IDB

IDB WORKING PAPER SERIES No. IDB-WP-273

Does Contracting-Out Primary Care Services Work?

The Case of Rural Guatemala

Julián P. Cristia
William N. Evans
Beomsoo Kim

November 2011

Inter-American Development Bank
Department of Research and Chief Economist

Does Contracting-Out Primary Care Services Work?

The Case of Rural Guatemala

Julián P. Cristia*
William N. Evans**
Beomsoo Kim***

* Inter-American Development Bank
** University of Notre Dame
*** Korea University

Inter-American Development Bank
2011

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Cristia, Julián P.
Does contracting-out primary care services work? : the case of rural Guatemala / Julián P. Cristia, William N. Evans, Beomsoo Kim.
p. cm. (IDB working paper series ; 273)
Includes bibliographical references.
1. Child health services—Guatemala. 2. Infant health services—Guatemala. 3. Immunization of children—Guatemala. 4. Children—Services for—Guatemala. 5. Medical care—Guatemala. I. Evans, William N. II. Kim, Beomsoo. III. Inter-American Development Bank. Research Dept. IV. Title. V. Series.

http://www.iadb.org

Abstract*

This paper estimates the impact of a large-scale contracting-out program in Guatemala, using two waves of living standard measurement surveys which collected data before and after the expansion of the program and exploiting variation in the timing of the program to estimate treatment effects. Results indicate large program impacts on immunization rates for children and prenatal care provider choices. The program increases substantially the role of physician and nurses as prenatal care providers at the expense of traditional midwives. There is no evidence of effects in family planning outcomes. Taken together these results suggest a potential effective role of contracting-out in the provision of health care.

JEL classifications: I18, I12
Keywords: Contracting-out, Health care provision, Guatemala

* The author wishes to thank Nohora Alvarado, Hedi Deman, Ariadna Garcia Prado, Gerard LaForgia, Roberto Iunes, Cristina Maldonado, Olga Namen, Isabel Nieves, Silvia Raw, José Rodas, Tomás Rosada, Guilherme Sedlacek, Yuri Soares, Rita Sorio and Federico Volpino as well as seminar participants at the Inter-American Development Bank, the 2009 Population Association of America Annual Conference and the 2009 Latin American and Caribbean Economic Association Annual Conference for invaluable comments and suggestions. Nicolas Bottan and Ted Enamorado provided excellent research assistance throughout this project. The opinions expressed herein are those of the authors and do not necessarily reflect those of the Inter-American Development Bank.

1. Introduction

A third of the welfare gains in developing countries in the last four decades can be attributed to improvements in longevity and health (Becker, Philipson and Soares, 2005) and reductions in child mortality have played a key role in these changes. These advances notwithstanding, the 2003 Lancet Series on Child Mortality drew significant attention by noting the sad and staggering fact that each year over six million children worldwide die from certain diseases, such as diarrhea, malaria, and pneumonia that can be substantially prevented or successfully treated by cheap and simple measures (Jones et al., 2003).

While medical and epidemiological studies have produced clear evidence on what medical strategies can be used to combat deadly diseases, it is less clear how these services can best be delivered to the poor. The previously mentioned 2003 Lancet Series stated "The knowledge base for designing, implementing, and sustaining effective delivery is scattered and in most cases context-specific." The gaps in the delivery of medical care are evident in empirical work that finds a small impact of increases in health spending on health improvements (Musgrove, 1996; Filmer and Pritchett, 1999). Filmer and Pritchett (2000) suggest that these negative results are due to the challenges faced by governments in inducing appropriate behavior by public employees and the significant "crowding out" of private provision of care.

The need to scale-up global initiatives targeting certain health problems, together with the constraints inherent in public sector programs in developing countries, has fueled an interest in experimenting with "contracting-out" the provision of health services to private suppliers. In these arrangements, the public sector purchases the provision of a limited range of health services for a targeted population to private suppliers, typically NGOs. It is thought that contracting-out would increase access by allowing contractors to compete through results-based management and through strong incentives linking supplier payments to the achievement of pre-defined targets.

In contrast to the promise of the benefits of private supply, Liu, Hotchkiss and Bose (2008) highlight a number of shortcomings of the contracting-out model. First, significant administrative costs are involved in this approach. Second, the assumptions about the number of potential providers in a market that would allow for competitive bidding have been shown to be unrealistic in many settings. Third, contracting-out can produce further fragmentation of the health system. Fourth, governments with weak capacity to provide services may be unable to properly monitor private providers. Finally, linking payments to pre-specified outcomes may give incentives to providers to game the system in ways such as moving resources from unmeasured to measured outcomes.

To date, there is limited evidence on the effectiveness of pay-for-performance programs. Basinga et al. (2010) examine the impact of a pay-for-performance health care delivery program in Rwanda. The authors identify the impact of the program by exploiting the gradual roll-out in a difference-in-difference model. They find that pay-for-performance increased institutional deliveries, the quality of prenatal care, and childhood preventive visits, but the program had no impact on prenatal care visits or immunizations.

Whether privately-provided but publicly-funded health care can improve health is ultimately an empirical question. Unfortunately, the evidence to date is limited in volume and quality. Liu, Hotchkiss and Bose (2008) reviewed 16 studies on the effectiveness of contracting-out primary care services and reported “most of the studies are descriptive, which does not allow one to control for the influence of potential confounders on program effects.” The authors also indicated that only four of the interventions reviewed had been analyzed using regression analysis. Similarly, England (2004) pointed out that although there is a growing literature on contracting-out of primary health care in developing countries, “few of these experiences have been subject to proper evaluation.”

The problems inherent in current evaluations are best illustrated by Bloom et al. (2006), who examine a contracting-out initiative begun in Cambodia in 1999. The program focused on child and maternal health and providers were given target goals for health utilization improvements. The program was begun in five randomly selected districts, and the authors use the assignment to estimate the intention to treat with a contracting-out facility as well as the treatment-on-the-treated impact of visiting a facility. Although the results suggest that the contracting-out model increased health care utilization, many results were statistically insignificant, suggesting that that experiment may have been under-powered. For example, although the authors estimate that contracting-out increases the probability that a toddler has a complete immunization record by 15 percentage points, the p-value on the estimate is 0.46. Likewise, contracting-out is estimated to have increased the chance of antenatal care by 26 percentage points, but the p-value on this estimate is 0.35. The authors did find statistically significant increases in Vitamin A use as well as an increased probability of using public facilities for health care, which were targeted health care objectives. Interestingly, the authors found not statistically significant changes in any non-targeted outcome.

This paper aims to fill this gap in the literature by analyzing a massive contracting-out experience in Guatemala. The *Programa de Extensión de Cobertura* (Coverage Extension Program, or PEC) was launched in 1997 after the Peace Accords ended a civil war that lasted more than three decades. The government contracted NGOs to provide health care services to a significant fraction of the population, mostly rural, poor and indigenous, that was underserved by the existing public health network. The set of health services covered stressed preventive actions and were primarily focused on improving maternal and child health. To provide services to the mostly rural target population, the government chose an outreach model in which NGOs set up medical teams that made monthly visits to the covered communities. Strong community support was sought by involving local leaders and a network of volunteers. The program expanded rapidly, covering about 3 million individuals by 2000, then stabilized until 2003 when it entered a second expansion wave, adding about 0.9 million people in newly covered communities by 2005 (see Figure 1).

The program's configuration is intended to surmount two significant barriers to attaining adequate coverage of preventive health services. The first barrier is the lack of adequate transportation system, which generates high costs of seeking care for the poor rural population. The second is information gaps among the target population regarding the medical benefits of the prioritized preventive health measures. These gaps were in part due to a historically strong reliance by large segments of the population on traditional medical methods as opposed to Western medicine.

In this paper, we estimate the impact of the program on intermediate health utilization such as prenatal care, childhood immunizations and family planning methods. We use data from the 2000 and 2006 Guatemalan Living Standards Measurement Surveys (LSMS) and restrict the sample to rural areas uncovered in 2003. We then estimate a difference-in-difference model that compares trends in outcomes between communities newly covered by the 2005 expansions (treatment group) and never-covered areas (control group). We demonstrate that pre-treatment trends for a large number of covariates are similar across both treatment and comparison samples, providing support for the empirical strategy.

We find that the PEC program was particularly effective in re-directing care towards trained professionals and at increasing immunization rates. Results indicate that the program did not change the fraction of women having prenatal check-ups but significantly altered the participation of skilled professionals in the provision of these services. Our estimates indicate that the fraction of women that received prenatal care from a doctor or nurse increased by 24 percentage points from a baseline level of 26 percent (p-value<0.1) and the fraction of women receiving three or more prenatal care visits from these providers increased by 31 percentage points from a baseline of only 19 percent (p-value<0.05). These changes indicate a substantial reduction in pregnancies cared for by traditional midwives. These estimates are remarkable, as the program in a short time span has been able to overcome very strong cultural resistance to Western medicine among the mostly indigenous rural population.

In terms of immunization, results point to large positive impacts in coverage of BCG, Measles, Polio and DPT on the order of 13 to 21 percentage points. Estimated effects on coverage of DPT and Polio boosters are even higher at about 30 percentage points. In contrast, we do not find statistically significant effects on knowledge or use of family planning methods.

Unfortunately, given existing data collection procedures in Guatemala, we are unable to examine short-term health outcomes such as infant mortality. Vital Statistics on birth and mortality data in Guatemala can, however, be used to identify whether a doctor or nurse was present at the birth and whether the child died within a year. Consistent with previous reports, we found significant measurement problems in the infant mortality variable, primarily because children delivered without the aid of trained medical professionals, who died early, are typically not registered in the national Vital Statistics records, understating mortality rates for births in historically underserved areas.

The main contribution of the paper lies in evaluating, with a credible empirical strategy, one of the largest examples to date of contracting-out health care services in developing countries. The contracting-out program evaluated here involved providing services to about 4.2 million people in Guatemala by 2006, about a third of the country population.[1] The findings suggest that large expansions of preventive health services are possible in a short time frame through the contracting-out model.

[1] As mentioned, the evaluation by Bloom et al. (2006) analyzed a large-scale health contracting experience in Cambodia though somewhat smaller in population covered (1.3 million) and percent of population covered (11 percent) than our study.

2. Background

2.1 Guatemala: Health Profile and Health System

In 2005, Guatemala's PPP GDP per capita in current dollars was $5,015, making it a middle low-income country (World Development Indicators, 2007). However, a high degree of inequality skews this number. More than half of the population is considered poor, and around 15 percent extremely poor. Guatemala is characterized by a young, primarily rural, ethnically diverse population, and poverty is highly concentrated in the rural and indigenous population. Moreover, the country's fertility rates are the highest in the Americas. The country fares poorly against the rest of the Central American countries in health indicators even though it has similar GDP levels to its neighbors. Infant and child mortality (25/1,000 and 41/1,000, respectively) are significantly higher in Guatemala compared to the rest of Central America (21/1,000 and 31/1,000, respectively).

Chronic malnutrition among children aged 3 to 59 months is the highest in the Western Hemisphere (49 percent) and comparable to African countries in much earlier stages of economic development. The main causes of death are communicable diseases such as diarrhea, respiratory problems, cholera, malnutrition and tuberculosis (World Bank, 2003). More than half of the deaths for babies one year of age and under are due to diarrhea and pneumonia. At the same time, access to health care services is highly unequal and concentrated in the urban and non-indigenous population. For example, in the 2006 LSMS, only 47 percent of mothers ages 15 to 49 reported having their last birth assisted by a physician. The corresponding statistic for women in rural areas was only 30 percent—and only 23 percent for indigenous women.

The healthcare system can be characterized as fragmented and with low levels of coordination. Different health suppliers act without proper and efficient functional division but rather cater to different target populations. The public health system is primarily composed of the Ministry of Health and the Guatemalan Social Security Institute (*Instituto Guatemalteco de Seguridad Social*, IGSS). The IGSS serves less than 20 percent of the population, and its beneficiaries are concentrated among high-income individuals who participate in the more formal economic sectors. The private system is comprised of hospitals, health clinics, pharmacies and labs. Among the rural indigenous population, Western medicine is usually shunned in favor of traditional medicines. Finally, there are a large number of NGOs providing health services in the country.

The Ministry of Health supplies three levels of services, according to the complexity of the care provided. In 2005, the first level comprised 926 health posts that were geographically distributed across the country but typically located in somewhat densely populated areas. These health posts are staffed by a certified nurse who provides basic preventive and curative services and refers the most difficult cases to higher levels of attention.[2] The second level is composed of 335 health centers that have different capability levels but are all staffed with at least one physician. These health centers are typically located in county capitals.[3] Finally, the third level of care is provided by 43 hospitals located in the most populated cities. The country's health infrastructure has remained virtually unchanged in the last 15 years, with the exception of the PEC program.

2.2 The Program[4]

After the 1997 Peace Accords that followed a lengthy civil war, the government of Guatemala launched the Coverage Extension Program (*Programa de Extension de Cobertura,* PEC) to rapidly scale-up the provision of primary health services in underserved rural areas. Under this program, NGOs were contracted to provide a basic package of health services in a set of assigned communities. NGOs established mobile medical teams composed of a physician or a nurse and a health assistant who were responsible for providing services. They visited the communities and were assisted by community facilitators that lived in the area and coordinated the preparations to make the visit efficient. NGOs were paid on a capitation basis and had to attain certain pre-established targets in their assigned geographical areas or risk the cancellation of their contracts. Services provided stressed preventive measures, especially for maternal and child health.[5]

The first phase of the program (1997-1999) can be characterized as one of rapid expansion in a weak management environment in terms of planning, supervision and monitoring. Elections in 2000 brought a new national administration that did not consider the program a

[2] This level also includes the NGOs contracted under PEC to provide mostly preventive services focused on mother and child health.

[3] There are 330 counties in the country (called municipalities in Guatemala).

[4] This subsection draws heavily from Danel and LaForgia (2005). Other sources used are government reports, personnel interviews and the qualitative evaluation done in 2008 by the consultant firm GETSA.

[5] Medical teams were encouraged to also provide curative services though in practice these actions were not prioritized given the absence of targets on these dimensions and the fact that the medical mobile teams were present only once a month in each village.

priority. Hence, PEC entered a second phase where the program experienced only a slight increase in enrollment and suffered some deep budget cuts. Reversing this trend, the presidential election in 2004 brought a new government that envisioned PEC as one of its key programs. The team of individuals that originally launched the program returned to the Ministry of Health. Population coverage started to increase again, reaching 3.8 million in 2005, and payments per capita increased substantially, returning to average 1997-1999 levels (around 8 dollars per capita annually). Supervision was strengthened and targets in some cases increased. For example, NGOs were initially requested to provide two prenatal care check-ups to at least 75 percent of pregnant women, but under the new plan, the number of visits was increased to three. Finally, an individual-level electronic medical records system was designed and implemented concomitant with increased funding to NGOs to contract personnel for data-entry functions. Beyond overall reporting the new computer system was used to track patients with scheduled services such as children needing vaccinations and pregnant women.

3. Data

For the empirical analysis we exploit the program expansion that took place in 2004-2005. The basic model will be a difference-in-difference specification that will use newly enrolled areas as the treatment group and using uncovered areas as a comparison sample. Given the dates of the expansion, the paper uses data from the 2000 and 2006 Living Standard Measurement Surveys (LSMS). The surveys contain information on a range of socio-economic dimensions such as housing conditions, family composition, income, consumption, employment and education. The LSMS also include a module on maternal and child health that provides information for mothers on prenatal care, birth delivery methods and postnatal care for the last pregnancy. Also include are vaccination records for children ages 0 to 5 and, for all women ages 15 to 49, knowledge and use of family planning services. These surveys over-sampled urban residential areas (which are relatively inexpensive to survey), but the data are representative of both urban and rural areas.[6]

To determine coverage of the program over time, we obtained administrative records containing the list of covered communities in 2003 and 2005.[7] These lists were matched to the

[6] All the results in the paper are obtained using the weighs provided to take into account this issue.
[7] Unfortunately, we could only access administrative records for these years.

census of communities in 2002, which in turn was matched to the 2000 and 2006 LSMS.[8] Program administrators additionally informed us that between 1999 and 2003 there were few changes in geographic coverage.

Table 1 presents summary statistics for women ages 15 to 49 for the pooled 2000 and 2006 LSMS. Column (1) shows statistics for the whole country and documents the high prevalence of indigenous population, low levels of education and deficient coverage in basic housing services. Columns (2) and (3) highlight that the rural population, when compared to its urban counterpart, has a much higher share of indigenous residents, presents even lower levels of education (40 percent have no formal education) and has lower access to basic housing services. Column (4) documents that by 2005, the areas covered by the program tended to have a high share of indigenous residents and a population with low levels of educational attainment; this population also had slightly lower access to basic housing services when compared to the overall rural population.

Given our empirical model, we restrict the analytical sample to rural communities that were not covered by 2003 which we could safely assume that were not covered by 2000. Many of these areas will then become covered by the PEC expansions of 2004/05. Table 2 presents summary statistics for the three samples analyzed in the paper. Column (1) shows statistics for the sample used to estimate effects in prenatal care, which includes women ages 15 to 49 who gave birth in the 12 months prior to the surveys.[9] A large fraction of these women (53 percent) are indigenous, have no formal education (45 percent) and typically have low access to housing services. Regarding prenatal care, 75 percent of them receive some type of prenatal care service, though only half of them are attended by a physician or nurse (the rest are attended by traditional midwives or other individuals such as relatives). On average women make 3.6 prenatal care visits, but only 31 percent of them have three or more prenatal care visits with a physician or nurse.

[8] To match the lists of covered communities by the program with the census of communities, we searched for communities located in the same municipality and with the exact same name, and for unmatched locations we then manually searched for communities with slightly different spelling. We were able to match close to 75 percent of localities, which accounts for around 80 percent of the population in both the 2003 and 2005 registries. Correlations between population totals reported in the 2002 Census and in the program covered lists were around 0.65.

[9] Communities were added to the program in January 2004 and January 2005, and the second survey was implemented in September 2006. Therefore, this sample restriction assures that women in treatment areas had their full pregnancy covered by the program.

To examine the program impacts on vaccination rates, we use a sample of children 0 to 5 years old. Column (2) presents statistics for this group. We have outcomes for four initial doses vaccinations and two booster vaccinations. Bacillus Calmette-Guérin (BCG) is the vaccine against tuberculosis which should be administered right after birth in Guatemala.[10] Diphtheria, Pertussis (whooping cough) and Tetanus (DPT) is a vaccine against infectious diseases. Three doses are scheduled to be given at 2, 4 and 6 months after birth.[11] Polio is the vaccine against the disease of the same name which can cause paralysis or even death, and four doses are administered starting two months after birth. Measles is an infectious disease located in the respiratory system, and its vaccine is administered when the child turns one year old. Vaccination rates in the analyzed sample hover between 80 to 90 percent for DPT, Polio and BCG, although coverage for Measles and boosters for DPT and Polio present significantly lower rates.

Finally, we analyze impacts on family planning by focusing on all women ages 15 to 49 years old. Column (3) presents statistics for this group. Women in the group are mostly indigenous, with low educational levels and less access to housing services, although these characteristics seem to be slightly less prevalent than for the group that had a birth in the 12 months prior to the survey. Family planning coverage levels are quite low, with just half of women knowing about birth control methods and about 37 percent using them.

4. Identification Strategy

For each of the analytical samples, we have information from two groups: those that will be exposed to PEC as a result of the 2004-2005 expansion and those that never have PEC. We also have data for only two periods, before and after the expansion. Data with this structure lend themselves particularly well to a simple difference-in-difference specification.

Let y_{it} be the outcome for person i in time period t. Define $\overline{y}^{tb}$ and $\overline{y}^{ta}$ as the mean outcomes for the treatment group before and after the intervention, respectively. Likewise, let $\overline{y}^{cb}$ and $\overline{y}^{ca}$ be the same values for the comparison sample (again before and after, respectively). A simple difference-in-difference estimate is calculated with these four means as simply

[10] In some countries, like the United States, these mass vaccinations of BCG never happened.
[11] The recommended vaccination schedule can be found at: http://www.cdc.gov/vaccines/recs/schedules/downloads/child/0-6yrs-schedule-pr.pdf

$$(1) \qquad \hat{\beta} = (\bar{y}^{ta} - \bar{y}^{tb}) - (\bar{y}^{ca} - \bar{y}^{cb})$$

where the first difference measures the change over time in the treated group while the second difference measures the amount of the change that can be attributed to secular changes in the economy. Econometrically, the estimate for (1) can also be captured in a regression model of the form

$$(2) \qquad y_{it} = \alpha + Treated_{it}\delta + Post_{it}\gamma + Treated_{it} * Post_{it}\beta + \varepsilon_{it}$$

where $Treated_{it}$ is an indicator for whether the respondent lives in a community that will receive PEC by 2006, $Post_{it}$ an indicator for year 2006. The scalar β is the parameter of interest which estimates the average treatment effect for the selected sample. Note that because this data set is constructed by pooling two cross-sectional data sets, individuals and communities in almost all cases do not show up in the data set in both time periods. We will get unbiased estimate of β using this specification so long as the comparison sample provides an accurate estimate of the secular change in outcomes that would have occurred in the absence of the intervention. As we noted above, however, communities were selected for PEC because of their lack of care, so there is a concern that since the levels of outcomes varied across treatment and comparison samples, so too did the trends. We will return to this point later in Table 6 and provide some evidence that the growth in outcomes not expected to be impacted by the program grew at similar rates for treatment and comparison samples.

A limitation of the specification above is that it does not control for any differences across people in observed characteristics. This can easily be incorporated into the model by estimating an equation of the form

$$(3) \qquad y_{it} = \alpha + Treated_{it}\delta + Post_{it}\gamma + Treated_{it} * Post_{it}\beta + x_{it}\theta + \varepsilon_{it}$$

where the vector x_{it} captures characteristics of the individual. In the prenatal care and family planning models, we add variables that measure the woman's age in years, dummy variables for whether they were married, indigenous, currently employed, and two dummy variables for years of education (1-3 years and ≥ 4 years with no education being the reference group). We also control for some measures of wealth by adding separate dummy variables for whether the family home has running water, a flush toilet, electricity or a cement floor. For the vaccination sample we use a restricted set of controls since some variables (e.g., education) are not relevant for this age group.

The PEC program was implemented in many small communities throughout rural Guatemala. Because there are potential omitted local characteristics that may be correlated with both the PEC intervention and the growth in health outcomes, we want to control for local, time-invariant characteristics as much as possible. In the best of all worlds, we would prefer that the survey sample respondents from communities before and after the intervention. Unfortunately, given the sampling frame for the LSMS, this did not occur and few communities were sampled in both 2000 and 2006. We can, however, control for higher level geographic areas in the model. Specifically, Guatemala is divided into 22 departments. Let $u(j)_i$ be a dummy variable that equals 1 if person *i* lives in department *j*. We will add a set of 21 department dummy variables to the model, and the estimating equation will take the form

$$(4) \qquad y_{it} = \alpha + Treated_{it}\delta + Post_{it}\gamma + Treated_{it} * Post_{it}\beta + x_{it}\theta + \sum_{j=1}^{21} \mu(j)_i\psi_j + \varepsilon_{it}$$

Because there is variation within the state in communities that are treated and not, we can add the variable *Treated* and the department effects to our model.

As a further effort of controlling for local time-invariant conditions, we restrict the sample to include only observations from counties that are observed both in the pre- and post-period. We repeat this procedure for each of the three analyzed samples. For these samples, we can add fixed effects at the county level. Under this specification, we cluster the standard errors at this level of geographic aggregation. The estimating equation is

$$(5) \qquad y_{it} = \alpha + Treated_{it}\delta + Post_{it}\gamma + Treated_{it} * Post_{it}\beta + x_{it}\theta + \sum_{k=1}^{68} \mu(k)_i\psi_k + \varepsilon_{it}$$

Bertrand, Duflo and Mullainathan (2004) note that many difference-in-difference models are possibly subject to high Type I error rates due to autocorrelation in the outcome of interest. This should not pose too much of a problem with our estimates since our observations are six years apart and any autocorrelation should dissipate over time. They recommend clustering the standard errors over the dimension by which the covariate of interest is varying. In this case, PEC is instituted at the community level so, following this suggestion, we should cluster at this level. Unfortunately, as we noted above, the sample frame for the LSMS survey is such that we do not observe many communities in both the 2000 and 2006 surveys; therefore clustering at this level will not capture autocorrelation. However, if we move up to a higher geographic level (county or department), we can capture potential correlation in behavior in communities within this level of geography. Hence, in all models, the standard errors are calculated allowing for arbitrary correlation in errors within a department or a county over time.

5. Results

In this section we describe the main results of the paper. Table 3 shows the results for prenatal care. The first column presents the results for the simple difference-in-difference specification formalized in equation (2) above where errors are clustered at the department level. The results suggest that the program did not affect the share of women receiving prenatal care services nor the number of average visits. However, the results indicate that the share of women receiving prenatal care visits by a physician or nurse increased by 24 percentage points (p-value < 0.10). Even larger effects, closer to 31 percentage points, are found when analyzing the share of women receiving three or more prenatal care visits by a physician or nurse. To explore the robustness of the findings, in columns (2) to (4) we implement the empirical specifications described in the previous section in equations (3) to (5). Column (2) presents estimates when adding a number of covariates at the individual level, whereas columns (3) and (4) show results when adding fixed effects at the department and county level, respectively. Estimated coefficients are quite similar across specifications, though standard errors increase substantially in the last specification due to the reduction in sample size and hence, the impact regarding women having three or more prenatal care visits by a physician or nurse loses statistical significance.

To explore these findings in more depth, Figure 2 depicts the distribution of women with respect to the type of prenatal care provider in 2000 and 2006 by treatment/comparison group. The top panel shows that during this period there were no major changes for women in the comparison sample, with the exception of a 13 percentage-point shift from no prenatal care to physician-attended care. In contrast, women in the treatment areas experience major changes. Most notably, the share of women receiving prenatal care by a physician or nurse increases by about 35 percentage points, and this increase is reflected in large decreases in the fraction of women receiving no care, care by a traditional midwife and by other providers (e.g., relatives).

These important changes are also present when considering the distribution of women with respect to the number of prenatal care check-ups performed by physician and nurse (Figure 3). Whereas in the comparison sample there is not much change between 2000 (pre-treatment) and 2006 (post-treatment), the distribution in treatment areas is significantly altered during the period of analysis. As noted, there is a drastic reduction in the fraction of women receiving no prenatal care by a physician or nurse. But for those being attended by a physician or nurse, the distribution became heavily concentrated in the three to five visits categories. This is consistent with aggressive and strategic actions by the NGOs to achieve the target of having 75 percent of women receiving three or more prenatal care visits.

As part of improving child and maternal health, the program aimed to increase vaccination rates. Table 4 presents estimated impacts for this set of outcomes. As before, we present in column (1) results from the simple difference-in-difference specification and in columns (2) to (4) estimates for the additional specifications. Results suggest that the program was very effective in increasing vaccination rates across the board. For initial dose vaccines there are positive statistical significant impacts in the range of 14 percentage points (BCG) to 22 percentage points (Measles). Estimated effects for DPT and Polio boosters are larger and hover around 30 percentage points. Across the six analyzed outcomes, larger effects are found for those indicators with lower baseline levels, which may explain why larger impacts are present for boosters than for initial dose vaccinations. In turn, baseline vaccination rates are lower for vaccines administered at an older age.[12] Results are robust to the different specifications, though standard errors are larger when restricting the sample to observations from counties that were observed both in the pre and post periods (column (4)).

We now turn to family planning outcomes. In general estimated coefficients are quite small in magnitude, and in no cases are results statistically significant, with the sole exception of the county fixed-effects specification that yields a negative impact for knowledge at the 10 percent level. Taken together, these results suggest that the program was not effective in improving these outcomes. One potential explanation could be that the indicators that the NGO had to achieve in this area were not well aligned with the outcomes measures. In particular, the NGO targets did not include indicators related to knowledge of birth control methods.

[12] BCG is administered right after birth, DPT and POLIO when children turn 2 months and Measles at 12 months.

As noted above, we allow for arbitrary correlation in the errors for observations from the same department, and there are 22 departments in Guatemala.[13] The asymptotic properties for these procedures are demonstrated for fixed group size and as the number of panel increases to infinity. Research has documented that these procedures tend to perform poorly when the number of panels is small (Wooldridge, 2005). What constitutes "small" is, however, open to debate, and some research suggests that 22 groups may be a concern. Cameron et al. (2008) have developed a wild bootstrap procedure to produce p-values of the test of the null in the presence of within-group correlation in errors that in Monte Carlo simulations has shown to have low Type I error rates, even in the presence of small numbers of groups. We use the wild bootstrap procedure with 1,000 replications to produce p-values for our statistically significant results from Tables 3-5, and we find that the p-values from the regular clustered standard errors do not change much with the wild bootstrap procedure. The p-value from the (regular clustered standard errors) and the [wild bootstrap] procedure generate the following results: three or more prenatal care visits by physician or nurse (0.076) and [0.036], BCG vaccinations (0.007) and [0.002], DPT vaccinations (0.033) and [0.034], Polio vaccinations (0.013) and [0.044], Measles vaccinations (0.003) and [0.008], DPT booster (0.011) and [0.012], and Polio booster (0.001) and [0.008].

The identifying assumption for the difference-in-difference framework is that, in the absence of treatment, outcomes variables will have evolved similarly in the treatment and comparison samples over the analyzed period. Although this assumption is untestable, we can nonetheless check whether other covariates, which should have been unaffected by the program, have evolved similarly over time in both groups. To that end, we run difference-in-difference models but using as outcome variables the main covariates used in the analysis. Table 6 presents the results. The first column shows results when using all observations and the second column when restricting to those from counties present in the data set in both the pre and post period. For 24 (27) variables out of 29 analyzed we cannot reject that similar trends are present for both groups when using the whole sample (restricting to counties present in both years) at the five percent level. Though not definite, this evidence gives some weight to the identifying assumption used in the paper.

[13] In the case of the last specification (column (4)) in Tables 3 to 5, we cluster standard errors at the county level. Because in this case the number of clusters is significantly larger (68 to 83, depending on the sample) the estimation of standard errors is safely executed using standard statistical procedures.

6. Conclusions

Across developing countries there is growing interest towards providing public health care services through contracting-out to the private sector. However, the evidence available in this area is limited in quantity and quality. This paper aims to contribute to this literature by evaluating a large-scale contracting-out experience by the Guatemalan government. To that end, the paper exploits the 2004-2005 expansion of this program together with 2000 and 2006 LSMS data and a difference-in-difference framework to analyze the program's impacts on prenatal care, child vaccination and family planning measures. The results indicate that the program was effective in inducing a significant increase in the use of physicians and nurses as main providers of prenatal care as well as large increases in coverage of vaccination rates for children. In contrast, the results indicate a lack of impact on knowledge and use of family planning methods.

The effectiveness of this program in producing large improvements in key health care utilization in a short time span provides some hope to policymakers around the world frustrated by the failure of many previous efforts. However, more research is needed in this area. An improved understanding of the effects of programs that involve contracting-out services to the private sector will be particularly valued by policymakers searching for effective ways to increase health coverage and, in turn, reduce the burden of disease in developing countries.

References

Basinga, P. et al. 2010. "Paying Primary Health Care Centers for Performance in Rwanda." Policy Research Working Paper 5190. Washington, DC, United States: World Bank.

Becker, G., T. Philipson and R. Soares. 2005. "The Quantity and Quality of Life and the Evolution of World Inequality." *American Economic Review* 95(1): 277-291.

Bertrand, M., E. Duflo and S. Mullainathan. 2004. "How Much Should We Trust Difference in Difference Estimates." *Quarterly Journal of Economics* 119(1): 249-276.

Bloom, E. et al. 2006. "Contracting for Health: Evidence from Cambodia." Washington, DC, United States: Brookings Institution. Working Paper.

Cameron, A., D. Miller and J. Gelbach. 2008. "Bootstrap-Based Improvements for Inference with Clustered Standard Errors." *Review of Economics and Statistics* 90(3): 414-427.

Danel, I., and G. LaForgia. 2005. "Contracting for Basic Health Services in Rural Guatemala: Comparison of Performance of Three Delivery Models." In: G. LaForgia, editor. "Health System Innovations in Central America: Lessons and Impact of New Approaches." World Bank Technical Paper 54. Washington, DC, United States: World Bank.

England, R. 2004. "Experiences of Contracting with the Private Sector: A Selective Review." Issues Paper—Private Sector. London, United Kingdom: DFID Health System Resource Centre.

Filmer, D., J. Hammer and L. Pritchett. 2000. "Weak Links in the Chain: A Diagnosis of Health Policy in Poor Countries." *World Bank Research Observer* 15(2): 199-224.

Filmer, D., and L. Pritchett. 1999. "The Impact of Public Spending on Health: Does Money Matter?" *Social Science and Medicine* 49(10): 1309-1323.

Jones, G. et al. 2003. "How Many Child Deaths Can We Prevent this Year?" *The Lancet* 362: 65-71.

Liu X., D. Hotchkiss and S. Bose. 2008. "The Effectiveness of Contracting-Out Primary Health Care Services in Developing Countries: A Review of the Evidence." *Health Policy and Planning* 23: 1-13.

Musgrove, P. 1996. "Public and Private Roles in Health: Theory and Financing Patterns." World Bank Discussion Paper 339. Washington, DC, United States: World Bank.

Wooldridge, J. 2005. “Cluster Sample Methods in Applied Econometrics.” *American Economic Review Papers and Proceedings* 93(2): 133-138.

World Bank. 2003. “Guatemala Poverty Assessment Program.” Washington, DC, United States: World Bank.

Figure 1. Population Covered by PEC

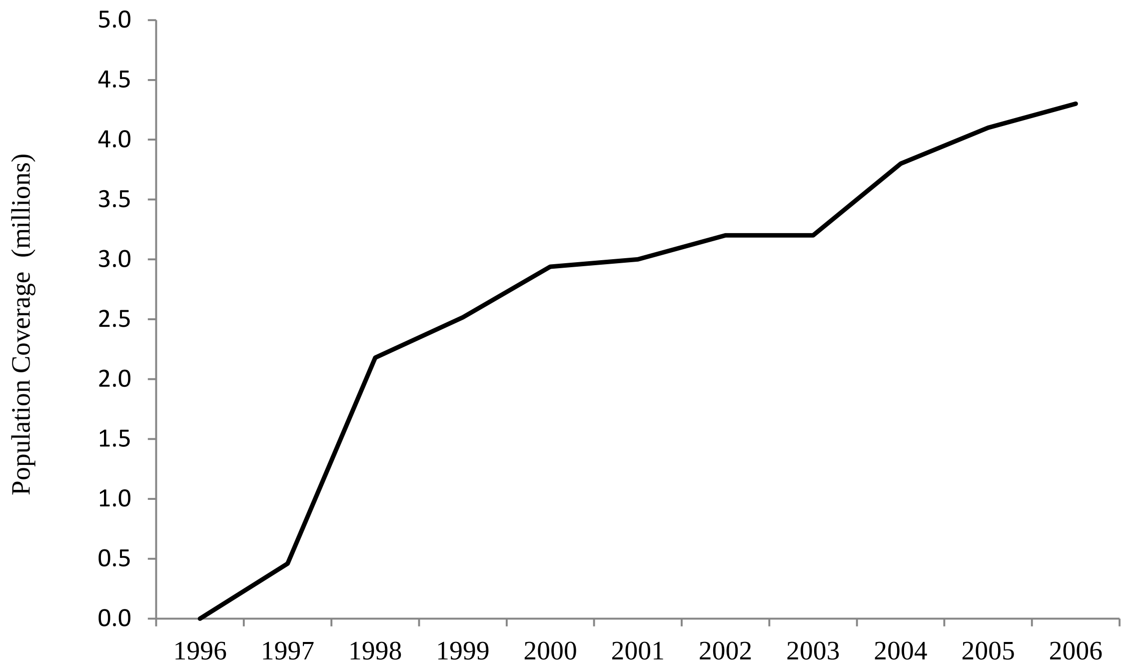
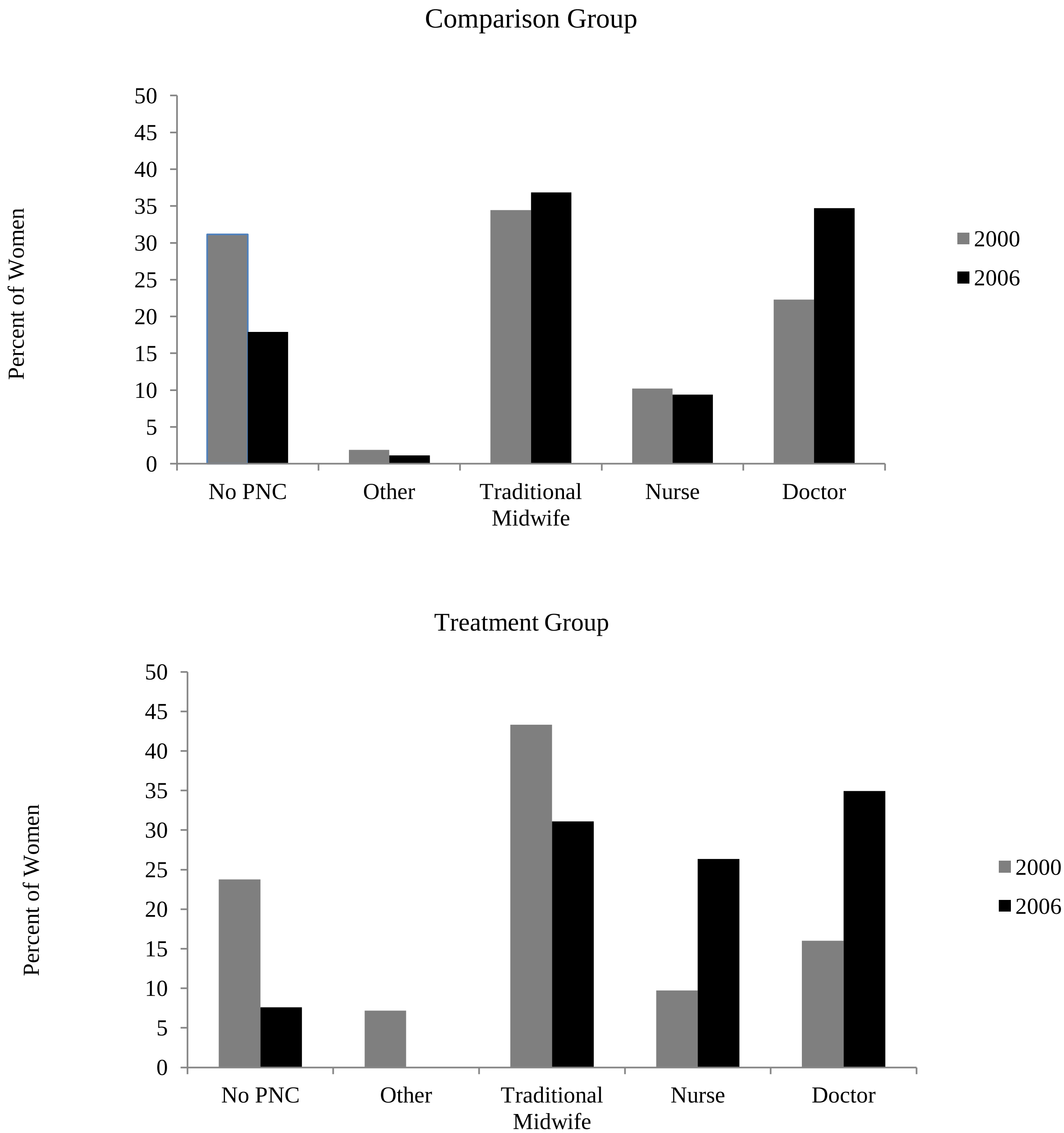
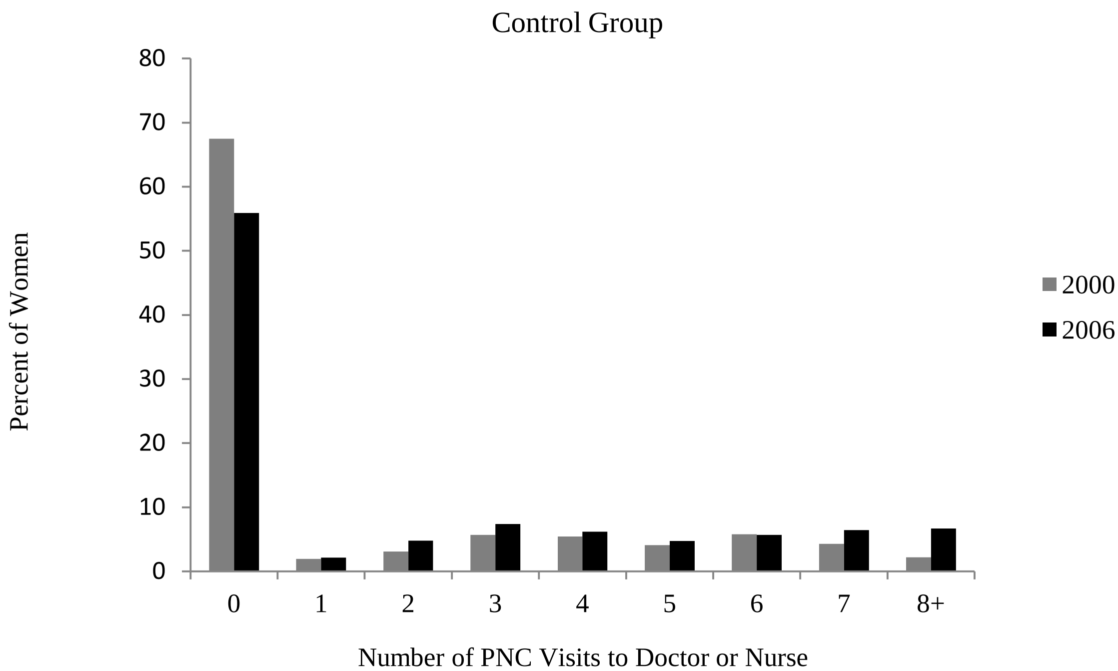


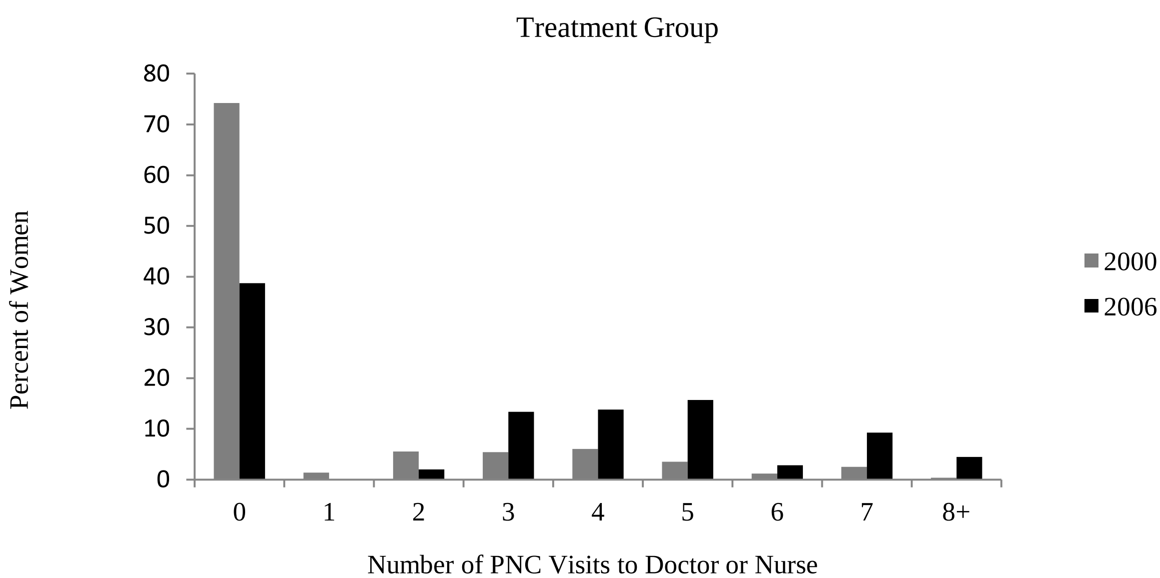
Figure 2. Choice of Provider or Use of PNC



Notes: The sample includes women that gave births in the last 12 months prior to the survey. The comparison group includes women in communities covered neither in 2003 nor in 2005. The treatment group includes women that were in communities not covered in 2003 but covered in 2005. Sample weights are used for calculations.

Figure 3. Number of PNC Visits to Doctor or Nurse





Notes: The sample includes women that gave births in the last 12 months prior to the survey. The comparison group includes women in communities covered neither in 2003 nor in 2005. The treatment group includes women that were in communities not covered in 2003 but covered in 2005. Sample weights are used for calculations.

Table 1. Descriptive Statistics, Women 15-49,
Living Standard Measurement Surveys (LSMS), 2000 and 2006

	All	Urban	Rural	Treated Area by 2005
Age	28.42	28.85	28.05	27.63
Married?	0.597	0.553	0.634	0.641
Indigenous?	0.419	0.330	0.495	0.646
Currently employed?	0.460	0.555	0.379	0.356
Years of education?				
0 years	0.316	0.183	0.429	0.468
1-3 years	0.208	0.161	0.249	0.239
4+ years	0.476	0.656	0.322	0.293
House has running water?	0.746	0.898	0.616	0.622
House has flush toilet?	0.351	0.664	0.083	0.070
House has electricity?	0.779	0.934	0.646	0.588
House has cement floor?	0.392	0.428	0.361	0.367
Observations	22816	10198	12618	3564

Notes: Sample weights used in all calculations.

Table 2. Descriptive Statistics, 2000 and 2006 LSMS

	Samples and Sample Means		
	Women age 15-49, gave birth in past 12 months	Children age 0-5	Women age 15-49
Demographic characteristics			
Female?	1.00	0.502	1.00
Age	26.16	2.03	28.02
Married?	0.889		0.696
Indigenous?	0.531	0.549	0.478
Currently employed?	0.255		0.387
Years of education?			
0 years	0.453		0.412
1-3 years	0.286		0.255
4+ years	0.261		0.333
House has running water?	0.526	0.535	0.621
House has flush toilet?	0.074	0.065	0.095
House has electricity?	0.521	0.537	0.648
House has cement floor?	0.294	0.288	0.347
Outcomes			
Prenatal care			
Any prenatal care (PNC) visits?	0.750		
# of PNC visits	3.614		
PNC by physician or nurse?	0.370		
$\geq$3 PNC by physician or nurse?	0.313		
Vaccination			
Had BCG?		0.893	
Had DPT?		0.802	
Had Polio?		0.835	
Had Measles?		0.670	
Had DPT booster?		0.519	
Had Polio booster?		0.529	
Family Planning			
Heard about birth control?			0.499
Used birth control?			0.372
Observations	1120	3978	7514

Notes: Sample weights used in all calculations.

Table 3. Difference-in-Difference-Estimates, Prenatal Care Sample, 2000 and 2006 LSMS

	All observations, 2000/2006			Obs. from counties observed in both years
		Regression adjusted		
Outcome [2000 mean for treated areas]	Diff-in-Diff Estimates (1)	Diff-in-Diff Estimates (2)	Diff-in-Diff Estimates (3)	Diff-in-Diff Estimates (4)
Any prenatal care (PNC) visits?	0.029	0.045	0.019	-0.072
[0.76]	(0.087)	(0.086)	(0.104)	(0.195)
# of PNC visits	0.421	0.433	0.107	-0.486
[3.5]	(0.515)	(0.571)	(0.597)	(1.423)
PNC by physician or nurse	0.239	0.218	0.158	0.223
[0.26]	(0.140)*	(0.139)	(0.124)	(0.165)
≥3 PNC by physician or nurse	0.308	0.299	0.230	0.291
[0.19]	(0.139)**	(0.138)**	(0.124)*	(0.182)
With state effects?	No	No	Yes	No
With county effects?	No	No	No	Yes
Observations	1120	1120	1120	682
# of communities	661	661	661	336

Notes: The sample includes women who gave birth in the last 12 months prior to the survey. The comparison group includes women in communities covered in neither 2003 nor 2005. The treatment group includes women in communities that were not covered in 2003 but were covered in 2005. In models (2) through (4), controls include age, married, indigenous, employed, education (1-3 years, >=4 years, no education as reference), running water, flush toilet, electricity, cement floor. Sample weights are used in all calculations. Standard errors, in parenthesis, are calculated allowing for arbitrary correlation in errors within a state (models 1-3) and within a municipality (model 4).

* 10%; ** 5%; *** 1%

Table 4.
Difference-in-Difference-Estimates, Childhood Vaccination Sample, 2000 and 2006 LSMS

	All observations, 2000/2006			Obs. from counties observed in both years
		Regression adjusted		
Outcome [2000 mean for treated areas]	Diff-in-Diff Estimates (1)	Diff-in-Diff Estimates (2)	Diff-in-Diff Estimates (3)	Diff-in-Diff Estimates (4)
Had BCG? [0.79]	0.136	0.126	0.156	0.206
	(0.068)*	(0.064)*	(0.053)***	(0.084)**
Had DPT? [0.78]	0.179	0.161	0.195	0.171
	(0.087)*	(0.087)*	(0.086)**	(0.105)
Had Polio? [0.78]	0.183	0.167	0.210	0.194
	(0.080)**	(0.080)**	(0.077)**	(0.108)*
Had Measles? [0.64]	0.218	0.185	0.232	0.234
	(0.070)***	(0.073)**	(0.068)***	(0.099)**
Had DPT booster? [0.44]	0.306	0.278	0.317	0.264
	(0.107)***	(0.115)**	(0.112)**	(0.113)**
Had Polio booster? [0.45]	0.307	0.278	0.310	0.244
	(0.080)***	(0.085)***	(0.080)***	(0.126)*
With state effects?	No	No	Yes	No
With county effects?	No	No	No	Yes
Observations	3978	3978	3978	2553
# of communities	602	602	602	372

Notes: Sample includes children age 0-5. The comparison group includes women in communities covered neither in 2003 nor in 2005. The treatment group includes women that were in communities not covered in 2003 but covered in 2005. In models (2) through (4), controls include age, married, indigenous, employed, education (1-3 years, >=4 years, no education as reference), running water, flush toilet, electricity, cement floor. Sample weights are used in all calculations. Standard errors in parenthesis are calculated allowing for arbitrary correlation in errors within a department (models 1-3) and within a municipality (model 4).
* 10%; ** 5%; *** 1%

Table 5.
Difference-in-Difference-Estimates, Family Planning Sample, 2000 and 2006 LSMS

	All observations, 2000/2006			Obs. from counties observed in both years
		Regression adjusted		
Outcome [2000 mean for treated areas]	Diff-in-Diff Estimates (1)	Diff-in-Diff Estimates (2)	Diff-in-Diff Estimates (3)	Diff-in-Diff Estimates (4)
Heard about birth control? [0.30]	-0.020	-0.043	-0.038	-0.165
	(0.080)	(0.052)	(0.037)	(0.085)*
Used birth control? [0.06]	-0.008	-0.013	-0.028	0.005
	(0.040)	(0.044)	(0.035)	(0.051)
With state effects?	No	No	Yes	No
With county effects?	No	No	No	Yes
Observations	7514	7514	7514	4337
# of communities	657	657	657	387

Notes: Sample includes women age 15-49. The comparison group includes women in communities covered neither in 2003 nor in 2005. The treatment group includes women that were in communities not covered in 2003 but covered in 2005. In models (2) through (4), controls include age, married, indigenous, employed, education (1-3 years, >=4 years, no education as reference), running water, flush toilet, electricity, cement floor. Sample weights are used in all calculations. Standard errors in parenthesis are calculated allowing for arbitrary correlation in errors within a department (models 1-3) and within a county (model 4).
* 10%; ** 5%; *** 1%

Table 6. Difference-in-Difference Estimates for Demographic Variables, 2000 and 2006 LSMS

	Difference-in-difference estimates	
	All observations 2000/2006	Observations from counties observed in both years
Women age 15-49, gave births in the past 12 months		
Age	-0.722 (0.497)	0.907 (1.832)
Married?	-0.075 (0.067)	0.007 (0.138)
Indigenous?	-0.234 (0.144)	0.044 (0.066)
Currently employed?	-0.153 (0.123)	-0.242 (0.175)
Education: 0 years	0.001 (0.113)	0.031 (0.187)
Education: 1-3 years	-0.190 (0.041)***	-0.199 (0.139)
Education: 4+ years	0.190 (0.100)*	0.169 (0.182)
House has running water?	-0.030 (0.115)	-0.153 (0.219)
House has flush toilet?	0.038 (0.044)	0.053 (0.092)
House has electricity?	-0.087 (0.112)	0.064 (0.147)
House has cement floor?	-0.028 (0.089)	-0.187 (0.148)
Observations	1120	682
Children age 0-5		
Age	0.150 (0.131)	0.202 (0.217)
Female?	0.094 (0.038)**	0.079 (0.100)
HH head is indigenous?	-0.045 (0.054)	0.182 (0.076)**
House has running water?	0.065 (0.154)	-0.006 (0.152)
House has flush toilet?	0.014 (0.038)	0.030 (0.059)
House has electricity?	0.140 (0.127)	0.189 (0.142)
House has cement floor?	0.064 (0.077)	-0.146 (0.123)
Observations	3978	2553
Women age 15-49		
Age	-1.213 (0.712)	0.268 (0.735)
Married?	-0.064 (0.064)	-0.053 (0.079)
Indigenous?	-0.175 (0.046)***	0.015 (0.055)
Employed?	-0.120 (0.089)	-0.258 (0.095)***
Education: 0 years	-0.063 (0.050)	0.041 (0.102)
Education: 1-3 years	-0.066 (0.015)***	-0.101 (0.065)
Education: 4+ years	0.129 (0.052)**	0.059 (0.104)
House has running water?	0.039 (0.172)	0.177 (0.163)
House has flush toilet?	0.092 (0.058)	0.015 (0.044)
House has electricity?	0.084 (0.112)	0.144 (0.168)
House has cement floor?	0.054 (0.060)	-0.109 (0.083)
Observations	7514	4337

Notes: See notes for Tables 3, 4 and 5 for sample construction and estimation. Weights are used in all calculations. Standard errors, in parenthesis, are calculated allowing for arbitrary correlation in errors within a state (column 1) and within a county (column 2). * 10%; ** 5%; *** 1%