Long-Run Effects of Temporary Incentives on Medical Care Productivity

By Pablo A. Celhay, Paul J. Gertler, Paula Giovagnoli, and Christel Vermeersch

Online Appendix

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 test for this behavior empirically. 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 14th 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 13th 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.

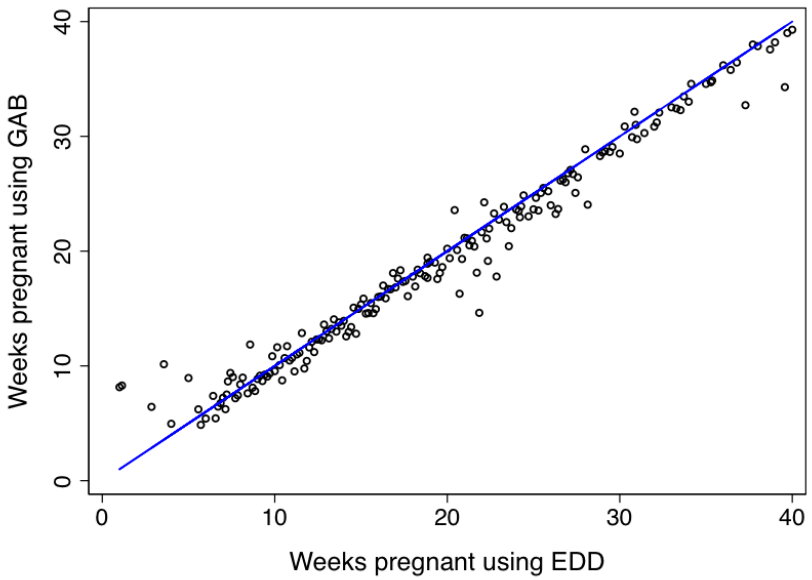

FIGURE A1. COMPARISON OF WEEKS PREGNANT AT 1ST PRENATAL VISIT BASED ON GESTATIONAL AGE AT BIRTH AND BASED ON DATE OF LAST MENSTRUATION

Note: The figure 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). The blue line is a 45 degree line.

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 **?** for manipulation at the threshold in studies that use Regression Discontinuity as their research design. The test is done separately for treated and control group clinics using only observations of the intervention period. The p-value at the discontinuity is 0.962 for the treatment group.

Furthermore, 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-differences regression:

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

where W_{ij}^{EDD} is weeks 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

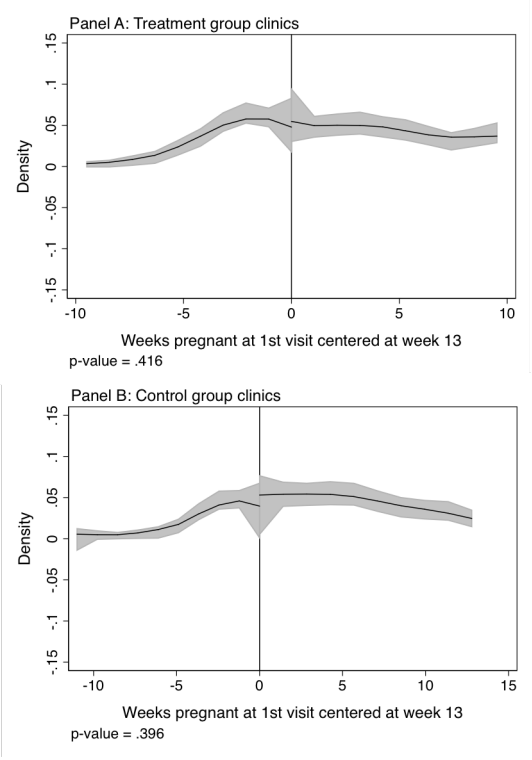

FIGURE A2. TEST FOR MISREPORTING WEEKS OF PREGNANCY AT THE THRESHOLD OF THE 13TH WEEK

Note: Authors' own elaboration based on **?** using default options of the procedure they recommend. We restrict the sample to the intervention period and estimate the densities for each group of clinics. The p-values at the bottom of each panel are for the discontinuity at the threshold of 13 weeks, centralized at 0 in the Figure.

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

In the absence of misreporting and no prematurity, there should be no difference between the two measures and β would have a coefficient of 1. However, because premature births occur before EDD, we expect β to be close to but less than one. Then we can interpret the other coefficients as the effect on $W_{ij}^{GAB} - \beta W_{ij}^{EDD}$ accounting for average weeks of prematurity. So the dependent variable is the error in EDD in forecasting actual delivery date. Equation (A1) takes on a difference-in-differences interpretation in the sense that 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).

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

Dependent Variable: Weeks Pregnant at 1st Prenatal Visit, by Gestational Age at Birth	(1)
Weeks Pregnant by EDD	0.90***
	(0.02)
1(Weeks Pregnant by EDD<13)	-0.18
	(0.32)
1(Weeks Pregnant by EDD<13) x 1(Treated=1)	-0.03
	(0.44)
Constant	1.33***
	(0.39)
Observations	1,670
Adjusted R2	0.82

Note: 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.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Appendix B: ITT Results and Descriptive Statistics of Clinic's Size

Table B1—Effects on Temporary Incentives on Timing of 1st Prenatal Visit

	(1) Intervention Period	(2) Post-Intervention Period I	(3) Post-Intervention Period II
A. Weeks Pregnant at 1st Prenatal Visit			
Treatment	-1.39**	-1.59**	-2.47**
	(0.67)	(0.73)	(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***	0.08**	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	0.08
Control Group Mean	0.31	0.34	0.27
Sample Size	769	1,269	710

Note: 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 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 (**?**). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (**?**) 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 period after the change in the coding of the first prenatal visit (April 2012 -December 2012). Standard errors are in parentheses. Standard errors are in parentheses.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B2—Cross-Price Effects of Temporary Incentives (Spillovers)

	(1) Intervention Period	(2) Post-Intervention Period I
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
B. 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

Note: This table reports ITT estimates of the treatment effect of the modified fee schedule on indicators of other services provided at the first visit. 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 (**?**). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (**?**) 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 period after the change in the coding of the first prenatal visit (April 2012 -December 2012). Standard errors are in parentheses. Standard errors are in parentheses.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table B3—Effects of Temporary Incentives Birth Outcomes

	(1)	(2)
	Intervention Period	Post-Intervention Period I
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
C. Premature Birth		
Treatment	0.03	-0.04*
	(0.03)	(0.02)
Large Sample p-value	0.31	0.08
Wild Bootstrapped p-value	0.32	0.09
Control Group Mean	0.09	0.12
Sample Size	414	708

Note: 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 differences are estimated from OLS regressions of the dependent variable on birth outcomes. 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 (**?**). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (**?**) 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 period after the change in the coding of the first prenatal visit (April 2012 -December 2012). Standard errors are in parentheses.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table B4—Correlation between Clinic's Size and Clinic's Treatment Effect

	A. Weeks Pregnant at 1st Prenatal Visit				B. First Prenatal Visit Before 13 of Pregnancy			
Regression #	Coefficient	St Err	N	Adj. R2	Coefficient	St Err	N	Adj. R2
(1)	-0.045**	(0.016)	15	0.330	0.002**	(0.001)	15	0.265
(2)	-0.055***	(0.014)	15	0.491	0.003***	(0.001)	15	0.435
(3)	-0.054***	(0.015)	15	0.466	0.003***	(0.001)	15	0.367
(4)	-0.053***	(0.016)	15	0.403	0.003**	(0.001)	15	0.342
(5)	-0.053***	(0.016)	15	0.406	0.002**	(0.001)	15	0.339
(6)	-0.053***	(0.016)	15	0.413	0.003**	(0.001)	15	0.355
(7)	-0.054***	(0.016)	15	0.409	0.003***	(0.001)	15	0.371
(8)	-0.053***	(0.016)	15	0.409	0.003**	(0.001)	15	0.350
(9)	-0.062***	(0.016)	15	0.488	0.003***	(0.001)	15	0.408
(10)	-0.053***	(0.016)	15	0.414	0.002**	(0.001)	15	0.348
(11)	-0.063***	(0.017)	15	0.489	0.003***	(0.001)	15	0.507
(12)	-0.054***	(0.016)	15	0.411	0.003***	(0.001)	15	0.372
(13)	-0.054***	(0.016)	15	0.442	0.003***	(0.001)	15	0.369
(14)	-0.050***	(0.015)	15	0.426	0.002**	(0.001)	15	0.342
(15)	-0.046**	(0.016)	15	0.325	0.002**	(0.001)	15	0.248
(16)	-0.044**	(0.018)	15	0.278	0.002*	(0.001)	15	0.192

Note: This table reports the coefficient of a simple OLS regression of clinic size against clinic's treatment effect for each of the main outcomes we analyse using the sample of treatment clinics. Each row corresponds to the correlation estimated after dropping one treatment clinic at a time.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

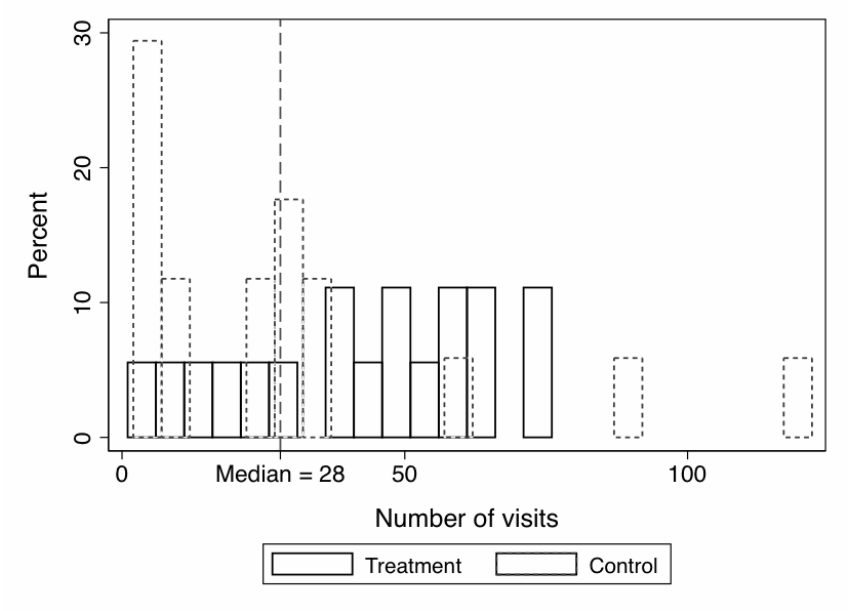

FIGURE B1. DISTRIBUTION OF CLINICS ACCORDING TO THE NUMBER OF VISITS AT BASELINE BY GROUP

Note: The histogram shows the distribution of clinics in each treatment group according to the size of the clinic. The size is calculated as the average number of women that each clinic attended in the pre intervention period.

Appendix C: Robustness Test Figures

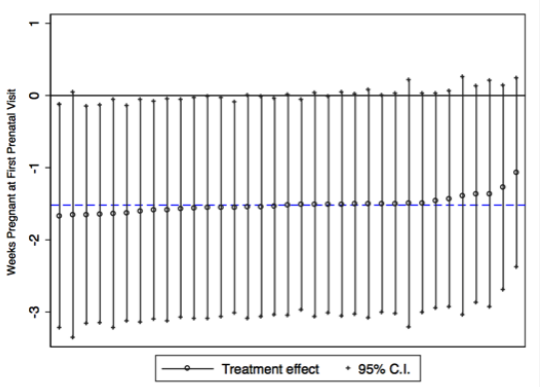

Figure C1. Estimates of Impact on Weeks Pregnant at 1st Prenatal Visit Dropping the Observations for Each Clinic One at a time

Note: This figure displays 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.

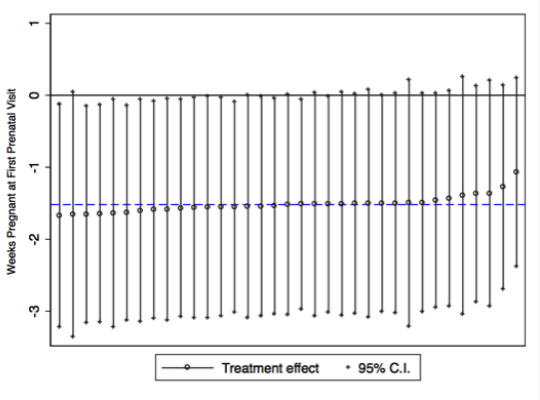

FIGURE C2. ESTIMATES OF IMPACT ON 1ST PRENATAL VISIT BEFORE WEEK 13 OF PREGNANCY DROPPING THE OBSERVATIONS FOR EACH CLINIC ONE AT A TIME

Note: This figure displays 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.

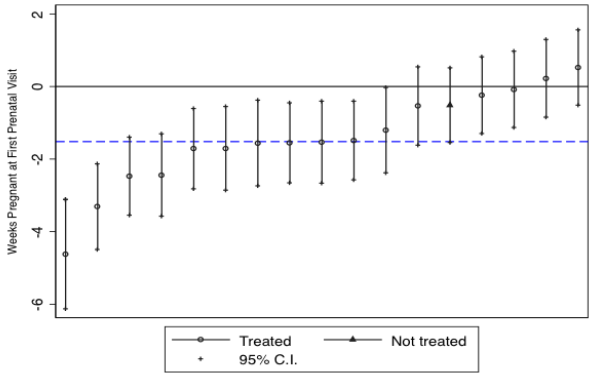

FIGURE C3. INDIVIDUAL CLINIC TREATMENT EFFECTS FOR WEEKS PREGNANT AT 1ST PRENATAL VISIT

Note: This figure displays 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.

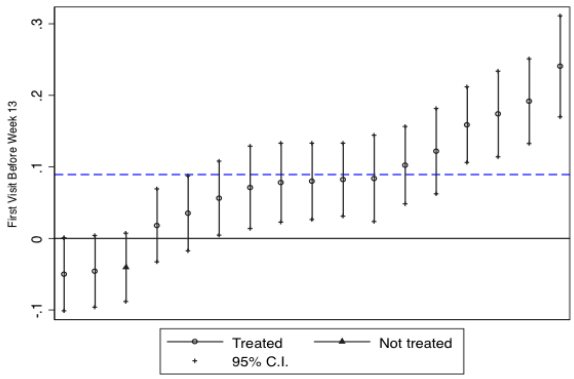

Figure C4. Individual Clinic Treatment Effects for 1st Prenatal Visit Before Week 13 of Pregnancy

Note: This figure displays 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 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.

Appendix D: Additional Robustness Test Tables

Table D1—Robustness Tests for Weeks of Pregnancy at 1st Prenatal Visit

	(1)	(2)	(3)
	Intervention Period	Post-Intervention Period I	Post-Intervention Period II
A. Results from Table **??**			
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

Note: This table reports LATE estimates of the treatment effect of the modified fee schedule on indicators of the timing of the 1st prenatal visit. The differences are estimated from OLS regressions of the dependent variable on birth outcomes. 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 (**?**). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (**?**) 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 period after the change in the coding of the first prenatal visit (April 2012 -December 2012). Standard errors are in parentheses.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE D2—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 **??**			
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

Note: This table reports LATE estimates of the treatment effect of the modified fee schedule on indicators of the timing of the 1st prenatal visit. The differences are estimated from OLS regressions of the dependent variable on birth outcomes. 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 (**?**). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (**?**) 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 period after the change in the coding of the first prenatal visit (April 2012 -December 2012). Standard errors are in parentheses.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table D3—Main Results Estimated at the Clinic Level

	(1)	(2)	(3)
	Intervention Period	Post-Intervention Period I	Post-Intervention Period II
A. Weeks Pregnant at 1st Prenatal Visit			
Treatment	-1.217**	-1.645**	-2.005*
	(0.518)	(0.610)	(1.031)
Conventional p-value	0.019	0.007	0.052
Permutation test p-value	0.082	0.001	0.002
Control Group Mean	17.80	17.90	20.10
Number of clinics	35	34	27
B. First Prenatal Visit Before Week 13 of Pregnancy			
Treatment	0.09***	0.075**	0.064
	(0.030)	(0.034)	(0.040)
Conventional p-value	0.003	0.026	0.106
Permutation test p-value	0.101	0.012	0.008
Control Group Mean	0.309	0.337	0.265
Number of clinics	35	34	27

Note: This table reports ITT estimates of the treatment effect estimated from a regression of the dependent variable on a binary indicator for random allocation of the treatment. 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 p-value after a permutation test of treatment assignment using 999 replications. Values are computed using number of pregnancies at baseline as sampling weights for each observation.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Table D4—Quartile Treatment Effects on Weeks of Pregnancy at the 1st Visit: IV Estimates

	(1) Intervention Period	(2) Post-Intervention Period I	(3) Post-Intervention Period II
1st Quartile (25th percentile)			
Treatment effect	-1.571**	-1.571***	-1.142
	(0.627)	(0.461)	(0.738)
Control group 25th percentile	12.00***	12.00***	13***
Large Sample p-value	(0.462)	(0.583)	(0.843)
Multiple hypothesis q value	0.01	0.001	0.12
2nd Quartile (50th percentile)			
Treatment effect	-1.714**	-1.714***	-3**
	(0.875)	(0.621)	(0.954)
Control group 50th percentile	16.86***	16.71***	19.29***
Large Sample p-value	(0.797)	(0.752)	(1.005)
Multiple hypothesis q value	0.03	0.01	0.002
3rd Quartile (75th percentile)			
Treatment effect	-0.857	-2.00**	-3.571**
	(1.044)	(0.789)	(1.158)
Control group 75th percentile	22***	23***	27.00***
Large Sample p-value	(0.754)	(1.143)	(1.572)
Multiple hypothesis q value	0.04	0.01	0.002
Sample Size	769	1296	710

Note: This table reports quantile regression results using random allocation of incentives as an instrument for actual take-up of the program by clinics at the 25th, 50th, and 75th percentiles of weeks pregnant at the first visit. We present large sample p-values after clustering standard errors at the clinic level. We also present the q-values obtained after controlling for the False Discovery Rate following the procedure in **?**, which adjusts p-values for the effects of testing multiple hypotheses. The control group quartile is computed as in Table **??**. Standard errors are in parentheses.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table D5—Main Results on the Subsample of Women with Data on Birth Outcomes

	(1) Baseline	(2) Intervention Period		(4) Post-Intervention Period I	(5)
	ITT	ITT	LATE	ITT	LATE
A. Weeks Pregnant at 1st Prenatal Visit					
Treatment	-0.37	-2.10***	-2.25***	-1.71*	-1.75*
	(0.59)	(0.63)	(0.65)	(0.87)	(0.88)
Large Sample p-value	0.53	0.001	0.00	0.05	0.05
Wild Bootstrapped p-value	0.58	0.004	0.01	0.08	0.09
Multiple Hypothesis q-value	1.00	0.02	0.02	0.02	0.06
Control Group Mean	17.92	18.33	18.33	18.46	18.46
Sample Size	931	555	555	802	802
B. First Prenatal Visit Before Week 13 of Pregnancy					
Treatment	0.04	0.12**	0.13**	0.07	0.07
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Large Sample p-value	0.30	0.002	0.001	0.15	0.14
Wild Bootstrapped p-value	0.31	0.01	0.01	0.17	0.15
Multiple Hypothesis q-value	1.00	0.02	0.02	0.05	0.08
Sample Size	931	555	555	802	802
Control Group Mean	0.31	0.30	0.30	0.33	0.33

Note: This table reports ITT and LATE estimates of the treatment effect of the modified fee schedule on indicators of the timing of the 1st prenatal visit for the subsample of women with data on birth outcomes The differences are estimated from OLS and IV 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 (**?**). Our Wild bootstrap procedure assigns symmetric weights and equal probability after re-sampling residuals (**?**) and uses 999 replications. We also present the q-values obtained after controlling for the False Discovery Rate following the procedure in **?**.Results for the post intervention period II are not shown since birth outcomes data for this period was not available. Standard errors are in parentheses.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

Appendix E: Analysis of Patient Volume Flow

Table E1—Effects of Temporary Incentives on the Volume of Patients

Dependent Variable:	(1) Number of first prenatal visits	(2)	(3) Log number of first prenatal visits	(4)
	Intervention Period	Post-Intervention Period I	Intervention Period	Post-intervention Period I
1(Treated=1) x After	0.105	0.429	0.093	0.042
	(0.455)	(0.438)	(0.229)	(0.220)
Observations	840	1,085	840	1,085

Note: The table shows the result of a difference-in-differences analysis. Each observation is a clinic-month. The regression controls for clinic and time fixed effects by including a binary indicator for each clinic and month. The dependent variable is the (logarithm of) the number of first prenatal care visits. "After" is an indicator variable that is equal to one if the visit occurs in the intervention (column 1) or post-intervention period (column 2) regression, and zero if the visit occurs during the pre-intervention period. Standard errors are in parentheses.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

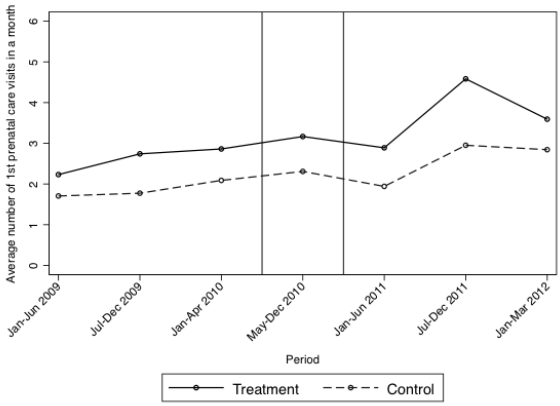

Figure E1. Average Number of 1st Prenatal care Visits by Treatment Group Before, During, and After the Intervention

Note: This figure plots the trends in the average number of 1st prenatal care visits for different periods before, during, and after the intervention for clinics in the treatment and control group.

Appendix F: Survey of Clinical Medical Directors

In collaboration with the Provincial Management Unit (UGPS) of *Plan Nacer*, we conducted a short survey of clinics that participated in the pilot. The survey (see below) 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 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 F1 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 F2 (**?**). We report results for both IPW and non-IPW regressions.

The graphs posed in Figure fig: fig8 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 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 Table F3.

Table F1—Baseline Characteristics of Clinics, by Response Status of the Online Survey

	Non respondent	Respondent	p-value of mean dif.	Sample Size
Percentage in Treatment Group	0.38	0.62	0.15	36
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 Women who are 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	0.06	0.07	0.73	31
% Premature Births	0.10	0.12	0.60	31

Note: 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 F2—Probability of Responding 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 Women who are 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

Note: 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 F3—Differences in Absolute Score and Relative Ranking of Early Prenatal Care

	(1)	(2)	(3)	(4)
	Absolute Score		Relative Ranking	
	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

Note: 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 (**?**). 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.

Box F1. 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:

1 25 years old

2 Living in the same neighborhood where your health facility is located

3 Without any apparent sign of disease

4 6 weeks pregnant

5 Had a previous low-risk pregnancy

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

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

	1	**2**	**3**	4	5	Not appropriate for a pregnant woman
Prenatal ultrasound	○	○	○	○	○	○
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