might have had a measurable impact on birth outcomes. One solution then would be to condition the incentives on the early initiation of high-risk women, but risk is difficult and expensive to identify and verify, and therefore may not be contractible.

Taken together these results are consistent with the presence of costs of adjustment as opposed to low perceived value inhibiting the diffusion of better quality of care practices. First, early initiation of care was both profitable at the lower fees and perceived to be important for health outcomes; yet, absent the incentives it was being provided at a low rate. Second, temporary incentives lead to the development of new outreach activities designed to identify and encourage pregnant women to seek care early resulting in a large and significant increase in the early initiation of care that persisted long after the incentives ended. Third, the new outreach activities were in and of themselves profitable.

Our results suggest that costs of adjustment maybe the mechanism behind recent evidence that permanent performance incentives improve access to higher quality medical care.[9] The standard explanation is that providers are reallocating their effort across services in response to the increased profit opportunities.[10] However, previous studies have been unable to distinguish between this mechanism and a cost of adjustment story. We are able to

---

[9] 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 Babiarz (2013) provide a review.

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

distinguish between the two mechanisms by observing 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 standard models but continue at the higher levels under the costs of adjustment model. Understanding the mechanism by which financial incentives work is 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 long-term boost in performance at a lower cost than permanent incentives.

Temporary incentives for technology adoption have been rarely studied.[11] Notable exceptions include Atkin et al. (2015) who examine how temporary financial incentives can overcome workers resistance to adopt a new and more efficient technology to produce soccer balls. The authors find that initial slow-downs in productivity from learning a new production process inhibited the adoption of the technology in firms where workers where compensated for performance. The results show that a short run financial incentive large enough to compensate workers for their short-term loss generated long run gains in productivity. In another related paper, Duflo et al. (2011) study the effect of providing short-term subsidies to purchase new more effective fertilizer by small farmers. The authors argue that even though new fertilizer is highly profitable, there might be important behavioral barriers and direct costs that inhibit their adoption. They show that small and temporary subsidies generated large increases in adoption, especially among impatient farmers. Finally, Bloom et al. (2012) show that management practices explain a great part of the differences in productivity among Indian firms in the textile industry. They show that providing managers with free short-term consulting on better management skills can create large gains in productivity in the long run.

The paper is organized as follows. Section II describes a simple model of technology adoption under fixed costs. Section III describes the intervention and the experimental design. Section IV describes the data. Section V explains the identification strategy and estimation methods. Section VI shows our main results on different outcomes and discusses the main

---

[11] There is also work on temporary incentives in the form of sales and coupons to market products—e.g. Blattberg and Neslin (1990); Kirmani and Rao (2000); and Dupas (2014). Similarly, there is a literature on the effect of temporary incentives for individuals to develop better health habits such as exercise and quitting smoking-- e.g. 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).

mechanism to explain the effects we find as well as alternative explanations to our results. Section XI and section XII discusses spill over effects and effects on birth outcomes, respectively. Section XIII concludes.

II.   Conceptual Framework

We develop a stylized model where clinics incur in different costs of adjustment to change clinical practice patterns. We consider 3 types of costs of adjustment: (i) monetary costs such as the purchase of complementary expenditures on design and testing, purchases of complementary technology and reduced productivity during a period of learning; (ii) psychological costs such as present bias and worker resistance to change; and (iii) uncertainty about the true marginal productivity and hence profitability of the innovation to the firm. We use this model to simply illustrate there are costs of adjustment when firms decide to adopt a new service into their set of practice patterns. 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 $\alpha \geq 0$ is the provider's intrinsic value of a unit of patient health.[12]   When α takes on value 0, the clinic is purely extrinsically motivated and as α rises the clinic is willing to sacrifice more income for patient health. 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.[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

$$H = \lambda_1 S_1 + \lambda_2 S_2 + \varepsilon \tag{1}$$

where ε is a mean zero random shock.

---

[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 Clemenes and Gotlieb (2014).

[13] Without some fixed costs of adjustment, both intrinsically and extrinsically motivated providers would still operate at the efficient frontier.

Clinical Practice Patterns-- Consider a clinic whose current clinical practice pattern is to provide S_1 to all patients. In this case, S_1 is the clinic's existing clinical practice pattern, 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 then it must overcome different barriers or costs. Clinics may have to incur an upfront one-time fixed cost *F.* Fixed costs include designing, testing, and learning how to best incorporate the delivery of the service into existing practice patterns, retraining, purchase of complementary medical equipment, and reduced productivity during a period of adjustment. Clinics have to learn how to best implement the new practice and hence they may be risk-averse towards it in that the marginal costs ex ante are not known or known imperfectly. Moreover, clinics may be present biased in that they may discount future profits highly. We incorporate these different costs to adoption into the profit function of each clinic.

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

$$\pi = \sum_{t=1}^{\infty} \beta^t \big[(p_1 - c_1) + EU_{c_2}(p_2 - c_2)S_2\big]N - FS_2, \tag{2}$$

where β is the clinic's discount rate. The discount rate may in part reflect present bias and psychological resistance to change; discounting future returns to an innovation at a higher rate thereby lowering the present value of an innovation. In this equation, c_2 is the stochastic component of the profit function, where the expected utility of providing S_2 is $EU_{c_2}(p_2 - c_2) = \int u(p_2 - c_2)dG(c_2)$, $u(\cdot)$ is a non-decreasing concave function, and c_2 is distributed according to $G(\cdot)$. To illustrate a measure of relative risk aversion we let $G(\cdot)$ be a mean preserving spread of another distribution $F(\cdot)$.[14]

Adoption-- The clinic adopts S_2 if

$$R(S_2 = 1) - R(S_2 = 0) \geq 0 \quad . \tag{3}$$

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

---

[14] It can be shown that since $G(\cdot)$ is a mean preserving spread of another distribution $F(\cdot)$, $U(F) = \int u(p_2 - c_2)dF(c_2) \geq \int u(p_2 - c_2)G(c_2) = U(G)$. Since $G(\cdot)$ is riskier than $F(\cdot)$, the latter is preferred for every risk averter (Mas-Colell, Winston, and Green, 1995).

$$\sum_{t=0}^{\infty} \beta^t \big[EU_{c_2}(p_2 - c_2) + \alpha\lambda_2 \big] N - F \geq 0 \tag{4}$$

Clinics are more likely to adopt S_2 if the profit margin from S_2 is higher, they have higher patient volumes, and they have lower discount rates. Clinics who are more intrinsically motivated (i.e. higher α) are also more likely to adopt 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). Finally, the probability of adopting S_2 decreases with lotteries of c_2 that are of higher risk or more uncertain.

Temporary Incentives-- Without loss of generality we can simplify the model to 2 periods with β as the discount rate. In this case, based on (3), an incentive θ that compensates for the utility differential of adopting the new service in period 1 is:

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

The temporary incentive, θ, at minimum covers the remainder of the cost of adjustment that is not paid by the discounted present value of the future stream of expected surplus generated from the provision of S_2. The incentive goes down with scale N, the expected profit margin $EU_{c_2}(p_2 - c_2)$, increases with uncertainty about c_2, and decreases with 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 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.

III. Context and 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 (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. Revenues from *Plan Nacer* are on top of clinic budgets that cover salaries as well as medical and non-medical supplies and materials. In practice, *Plan Nacer* payments top up these budgets by 5 to 7%. Health facilities are free to choose how to use realized revenues within relatively broad guidelines, and in Missiones clinics can and do use 50% of the *Plan Nacer* payments to pay bonuses to clinic staff. In this sense, all services, including prenatal care visits, provided to *Plan Nacer* beneficiaries are in the interest of clinic staff as 50% of the payment is used to increase staff bonuses.

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 identifies eligible women and children in order to enroll them into the program. Enrollment activities usually consist on door-to-door visits across a determined geographic area assigned to each clinic and defined by the Province.

Clinics can only provide services to the population within their area and enrolled beneficiaries can only obtain care from their assigned clinic. Outreach staff regularly contact beneficiaries to encourage them to take advantage of program benefits. *Plan Nacer* reimburses clinics for all outreach activities to the beneficiary population at a rate higher than the clinic's cost of outreach.

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, each 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 of pregnancy.

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] At this initial price prenatal care visits were profitable as 50% of this fee was used to increase staff bonuses. 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. Every other component of the *Plan Nacer* program remained the same. 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.

---

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

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. 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.

IV.   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 each service is provided; therefore we do not need to rely on maternal recall of these variables usually collected by surveys long after the visit. The data used in the analysis are extracted from individual clinical 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, each observation in our sample corresponds to a unique pregnancy by women who initiated their prenatal care in one of the primary care clinics of the sample.

A.   Analysis Sample

The timeline of the study and the availability of data is 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 the 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 that became pregnant before May 2011.

B. *Measurement of Weeks Pregnant at 1st Prenatal Visit*

Each observation in our sample corresponds to a different pregnancy that initiated prenatal care in one of the clinics included in the experiment. For each pregnancy we observe the date of the first prenatal visit and the date of the last menstruation period as recorded by the physician.   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]

---

[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.

One potential problem is that medical personnel in treatment facilities might misreport the date of the first visit as occurring before week 13 so that they could bill it to the program at a higher amount. We think this is unlikely for the following reasons. First, the week of pregnancy at the first 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 that could turn into substantial financial penalties for the provinces. Third, while there may have been an incentive to misreport during the intervention period, there was no financial return to misreporting in the post-intervention period once the incentives were removed. It also was unlikely that it was worth the clinics' time to set up elaborate procedures for falsifying records when they knew the incentives were only in place for 8 months. Finally, clinical records are legal documents in Argentina and practitioners could lose their medical license if caught systematically misreporting for financial gains.

To corroborate our belief that false reporting of 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 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.

We also explore whether there is any manipulation of the data at the threshold of the 13th week of pregnancy. Appendix Figure A2 shows that there is no discontinuity at this threshold using the test proposed by McCrary (2008) for manipulation at the threshold in studies that use Regression Discontinuity as their research design.

15

C. Descriptive Statistics and Baseline Balance

Table 2 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.

V.   Identification and Estimation

We estimate both the intent-to-treat (ITT) and local average treatment (LATE) effects of the incentives on outcomes (Imbens and Angrist 1994). The ITT is the effect of assigning a clinic to treatment on outcomes, regardless of compliance. The LATE is the effect of a clinic actually receiving the incentives. 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]

The ITT estimate 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 using the following specification

$$y_{ijt} = \alpha_t + \beta_t T_{jt} + \epsilon_{ijt} \quad , \tag{5}$$

where y_{ijt} is the outcome of individual i receiving care in clinic j in period t, T_j is a dummy variable taking on the value 1 if the clinic was assigned to the treatment group and 0 otherwise, and ϵ_{ijt} is a zero mean random error. Notice that parameters are allowed to vary by period. We work with four different periods of analysis: an 18-month pre-intervention period, an 8-

---

[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.

month intervention period, a 15-month post intervention period I, and an 8-month post intervention period II. We estimate separate models for each of these periods. In the LATE model we replace T_j the "assigned to treatment" variable with an indicator of being actually treated and use the clinic's randomized assignment status as an instrumental variable for actual treatment.

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 randomization inference methods that are robust to randomized assignment of treatment among a small number of clusters. Specifically, we use the Wild bootstrap to 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).

VI.   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.

A.   Densities

Figure 2 compares the densities of weeks pregnant at the time of the first prenatal visit for the clinics assigned to the treatment and control groups and reports p-values for Kolmogorov-Smirnov tests of equality of the distributions. 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 for 24 months and more after the higher fees were removed.

B. *Short Run Effects*

Table 3 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.

C. *Long Run Effects*

Our model in Section II 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 3 reports the estimated impact of the temporary fee increase on early initiation of care in the 15-month period after the fees were removed. On average, pregnant 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.

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 explore 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 3).[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 4 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.

>   D.   *Longer run effects*

The period of analysis in our main results is restricted to January of 2009 to March of 2012. Recall that starting in April 2012, the visit coding system changed. Hence starting in April 2012 what is reported as first visit in the data is actually a mix of first and second visits. As a result the average of weeks pregnant at the first visit increases and the share of pregnant women whose first visit was before week 13 falls relative to previous periods. Column 3 in Table 3 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 3 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

---

[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.

pregnant women who had their first visit before week 13. These results suggest that the improved productivity persisted at least 24 months after the fees were removed.

E.   Robustness

We implement a number of different robustness checks to our results. 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 had 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, 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 3.

Second, 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 in two ways. First, we re-estimate the models by dropping all the observations from one clinic one at a time. This produces 36 different estimated treatment effects, which we picture in appendix Figures B1 and B2 for the outcomes of weeks pregnant at the time of the first prenatal visit (B1) and for the probability 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. Notice that there is almost no difference in any of the estimates implying no one clinic drives the estimated effects in Table 3.

Finally, we estimate clinic-specific treatment effects whereby we compare each treated clinic individually to the control clinics as a group. Appendix Figures B3 and B4 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 again 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 and the solid black line represents a zero treatment effect. 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.

VII.   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 medical professionals and directors of the treated clinics. In this section we first report what we learned from these interviews. In summary, all of the clinics reported developing a new set of community outreach activities designed to identify *Plan Nacer* beneficiary women early in their pregnancies and reach out to them to encourage early initiation of prenatal care. The design and installation of these outreach activities into clinic routines involved nontrivial fixed costs and the delivery of those services created new variable costs. Moreover, the interviews reflect that during an experimentation period, clinics where able to design profitable strategies to implement in the future so that they significantly reduced the uncertainty embedded in incorporating and sustaining the new practice. Knowledge about which practices worked best also allows reducing present bias towards of old routines. We then use the whole sample to analyze the impact of the temporary incentives on community outreach activities.

Developing New Outreach Strategies— All of the clinics reported organizing a team meeting with the staff at the beginning of the intervention in order to discuss and brain storm strategies to respond to the new incentive scheme. They developed innovative data driven strategies to identify women who were likely to be pregnant. The clinics then typically sent staff to inquire about last menstruation date and offer an instant-read pregnancy test to those women whose menstruation was overdue. If pregnant, they then encouraged the expectant mothers to start prenatal care quickly.

Much of this involved experimenting with different strategies until they found what worked best. For instance, health workers started to monitor women who used birth control

pills[21] and prioritize home visits to women who were late in picking up their pill refills. Second, clinics targeted mothers who already have children, as they are less likely to initiate their prenatal visits early in a new pregnancy, by meeting them at free milk distribution centers.[22] Third, health workers noted that adolescents are less willing to reveal a pregnancy in the presence of their parents. Clinics therefore changed the timing of home visits so as to increase the chance of finding adolescents by themselves. Clinics' work schedules were also modified so as to ensure predictable availability of a gynecologist on certain days of the week so health workers could better schedule patient appointments. Other clinics started keeping track of visits to "at risk" patient and map clinic catchment areas with corresponding (potential) pregnancies so as to more efficiently organize home visit routes.

Clinics invested in testing a number of different strategies designed to reach out pregnant women and encourage them to initiate care early in order to learn which of these strategies worked best. Absent the incentives, such experimentation was perceived as being too costly and is likely an important costs of adjustment. These types of costs include monetary investments to put in extra hours of work thinking and planning new strategies, overcoming uncertainty about the effectiveness of new practices, and adjusting to working under a new set of outreach activities.

These types of costs of adjustment are anchored by old practices (habits) similar to the status quo bias (Thaler and Sunstein, 2009, pp. 35). Overcoming uncertainty, adjusting to new practices, and the expense of investing time to plan them, are all simultaneously determined and we are not able to separately identify one or the other. In fact, we interpret that the temporary incentives motivating clinics to overcome such costs by providing the clinics with the opportunity to test and experiment thereby overcoming all of these potential barriers.

Implementation — Clinics used Community Health Workers (CHWs) to implement these new outreach activities.[23] Normally, CHWs carry out community outreach activities including

---

[21] Birth control pills are dispensed free of charge by each health facility. Women cannot collect more than a months supply at any one time and must return each month for refill. The pharmacy unit keeps records of birth control pill collections.

[22] *Plan Nacer* beneficiaries with young children are eligible for free milk weekly and mothers collect the milk at distribution centers.

[23] The Ministry of Health created CHW as a job category in 2005 as part of a 3-year associates degree program in Primary Health Sector Management from the Ministry of Health. CHWs have classes at least 4 hours per

promotion of preventive health, follow-up of patients in treatment including pregnant women, follow-up of immunization status of children, health data management, early detection of malnutrition in children, among others as well as periodically updating the roster of residents in the clinic's catchment areas. Since its rollout in 2005, *Plan Nacer* has reimbursed clinics for outreach activities at a profitable rate.[24] CHWs work under temporary contracts of variable length with the facilities and are not part of the formal civil service subject to more rigid labor laws. As such, clinics can easily and quickly expand and contract the amount of CHW labor they employ. During the intervention, clinics reported expanding CHW activities by increasing the hours of existing CHWs as opposed to hiring new CHWs and paid incentive bonuses to CHWs for getting pregnant women into prenatal care.[25]

Impact of Temporary Incentives on Outreach Activities — Based on qualitative evidence, in the last section we learned that clinics reported to have developed and implemented varies strategies that increased the amount of outreach activities once the incentives were active. To investigate this further, we accessed administrative data that records all community outreach activities performed by clinics in Misiones, and test whether there were significant increases in the amount of outreach activities to pregnant women in treatment clinics relative to control clinics.[26] Figure 5 displays the average and median number of CHW outreach activities that resulted in maternal care visits for the pre-intervention, intervention, and post-intervention I periods. In this case the medians are better measures of central tendency as the densities of both activities are asymmetric heavily skewed to the right. The results show that there is no difference in outreach activities between treatment and control clinics in the pre-intervention period. In the intervention period the treatment group had substantially more activities than the control group, and this difference continued through the post-intervention period.

---

week and are required to work at least 21 hours a week as interns in a local clinic or hospital. The interns are paid an hourly stipend that is less than the minimum wage.

[24] From administrative records we the average cost of outreach activities to $1 USD as CHWs are paid $2 USD per hours and complete on average 2 outreach activities per hour. *Plan Nacer* pays $2.5 USD for outreach activities to pregnant women, so that each outreach activity generates a profit of $1.5 USD.

[25] Until 2013 health facilities participating in *Plan Nacer* in Misiones was able to use up to 50% of their of *Plan Nacer* funds to pay bonuses to health professionals. The bonuses could be assigned to any person working at the health facility, including CHWs

[26] *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.

We use these data to estimate the differences in the logarithm of number of activities between the treatment and control groups using the same methods in Table 3. 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 both the intervention and post-intervention periods (Table 4). Outreach activities doubled in the treatment clinics relative to the controls in both the intervention and post intervention periods suggesting that the temporary incentives significantly raised CHW outreach activities to a level that persisted at least 15 months after the temporary incentives were removed.

VIII.   Profitability of Prenatal Care and Outreach Activities

We have shown so far that the temporary incentives led to a long-term increase in early initiation of prenatal care through increased outreach activities by CHWs. We have also shown that clinics invested in developing and testing new data driven outreach activities in order to locate and encourage pregnant women to seek care early. To be completely consistent with our model we also need to show that prenatal care and the outreach activities used to encourage early initiation of care were known to be profitable before the temporary incentives were rolled out.

Prenatal care has always been profitable for clinics since *Plan Nacer* started. *Plan Nacer* pays clinics an additional AR\$ 40 for each prenatal care visit and half of that was used to pay bonuses to the medical care providers. Revenues from *Plan Nacer* are on top of clinic budgets that cover salaries as well as medical and non-medical supplies and materials. Clinics had been enrolled in *Plan Nacer* for over 5 years before the intervention. *Plan Nacer* trained clinics on billing procedures and clinics had been billing *Plan Nacer* for prenatal care visits over the whole period.

Clinics relied on Community Health Workers (CHWs) for outreach activities including locating and encouraging pregnant women to initiate prenatal care initiation. The Ministry of Health created CHW as a job category in 2005 as part of a 3-year associates degree program in Primary Health Sector Management from the Ministry of Health. CHWs have classes at least 4 hours per week and are required to work at least 21 hours a week as interns in a local clinic or hospital. The interns are paid an hourly stipend that is less than the minimum wage. Their normal activities include promotion of preventive health, follow-up of patients in treatment

24

including pregnant women, follow-up of immunization status of children, health data management, early detection of malnutrition in children, among others as well as periodically updating the roster of residents in the clinic's catchment areas.

Since its rollout in 2005, *Plan Nacer* has reimbursed clinics for outreach activities at a profitable rate. From administrative records we calculated that the average cost of outreach activities is $1 USD as CHWs are paid $2 USD per hour and complete on average 2 outreach activities per hour. Plan Nacer pays $2.5 USD for each outreach activity to pregnant women, so that each activity then generates a profit of $1.5 USD.

The profitability of prenatal care and outreach activities were well known to the clinics long before the implementation of the temporary incentives. Clinics had been using CHWs for other outreach activities such as promotion of preventive health, follow-up of patients in treatment including pregnant women, follow-up of immunization status of children, health data management, early detection of malnutrition in children, among others as well as periodically updating the roster of residents in the clinic's catchment areas. Clinics had been billing *Plan Nacer* for outreach activities from the beginning of the program, so that clinics should have been aware of the profitability long before the intervention. Indeed, figure 5 reports that clinics were heavily engaged in outreach in the year before the temporary incentives were rolled out.

Finally, while clinics may have known the prices paid for these services ex ante, they may have not known the costs prior to experimenting with alternatives outreach strategies. Specifically, the may not have known how many women that CHWs could visit and implement the strategy per day. Indeed, clinics could have been inhibited from trying these strategies because of the uncertainty in costs. At the higher reimbursement rate for early prenatal care during the incentives, clinics might have found it cost-effective to try out new strategies that were otherwise too risky. After learning the costs of the best strategies and confirming profitability, clinics continued to perform it in the longer run. As such, the fact that clinics expanded outreach activities is also consistent with a process of learning about the profitability of such strategies.

IX.   Knowledge and Salience of the Importance of Early Initiation of Prenatal care

A key aspect of the argument that fixed costs of adoption inhibited clinics from adopting services to increase the early initiation of prenatal care is that clinics valued early initiation

enough to have had adopted these services without the fixed costs. It is possible, however, that fixed costs of adoption were not inhibiting adoption, but rather clinics did not know the importance that the health profession places on the early initiation of care.

The temporary incentives could have informed providers about the health benefits of early initiation of prenatal care. Hence, the incentives might have worked through a "knowledge channel" to change practice patterns. This seems unlikely as early initiation of prenatal care has long been part of the Argentine CPGs for prenatal care, and is part of standard training in Argentine medical and nursing schools. Providers are taught that prenatal care by skilled health professionals beginning in the first trimester of pregnancy is essential for good maternal and newborn health outcomes as it is argued in the medical literature (Schwarcz et al. 2001; Carroli et al. 2001a; Campbell and Graham 2006; WHO 2006).

Related to the knowledge channel is saliency. Providers may have intellectually known the importance of early care, but may not have sufficiently valued it enough to invest in these services without the increased fees. The temporary incentives might have just made early initiation of care more salient[27] and thereby increased the importance of early initiation of care in the staff's minds so that it became a higher priority for action. 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 as opposed to fixed costs of adoption then the observed treatment effects may be explained by the fact that temporary incentives help to overcome this type of psychological resistance to change.

While we do not have information on the knowledge and salience of early initiation of care before the experiment, we are able to 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. To test for these hypotheses we administered a survey to the chief medical officer

---

[27] 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.

(CMO) of each clinic about the absolute and relative importance of seven different prenatal care procedures including initiating prenatal care prior to week 13 of pregnancy (see Appendix D).[28]

Figure 6 compares 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. Panel A in Figure 6 shows that the absolute score assigned by medical directors to early prenatal care is on average 4.8 in the treatment group and 4.7 in the control group. Panel B in Figure 6 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 very high absolute and relative importance, and that the temporary fees did not have an effect on either the absolute or relative importance of this practice.

X.   Other Alternative Explanations

Performance Indicator.- Financial incentives could have made treated clinics believe that early prenatal care would be a quality indicator on which they will be held accountable in the future.   This is unlikely for a number of reasons. First, when the temporary fees were introduced to the treatment clinics no mention was made of the importance of early initiation of prenatal care or that this was an indicator of quality. They were only informed about the change in the fee structure and its timing. Second, there is a published well-established list of criteria on which clinics and personnel are evaluated (National Ministry of Health 2009, 2010). Early initiation of prenatal care is not on this list. *Plan Nacer* regularly retrains clinics on these criteria. Neither the list of criteria nor the training changed around the time of the intervention.

Substitution.- 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

---

[28] We want to study the behavior of the clinic as a unit instead of a typical prenatal care provider, such as an OBGYN, since the financial incentives were designed to affect the behavior of the whole team rather than a particular individual. As such, CMOs are more representative within a clinic, and were involved in both managing the clinical team (e.g. community health workers and medical staff) and providing health care.

would have used other clinics. This is unlikely to be true as beneficiary women are assigned to specific clinics when enrolled in *Plan Nacer* and cannot simply go to another clinic to receive care. Moreover, clinics and their CHWs have specific geographic areas assigned and do not conduct outreach activities outside of those areas. Finally, 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.

Information Spillovers.- Another 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 women who became pregnant during the post-intervention period. Hence, the persistence of the effect of the incentives might be caused by an informational spillover. This is a possible argument and one that would be consistent with finding higher rates of effects of early prenatal care in the long run in treated clinics. However, the higher amount of 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. If the long-run effects where to be explained by informational spillovers the production of community outreach activities would have to present rapid diminishing marginal returns which is unlikely given their high productivity as evidenced by the field interviews. Hence, if there were large 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, since they would add to the effects of outreach activities. Moreover, if beneficiary women were passing information to others, then we could expect to see some changes in control clinics as well. However, the empirical analysis of long-run trends suggests that there is no "catch-up" effect of control clinics.

Labor Contract Frictions.- An additional candidate explanation for not reducing outreach activities after the temporary fees disappeared was not that clinics valued early initiation of prenatal care without the fee increase, but rather that CHW employment contracts were sticky so it was hard and costly to reduce CHWs. This is unlikely to be true for a number of reasons. First, continuing to provide unproductive outreach services was costly and clinics could have reassigned CHWs to other tasks or reduced their use. Second, most of the CHW expansion was

through increasing CHW hours and not hiring new CHWs so it should have been easy to reduce hours to pre-intervention levels. Third, any new CHW hired could have been easily dismissed. CHWs work under temporary contracts of variable length with the facilities and are not part of the formal civil service subject to more rigid labor laws. As such, clinics can easily and quickly expand and contract the amount of CHW labor that they employ. Fourth, since clinics knew that the temporary fees only lasted eight months when the program started, they would not have hired new CHWs on contracts for longer than that period. Contracts would then have had to be renewed or new CHWs hired in order to continue the higher level of outreach activities. Finally, even if contracts were sticky clinics should have been able to reduce some of the outreach activities in the follow-up period but we see no reduction.

Career incentives.- It is also possible that even if clinic directors did not value early initiation of prenatal care, they did not reduce outreach activities in the post intervention period because they were worried that these reductions would harm their careers (Ashraf et al. 2105). This is highly unlikely for a number of reasons. First, the Government regularly monitored clinic performance on a large number of indicators, none of which was early initiation of prenatal care. With career incentives at play money spent on outreach activities could have been better used on services for which the clinic was explicitly accountable. Second, the intervention only changed the fees for a short 8-month period of time and not a permanent change that might also signal a long-term change in priorities that might have shifted beliefs of clinical directors and staff about the importance of early prenatal care. Third, there was no accompanying information explaining the reason for the temporary fee change or that clinics would be assessed on this indicator.

XI.   Cross-Price Effects

While the modified fee schedule was designed to affect the timing of the first prenatal visit, providers may have reduced 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 5 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.

XII.   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 6 we find no effect of the incentives on birth outcomes in either the intervention period or 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 3 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.[29] Third, despite the medical literature and CPG recommendation, it is possible that early initiation of care matters only for a small amount of the general population of pregnant women, such as 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 little 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

---

[29] Results of this analysis are available upon request.

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 their 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 from it. 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.

XIII. Discussion

In this paper we examined the effects of temporary financial incentives for medical care providers to adopt better quality practices. We used this analysis to investigate whether slow diffusion of better quality practices is driven by perceived low-returns or high costs of adjustment for adopting high return practices. The results suggest that the slow diffusion is driven by high costs of adjustment as opposed to low returns.

We addressed this question in the context of a randomized field experiment in Misiones, Argentina. The intervention randomly allocated 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 estimated both the short-term effects of the incentive and whether the effects persisted once the direct monetary compensation disappears. We found that pregnant women who attended clinics in the treatment group were 34% more likely to initiate prenatal care before week 13 and that the higher levels of early initiation of care persisted for at least 24 months after the incentives ended.

We also showed that the temporary incentives motivated clinics to design and experiment with new outreach strategies to locate and encourage pregnant women to start care early. For instance, some coordinated with local pharmacies to find out when a woman was late in picking up contraceptive pills, then send community health workers to inquire about last menstruation date, offer instant-read pregnancy tests, and finally encourage the

31

expectant mothers to start prenatal care quickly. We show that outreach activities for pregnant women doubled in the treatment group.

Finally, we provided evidence that in the absence of adjustment costs of adoption, it was in the clinics' interest to have provided these outreach services. First, clinic medical directors rank early initiation of care as one of the highest of health priorities among all prenatal care services. Second, outreach activities are reimbursed at a higher rate than their cost for a long period before the experiment. Likewise, before the temporary fee increase, clinics were paid for each prenatal care service, and 50% of these additional resources were used to pay staff bonuses. As such, the temporary incentives helped to overcome adjustment costs of developing and experimenting with new outreach strategies for early prenatal care. Once clinics learned what worked best they continued to provide outreach activities to encourage early prenatal care services because they are profitable and valuable.

Our results have a number of important policy implications. First, they 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.

Ashraf, N., Bandiera, O., & Lee S. (2014). "Do-gooders and go-getters: career incentives, selection, and performance in public service delivery," STICERD - Economic Organization and Public Policy Discussion Papers Series 54, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Atkin, D., Chaudhry, A., Chaudry, S., Khandelwal, A., & Verhoogen, E. (2015). "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan," Working Paper 21417. National Bureau of Economic Research.

Baker, G. P., Jensen, M. C., & Murphy, K. J. (1988). "Compensation and incentives: practice vs. theory," *The Journal of Finance*, 43(3), 593-616.

Baker, L. C. (2001). "Managed care and technology adoption in health care: evidence from magnetic resonance imaging," *Journal of Health Economics*, 20(3), 395-421.

Baker, L. C., & Phibbs, C. S. (2002). "Managed care, technology adoption, and health care: the adoption of neonatal intensive care," *The RAND Journal of Economics,* 33(3), 524-548.

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.

Bech, M., Christiansen, T., Dunham, K., Lauridsen, J., Lyttkens, C. H., McDonald, K., & McGuire, A. (2009). "The influence of economic incentives and regulatory factors on the adoption of treatment technologies: a case study of technologies used to treat heart attacks," *Health Economics*, 18(10), 1114-1132.

Berwick, D. M. (2003). "Disseminating innovations in health care," *JAMA*, 289(15), 1969-1975.

Blattberg, R. C. & Neslin, S. A. (1990). "Sales promotion: concepts, methods, and strategies," Englewood Cliffs, Prentice Hall, New Jersey.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2012. "Does Management Matter? Evidence from India," *The Quarterly Journal of Economics*, 128(1), 1-51.

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.

Bresnahan, T., and Trajtenberg, M. (1995). "General Purpose Technologies 'Engines of Growth'?," *Journal of Econometrics* 65 (1): 83–108.

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 industrie*s*," Princeton University Press.

Carroli, G., Villar, J., Piaggio, G., Khan-Neelofur, D., Gülmezoglu, M., Mugford, M., & Bersgjø, 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.

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

Clemens, J. & Gottlieb, J. D. (2014). "Do physicians' financial incentives affect medical treatment and patient health?" *The American Economic Review*, *104*(4), 1320-1349.

Coleman, J. S., Katz, E., & Menzel, H. (1966). *Medical innovation: A diffusion study.* Second Edition, Indianapolis: Bobbs-Merrill.

Comin, D. & Hobijn, B. (2010). "An exploration of technology diffusion," *The American Economic Review*, 100(5): 2031-59.

Conley, T. G., & Udry, C. R. (2010). "Learning about a new technology: Pineapple in Ghana," *The American Economic Review*, 35-69.

Cutler, D. M., & Huckman, R. S. (2003). "Technological development and medical productivity: the diffusion of angioplasty in New York state," *Journal of Health Economics*, 22(2), 187-217.

Cutler, D. M. (2007). "The lifetime costs and benefits of medical technology," *Journal of Health Economics*, 26(6), 1081-1100.

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.

David, P. (1990). "The dynamo and the computer: an historical perspective on the modern productivity paradox," *The American Economic Review* 355-361.

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.

DellaVigna, S. (2009). "Psychology and economics: evidence from the field," *Journal of Economic Literature*, *47*(2), 315-372.

Duflo, E., Kremer, M., and Robinson, J. (2011). "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," *The American Economic Review* 101 (6): 2350–90.

Dupas, P. (2014). "Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment," *Econometrica, 82*(1), 197-28.

Eldridge C and Palmer, N. 2009. "Performance-based payment: some reflections on the discourse, evidence and unanswered questions." *Health Policy and Planning*, 24:160-6.

Evans W.N. and Stech Lien D. 2005. "The benefits of prenatal care: Evidence from the PAT Bus Strike," *Journal of Econometrics*, 125(1), 207-239.

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.

Foster, A. D., & Rosenzweig, M. R. (1995). "Learning by doing and learning from others: Human capital and technical change in agriculture," *Journal of political Economy*, 1176-1209.

Gelbach, J. B., Klick, J., & Stratmann, T. (2009). "Cheap donuts and expensive broccoli: the effect of relative prices on obesity," *Working Paper*.

Geroski, P. A. (2000). "Models of technology diffusion," *Research policy*, 29(4), 603-625.

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., & 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.

Griliches, Z. (1957). "Hybrid corn: An exploration in the economics of technological change," *Econometrica*, 501-522.

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.

Grossman, M., and Joyce, T. (1988). "Unobservables, pregnancy resolutions, and birthweight production functions in New York City," *Journal of Political Economy,* 98:983-1007.

Hall, B., & Khan, B. (2003). "Adoption of New Technology," in *New Economy Handbook*, ed. by D. C. Jones. Academic Press, San Diego.

Hall, B. (2005). "Innovation and Diffusion," in *Oxford Handbook of Innovation*, ed. by J. Fagerberg, D. C. Mowery, and R. R. Nelson, pp. 459-484. Oxford University Press.

Hannan, M. T., & Freeman, J. (1984). "Structural inertia and organizational change," *American Sociological Review*, 149-164.

Hawkes, S. J., Gomez, G. B., & Broutet, N. (2013). "Early antenatal care: does it make a difference to outcomes of pregnancy associated with syphilis? A systematic review and meta-analysis," *PloS one*, *8*(2), e56713.

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," W*orking 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.

Kirmani, 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.

Lazonick, W. (1979). "Industrial Relations and Technical Change: The Case of the Self-Acting Mule," *Cambridge Journal of Economics* 3 (3): 231–62.

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*.

McCrary, Justin. "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics* 142, no. 2 (2008): 698-714.

Mansfield, E. (1961). "Technical change and the rate of imitation," *Econometrica*, 741-766.

Mas-Colell, A., Whinston, M., & Green, J. (1995). Microeconomic Theory, Oxford, England: Oxford University Press.

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*.

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.

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.

Parente, S. L., & Prescott, E. C. (1994). Barriers to technology adoption and development. Journal of political Economy, 298-321.

Phelps, C. E. (2000). "Information diffusion and best practice adoption," *Handbook of Health Economics*, 1, 223-264.

Rosenberg, N. (1972). "Factors affecting the diffusion of technology," *Explorations in economic history*, 10(1), 3-33.

_____ (1982). Inside the black box: technology and economics. Cambridge University Press.

Rosenzweig, M.R., and Schultz P. T. (1983). "Estimating a household production function: Heterogeneity, the demand for health inputs, and their effects on birth weight." *The Journal of Political Economy,* 723-746.

Rous, J.J., Jewell, R.T. and Brown, R.W. (2004). "The effect of prenatal care on birthweight: a full-information maximum likelihood approach," *Health economics*, *13*(3), pp.251-264.

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*.

Ryan, B., & Gross, N. C. (1943). "The diffusion of hybrid seed corn in two Iowa communities," *Rural sociology*, 8(1), 15.

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.

Skinner, J., & Staiger, D. (2009). "Technology diffusion and productivity growth in health care," *Review of Economics and Statistics*, 97(5), 951-964.

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.
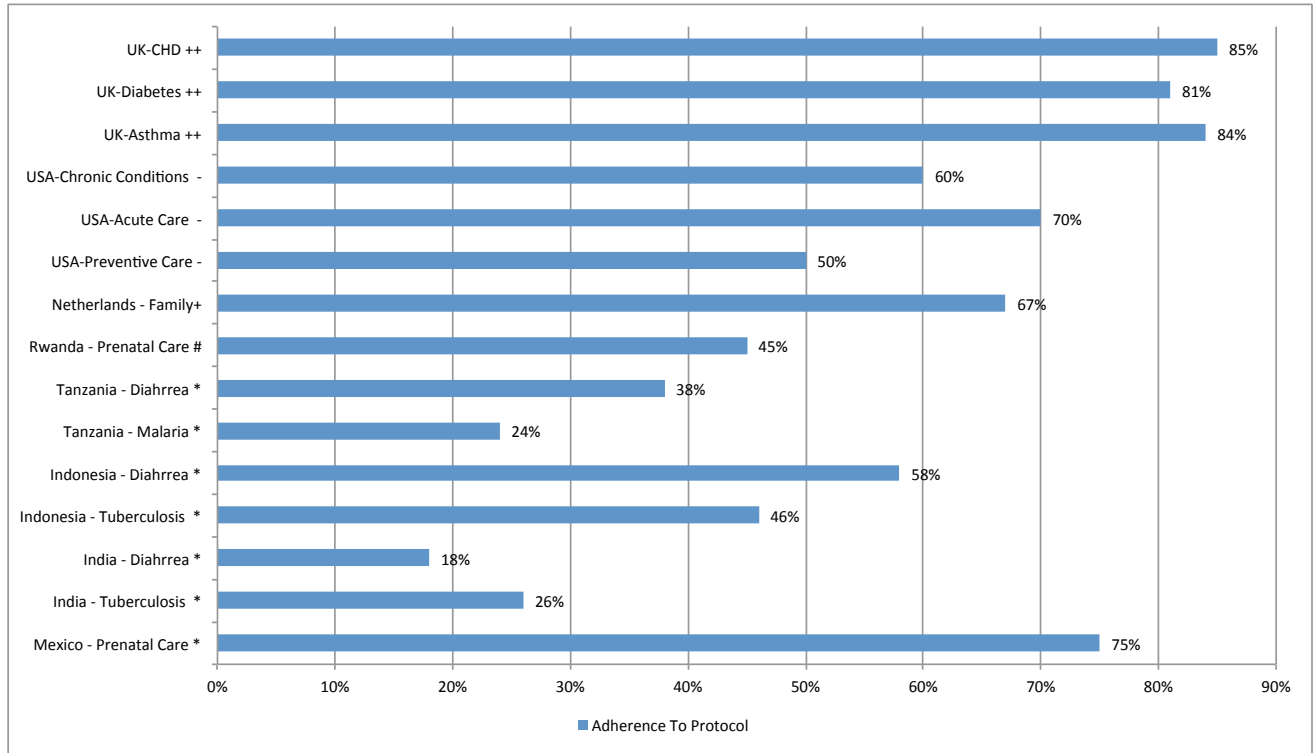
Figures and Tables

FIGURE 1: PROVIDER COMPLIANCE WITH CLINICAL PRACTICE GUIDELINES

Notes: The Figure shows the percentage of adherence to clinical practice guidelines for different countries and conditions obtained from several studies on the topic. Authors' elaboration based on (-) Schuster et al. (1998); (+) Grol (2001); (++) Campbell et al. (2007); (*) Das and Gertler (2007); and (#) Gertler and Vermeersch (2012). CHD: Coronary Heart Disease.
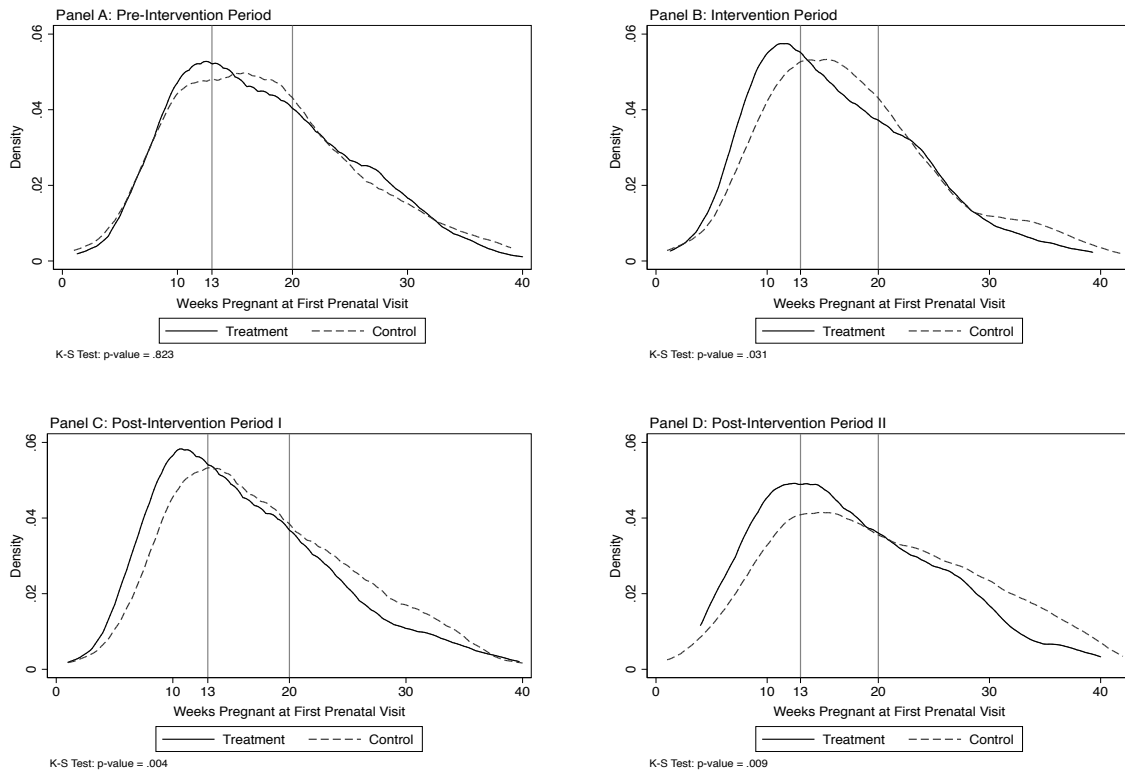
FIGURE 2: DENSITIES OF WEEKS PREGNANT AT 1ST PRENATAL VISIT

Notes: Densities estimated using an Epanechnikov kernel with optimal bandwidth. P-vales 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.
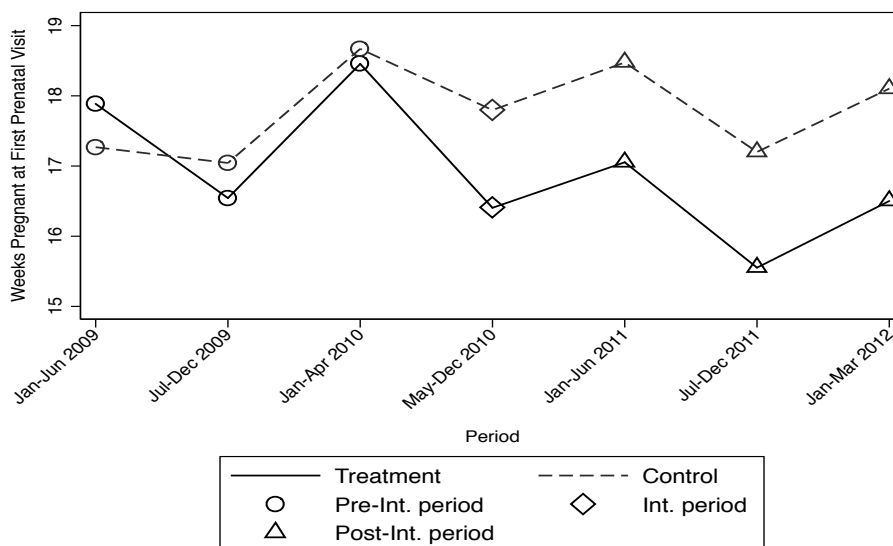
FIGURE 3: MEAN NUMBER OF WEEKS PREGNANT AT 1ST PRENATAL VISIT

Notes: The first two points (circles) are means for 6-month periods prior to the intervention period. The third point (Diamond) corresponds to the 8-month 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 4: 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 8-month 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.
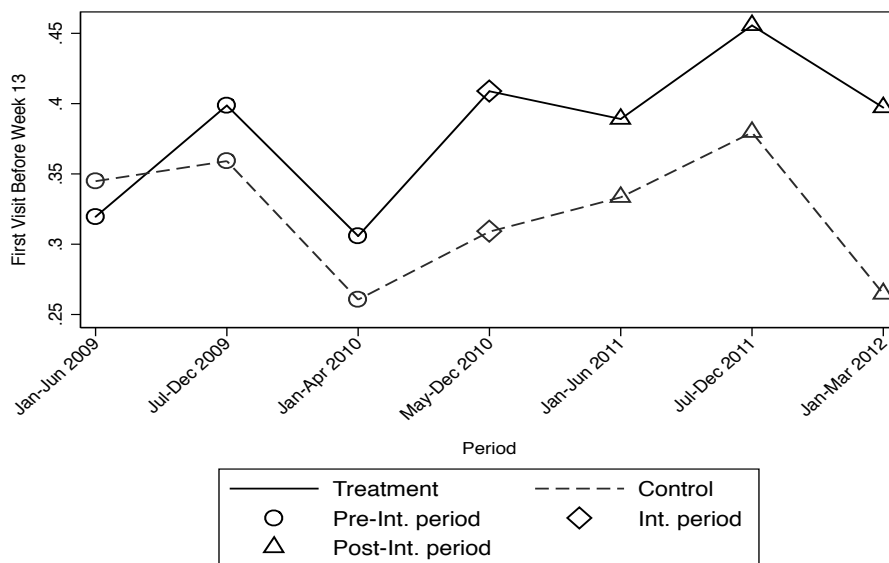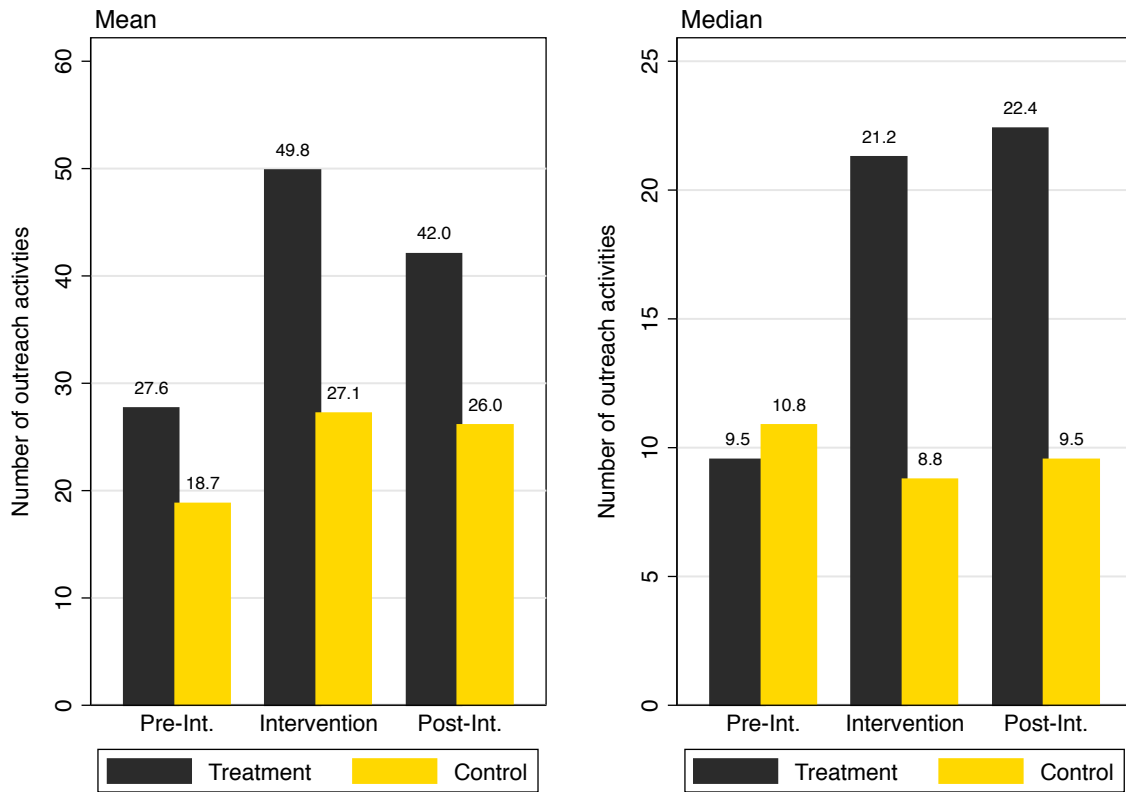
41

Figure 5: Number of Clinic Outreach Activities

Notes: 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)
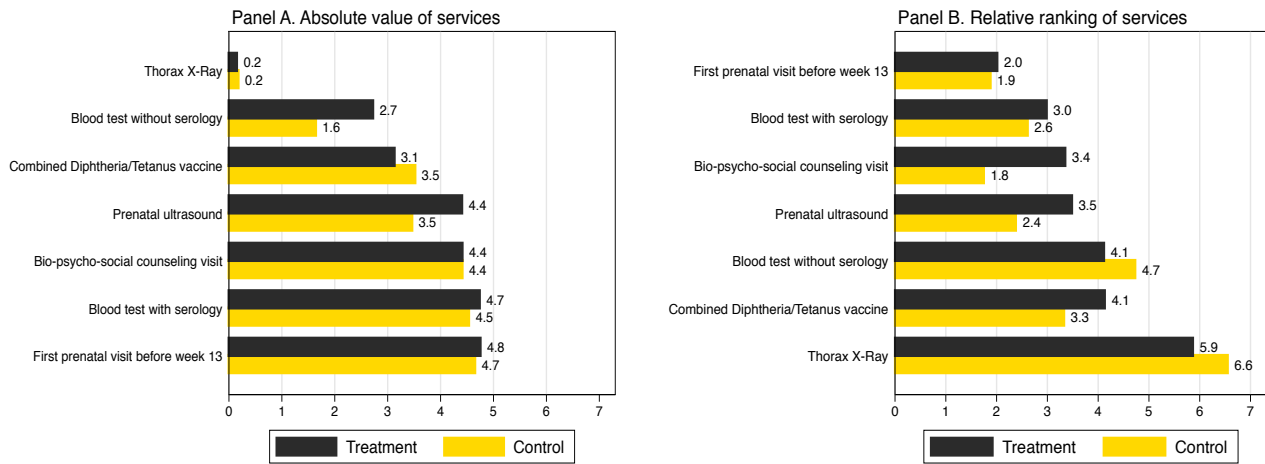
FIGURE 6: IMPORTANCE OF PRENATAL CARE SERVICES

Notes: Panel A and Panel B report the average of the absolute score and relative ranking, respectively, that measures the importance given by clinics to seven different prenatal care procedures including initiating prenatal care prior to week 13 of pregnancy (Appendix D). The absolute scores range from 1 to 5, with 5 being the highest score in terms of importance.
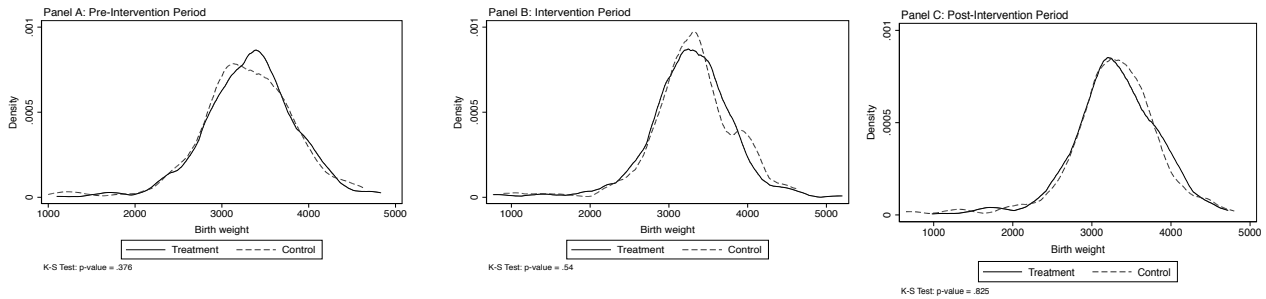

FIGURE 7: BIRTH WEIGHT DENSITIES

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

Time Period	Dates		Payment for 1st 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

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

TABLE 2: BASELINE DESCRIPTIVE STATISTICS

	Assigned Treatment Group		Assigned Control Group		*p*-Value for test of equality of means	
	Mean (s.d.)	N	Mean (s.d.)	N	Large sample	Wild Boot-Strapped
Weeks Pregnant at 1st Prenatal Visit	17.5 (7.48)	743	17.6 (7.74)	497	0.89	0.84
1st 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

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 3: EFFECTS ON TEMPORARY INCENTIVES ON TIMING OF 1ST PRENATAL VISIT

	(1)	(2)	(3)
	Intervention Period	Post-Intervention Period I	Post-Intervention Period II
A. Weeks Pregnant at 1st Prenatal Visit			
Treatment	-1.47**	-1.63**	-2.47**
	(0.71)	(0.75)	(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**	0.08**	0.08**
	(0.04)	(0.04)	(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 estimated from 2SLS regressions of the dependent variable on actual treatment status instrumented with clinic treatment assignment type. The p-values are for 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0$

TABLE 4: IMPACT ON LOG NUMBER OF OUTREACH ACTIVITIES

	(1)	(2)
	Intervention Period	Post-Intervention Period I
Treatment	0.47**	0.56**
	(0.23)	(0.22)
Large Sample p-value	0.04	0.01
Wild Bootstrapped p-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 estimated from 2SLS regressions of the dependent variable on actual treatment status instrumented with clinic treatment assignment type. The p-values are for 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).). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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

Notes Notes: This table reports LATE estimates of the treatment effect estimated from 2SLS regressions of the dependent variable on actual treatment status instrumented with clinic treatment assignment type. The p-values are for 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)
	Intervention Period	Post-Intervention Period I
A. Birth Weight		
Treatment	-37.34	25.11
	(48.61)	(40.67)
Large Sample *p*-value	0.44	0.54
Wild Bootstrapped *p*-value	0.49	0.51
Control Group Mean	3,304	3,279
Sample Size	555	802
B. Low Birth Weight		
Treatment	0.01	-0.01
	(0.02)	(0.02)
Large Sample *p*-value	0.63	0.60
Wild Bootstrapped *p*-value	0.61	0.56
Control Group Mean	0.05	0.06
Sample Size	555	802
C. Premature Birth		
Treatment	0.03	-0.04
	(0.03)	(0.02)
Large Sample *p*-value	0.31	0.08
Wild Bootstrapped *p*-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 estimated from 2SLS regressions of the dependent variable on actual treatment status instrumented with clinic treatment assignment type. The *p*-values are for 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.

We also explore whether there is any manipulation of the data at the threshold of the 13th week of pregnancy. Figure A2 shows that there is no discontinuity at this threshold using the test proposed by McCrary (2008) for manipulation at the threshold in studies that use Regression Discontinuity as their research design. The p-value at the discontinuity is 0.838.

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:

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

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, $I\left(W_{ij}^{EDD} < 13\right)$ 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).
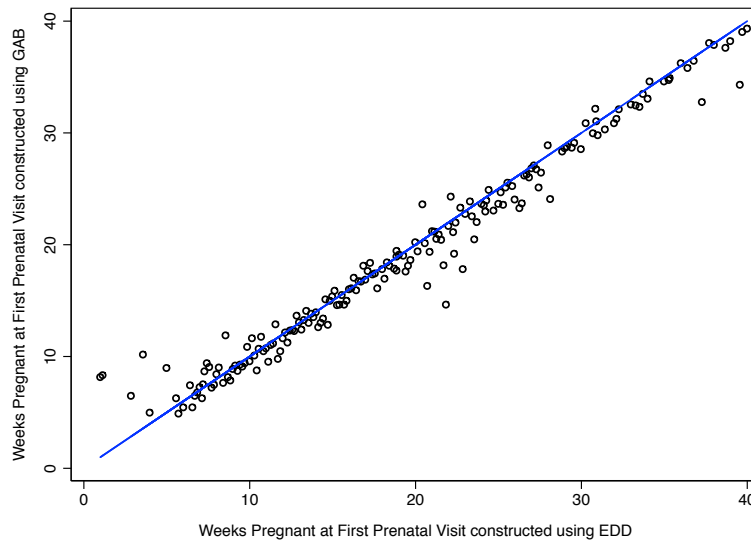

FIGURE A1: COMPARISON OF WEEKS PREGNANT AT 1ST PRENATAL VISIT BASED ON GESTATIONAL AGE AT BIRTH AND BASED ON DATE OF LAST MENSTRUATION

Notes: Authors' own elaboration based on data from the provincial medical record information system.
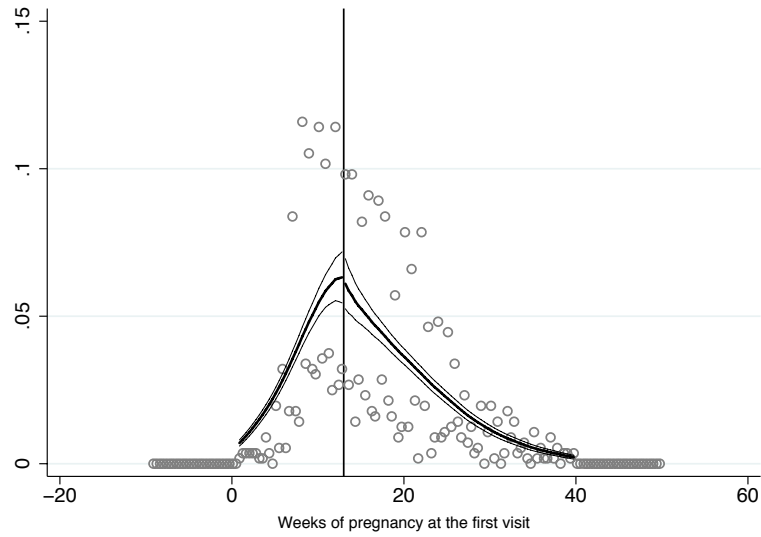
FIGURE A2: TEST FOR MISREPORTING WEEKS OF PREGNANCY AT THE THRESHOLD OF THE 13TH WEEK BASED ON THE "MANIPULATION" TEST IN MCCRARY (2008)

Notes: Authors' own elaboration based on McCrary (2008). The p-value at the discontinuity is 0.838.

TABLE A1: TEST FOR MISREPORTING WEEKS PREGNANT AT 1ST PRENATAL VISIT

Dependent Variable: Weeks Pregnant at 1st 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

Notes: 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
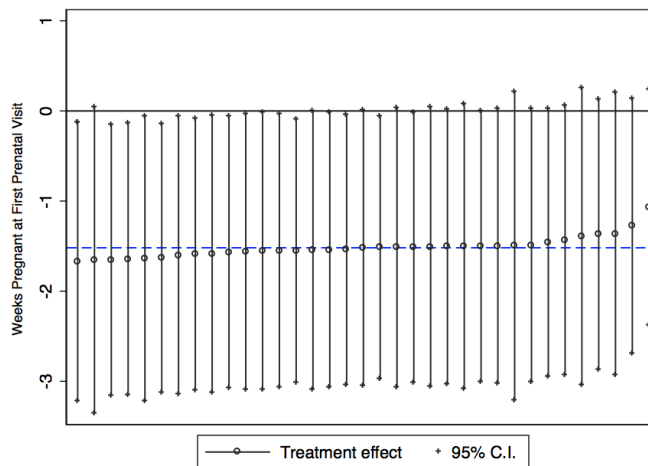
Appendix B: Robustness test results

FIGURE B1: ESTIMATES OF IMPACT ON WEEKS PREGNANT AT 1ST PRENATAL VISIT DROPPING THE OBSERVATIONS FOR EACH CLINIC ONE AT A TIME

Notes: This figure plots different treatment effects computed by dropping one clinic at a time for weeks pregnant at the first visit 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 post intervention period I (hence May 2010-March 2012). The x-axis is sorted from the lowest to the highest treatment effect. 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.
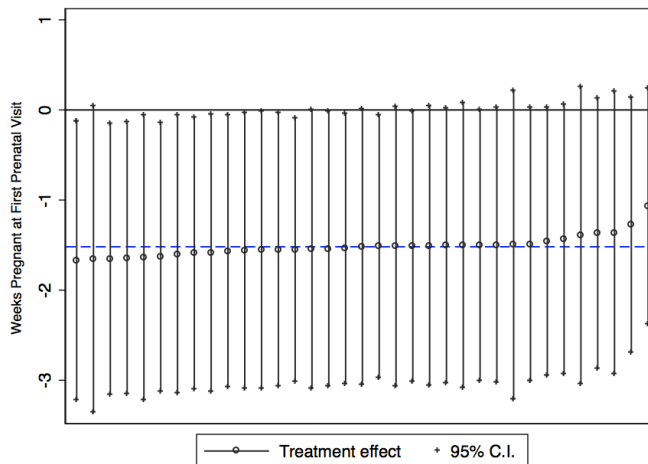
FIGURE B2: ESTIMATES OF IMPACT ON WEEKS 1ST PRENATAL VISIT BEFORE WEEK 13 DROPPING THE OBSERVATIONS FOR EACH CLINIC ONE AT A TIME

Notes: This figure plots different treatment effects computed by dropping one clinic at a time for 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). The x-axis is sorted from the lowest to the highest treatment effect. 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 B3: 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 (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.
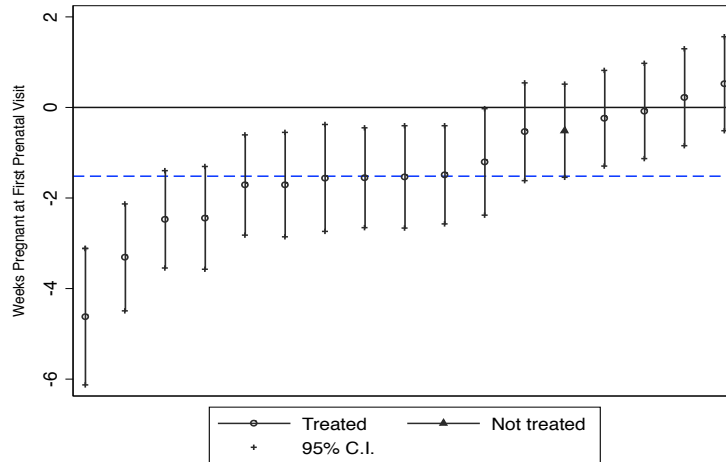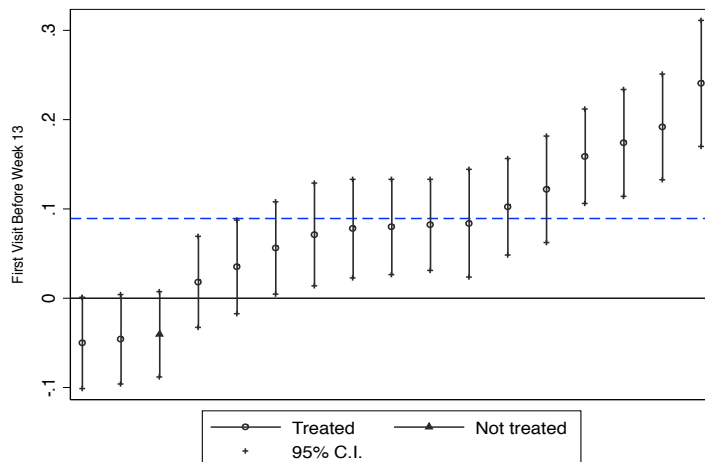
FIGURE B4: INDIVIDUAL CLINIC TREATMENT EFFECTS FOR 1ST PRENATAL VISIT BEFORE WEEK 13 OF PREGNANCY

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	Post-Intervention Period II
A. Results from Table 4			
Treatment	-1.47**	-1.63**	-2.47**
	(0.71)	(0.75)	(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*	-2.01***	-2.01*
	(0.77)	(0.70)	(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

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	Post-Intervention Period II
A. Results from Table 4			
Treatment	0.11**	0.08**	0.08**
	(0.04)	(0.04)	(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**	0.10**	0.10*
	(0.04)	(0.04)	(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

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

	(1) Intervention Period	(2) Post-Intervention Period I	(3) Post-Intervention Period II
A. Weeks Pregnant at 1st Prenatal Visit			
Treatment	-1.39** (0.67)	-1.59** (0.73)	-2.47** (1.02)
Large Sample p-value	0.04	0.03	0.02
Wild Bootstrapped p-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 p-value	0.01	0.02	0.04
Wild Bootstrapped p-value	0.03	0.05	0.08
Control Group Mean	0.31	0.34	0.27
Sample Size	769	1,269	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
A. Tetanus Vaccine		
Treatment	0.02	-0.02
	(0.07)	(0.05)
Large Sample *p*-value	0.76	0.62
Wild Bootstrapped *p*-value	0.80	0.59
Control Group Mean	0.79	0.84
Sample Size	769	1,053
A. Number of visits		
Treatment	0.37	0.50
	(0.32)	(0.57)
Large Sample *p*-value	0.24	0.38
Wild Bootstrapped *p*-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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1)	(2)
	Intervention Period	Post-Intervention Period
A. Birth Weight		
Treatment	-34.88	24.48
	(45.38)	(39.63)
Large Sample *p*-value	0.44	0.54
Wild Bootstrapped *p*-value	0.46	0.57
Control Group Mean	3304.82	3279.13
Sample Size	555	802
B. Low Birth Weight		
Treatment	0.01	-0.01
	(0.02)	(0.01)
Large Sample *p*-value	0.63	0.60
Wild Bootstrapped *p*-value	0.61	0.63
Control Group Mean	0.05	0.06
Sample Size	555	802
B. Premature		
Treatment	0.03	-0.04*
	(0.03)	(0.02)
Large Sample *p*-value	0.31	0.08
Wild Bootstrapped *p*-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: Survey of Clinic Medical Directors

In collaboration with the Provincial Management Unit of the program (UGPS), we conducted a short of clinics that participated in the pilot. The survey aimed 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). Fifty-five percent of the clinics responded to the survey, which reduces the sample to 20 clinics from the 36 clinics considered initially in the analysis. 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 7 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.

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 <u>lowest</u> importance
5 corresponds to a service to which you assign the <u>highest</u> importance

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

2. **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

Thorax X-Ray

First prenatal visit before week 13 of pregnancy

Bio-psycho-social pregnancy counseling visit

Combined Diphtheria/Tetanus vaccine

Blood test with serology

Blood test without serology

TABLE D1: BASELINE CHARACTERISTICS OF CLINICS, BY ONLINE SURVEY RESPONSE STATUS

	Non-respondent	Respondent	P-value	Obs.
Percentage in Treatment Group	0.38	.62	0.15	36

Number of Pregnant Women Attended per Year	48.60	54.90	0.33	36
Weeks Pregnant at 1st Prenatal Visit	17.04	16.77	0.15	36
1st Visit before Week 13 of Pregnancy	0.34	0.36	0.27	36
% of Pregnant Women who are *Plan Nacer* Beneficiaries	0.61	0.64	0.59	36
Tetanus Vaccine During Prenatal Visit	0.76	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	0.12	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

	Coefficient	Marg. Eff.
Treatment Group	1.498	0.274
	(1.111)	(0.180)
Birth Weight (grams)	0.100	0.018
	(1.076)	(0.196)
Weeks Pregnant at 1st Prenatal Visit	-0.594	-0.109
	(0.648)	(0.121)
1st Visit before Week 13 of Pregnancy	-3.590	-0.657
	(9.026)	(1.670)
% of Pregnant Women who are *Plan Nacer* Beneficiaries	1.620	0.296
	(4.359)	(0.774)
Tetanus Vaccine During Prenatal Visit	3.350	0.613
	(3.817)	(0.646)
Number of Prenatal Visits	-0.099	-0.018
	(0.559)	(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)	(2)	(3)	(4)
	OLS	OLS-IPW	OLS	OLS-IPW
Difference (Treatment – Control)	0.20	0.13	0.10	0.14
	(0.22)	(0.92)	(0.21)	(0.89)
Large Sample p-value	0.38	0.89	0.65	0.88
Permutation p-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.