Hospital Births, Adverse Selection And Infant Mortality

Edward N. Okeke*[1] and A.V. Chari[1]

[1]RAND Corporation

Abstract

In this paper we provide causal estimates of the effect of an institutional birth on newborn mortality. To address the endogeneity of choice of birth location, we exploit a random shock to institutional deliveries provided by the implementation of a government pay-for-performance (P4P) program in Rwanda. Using data on 11,153 births between 2000 and 2010 from the Rwandan Demographic Health Surveys, we first show that the P4P program increased rates of institutional deliveries by between 9-10 percentage points in treated relative to control districts. This is remarkably similar to the results from an earlier evaluation of the program despite the use of different data. Using program rollout as an instrument for facility deliveries, we find no statistically significant effect of institutional births on neonatal mortality. The coefficients are close to zero and robust to changes in specification. We also show that OLS models underestimate the treatment effect by several orders of magnitude, a result that is consistent with adverse selection into institutional deliveries.

*RAND Corporation, 4570 Fifth Avenue, Pittsburgh PA 15213. E-mail: ekeke@rand.org, achari@rand.org. **Preliminary and Incomplete. Please do not cite.**

1 Introduction

Nearly 4 million infants die every year within a month of being born. About 3 million of these deaths occur within the first week of life. Two-thirds of these deaths occur in Africa and Southeast Asia.[1] These deaths have been called "inexcusable" and "unconscionable" and reducing them has been referred to as a moral imperative (Lawn, Cousens and Zupan, 2005; Lawn et al., 2009). Leading causes of neonatal deaths include severe infection, asphyxia, preterm birth, and tetanus, with the first three accounting for about 87 percent of all neonatal deaths. Birth asphyxia[2] alone accounts for more than 900,000 deaths every year. Malaria in comparison, accounts for about 600,000 deaths annually (WHO, 2012). Neonatal deaths currently make up about 40 percent of under-5 mortality (You et al., 2010). Reducing neonatal mortality is therefore a critical part of the effort to achieve the MDG 4 target of reducing under-five mortality by two-thirds by 2015.

Access to skilled delivery care is believed to be important in improving newborn survival (Lawn, Cousens and Zupan, 2005; Ngoc et al., 2006). This is in part because the health of the newborn infant is closely linked to the health of the mother. Maternal risk factors such as anemia and hypertension, and delivery complications such as prolonged or obstructed labor have all been shown to be consistently associated with a higher risk of neonatal mortality (Bartlett, Paz de Bocaletti and Bocaletti, 1993; Chalumeau et al., 2000; Kusiako, Ronsmans and Van der Paal, 2000). Weiner et al. (2003) in a study in Kenya estimated that up to half of all perinatal deaths were attributable to labour complications.[3] Intrapartum risk factors are thought to play a larger role in perinatal or neonatal deaths relative to pre-pregnancy or antenatal factors (Lawn, Cousens and Zupan, 2005). Current global health strategies to reduce mortality therefore focus on increasing access to, and utilization of, ma-

[1]Ten countries – Afghanistan, Bangladesh, China, Democratic Republic of Congo, Ethiopia, India, Indonesia, Nigeria, Tanzania, and Pakistan – account for the majority of these deaths (Lawn et al., 2009).

[2]This is now referred to as an intrapartum-related neonatal death.

[3]This however was from a study of 910 births in a single district hospital in Kenya.

ternal health care services during pregnancy and delivery (Lawn et al., 2009). A central part of this strategy is promoting institutional deliveries (Filippi et al., 2006).

A growing number of countries have implemented programs that incentivize women to give birth in a health facility. The Janani Suraksha Yojana (JSY) in India is a well-known example. Others examples include the Safe Delivery Incentives Program (SDIP) in Nepal and the Demand-side Financing Maternal Health Voucher Scheme (MHVS) in Bangladesh. Funding for these programs run into the millions of dollars. The JSY for example has annual expenditures in excess of 200 million US dollars (Mazumdar, Mills and Powell-Jackson, 2011), and the SDIP has annual expenditures in excess of 1.2 million US dollars, a significant sum considering that the gross national income per capita in Nepal is $340 (Powell-Jackson et al., 2009). The returns to these expenditures largely depend on the effect of institutional births on health outcomes, and in particular on mortality. Estimates of this parameter are however hard to find.

While no one will argue that high quality obstetric care is likely to improve neonatal outcomes, more than a decade's worth of research from developing countries has shown that access to a health facility does not necessarily imply access to high quality care (Banerjee, Deaton and Duflo, 2004; Das, Hammer and Leonard, 2008). Low levels of observed quality are due to a combination of a lack of necessary clinical inputs, low levels of provider human capital, and low levels of provider effort (Chaudhury et al., 2006; Klemick, Leonard and Masatu, 2009). These problems are worse in rural communities, where infant mortality is concentrated. Locational preferences of health providers imply that rural providers are more likely to be drawn from the low end of the quality distribution. In a well-cited study by Walraven and Weeks (1999) they argue that the"skilled" attendant in the local clinic may be no more skilled than the traditional community midwife. Ronsmans et al. (2009) in a more recent study has shown that midwives are not skilled at managing complications, even when women seek help early. Harvey et al. (2007) also find significant knowledge and skill gaps in a study of skilled birth attendants in five countries. Skill deficits were much worse than knowledge deficits – the authors found that knowledge of a procedure was no guarantee that it could be performed correctly. These studies raise valid questions about the returns to institutional deliveries.

Estimating the causal effect of an institutional birth on newborn mortality is complicated by the fact that a mother's choice of delivery location is likely to be correlated with unobserved individual and household characteristics that also affect newborn health. Maternal risk is an important and often unobserved dimension along which women may differ. If women who deliver in a hospital are (predominantly) adversely selected then estimates of the effect of institutional deliveries on mortality are likely to be biased towards zero. Favorable selection will on the other hand result in estimates that are too large. It is well known for example that institutional deliveries are positively correlated with household income and maternal schooling (Gwatkin et al., 2000; Houweling et al., 2007). Bharadwaj and Nelson (2010) in a study in India show that women were more likely to give birth in a hospital when they were carrying a son. They also find that households were more likely to invest in complementary child health inputs when the fetus was male. Isolating the effect of an institutional delivery from the effect of other unobserved birth inputs that also enter into the child health production function is challenging.

In this paper we attempt to weigh in on this question. To solve the identification problem, we exploit a random shock to institutional deliveries provided by the implementation of a government pay-for-performance (P4P) program in Rwanda. What makes this program unique, and strengthens our identification, is that the program was phased in randomly. Twelve randomly chosen districts were selected to begin the program in 2006; the remaining districts joined the program in 2008. Using data on 11,153 births between 2000 and 2010 from the Rwandan Demographic Health Surveys, we first show that the P4P program significantly increased rates of institutional deliveries.[4] We find an increase of between 9 and 12 percentage points depending on the specification and the sample used. Using program rollout as an exogenous instrument

[4]Confirming the original finding by (Basinga et al., 2011).

for facility deliveries, we find no effect of an institutional birth on a child's probability of dying within the first week or within the first month. This result holds across all our specifications. We show that OLS models, even after controlling for an extensive set of individual and household characteristics, still underestimate the treatment effect by several orders of magnitude. These results are consistent with adverse selection into institutional deliveries.

The remainder of the paper is organized as follows: in Section 2 we discuss the existing literature, in Section 3 we describe the pay-for-performance program, in Section 4 we discuss the data, in Section 5 we lay out our econometric strategy, in Section 6 we discuss the results, and in Section 7 we conclude.

2 Existing Literature

Access to skilled attendance at delivery is considered to be critical in reducing deaths that occur during pregnancy, delivery and the post-partum period (World Health Organization, 1999). A Lancet series on the Millennium Development Goals has argued that MDG goals will best be achieved "by adopting a core strategy of health center-based intrapartum care" (Filippi et al., 2006). The rationale for advocating for institutional deliveries lies in the fact that mortality risk is concentrated around the time of birth – a quarter of all neonatal deaths occur within the first 24 hours, three-quarters occur within the first week (WHO, 2006). It is thought that institutional deliveries give women access to skilled providers who are better able to recognize and manage complications and perform essential interventions (including newborn resuscitation). In addition to having access to a skilled provider, it is also thought that complications are better managed in health care facilities and giving birth in hospital therefore minimizes potential delays in getting help. Despite the increasing emphasis placed on institutional deliveries in global health circles, there is surprisingly little evidence about its impact on mortality.

There are several related but distinct strands of the literature that bear on this research question. First is the epidemiologic literature on skilled birth attendance (Lawn, Cousens and Zupan, 2005; Ronsmans et al., 2008).[5] Skilled birth attendance and institutional deliveries are very highly correlated. If a birth takes place in a health facility, one can usually assume that the birth was attended by a skilled attendant – in Rwanda and India for example, greater than 99 percent of births in a health facility were attended by a doctor or nurse. For births taking place outside of a health facility, one can usually assume that a skilled provider was not in attendance, but this varies by region. In South Asia, it is more common for skilled providers to attend births at home – in India about 13 percent of non-facility births are attended by a doctor or nurse. This is in striking contrast to Rwanda where less than 1 percent of non-facility births are attended by a doctor or nurse.[6] In practice therefore skilled birth attendance is often synonymous with an institutional delivery (Campbell and Graham, 2006; Hofmeyr et al., 2009).

Two recent papers have reviewed the literature on skilled birth attendance. Darmstadt et al. (2005) finds inconsistent evidence of effects on neonatal mortality. They conclude that the overall quality of the evidence is low – eight of the ten studies reviewed were either historical studies or used a before-after design. A more recent review by Lee et al. (2011) found ten studies where the primary outcome was neonatal or perinatal mortality. They, like Darmstadt et al. (2005) observed that the quality of the available evidence was low. Based on a meta-analysis of the four studies that met their inclusion criteria, they concluded that skilled attendance conferred a small protective effect on newborns. Inference based on this literature is problematic however given their serious limitations. As we have argued estimated effects may be biased upwards if there is favorable selection into facility deliveries, or downwards, if women who choose facility deliveries are adversely selected.[7] Adverse selection may arise because women at higher risk of experiencing bad pregnancy out-

[5]The World Health Organization defines a skilled birth attendant as someone who has been trained to proficiency in the skills necessary to manage normal deliveries and diagnose, manage or refer complications. (World Health Organization, 1999). This definition includes doctors, midwives, and nurses but not traditional birth attendants.

[6]Authors' calculations from Demographic and Health Survey data for both countries.

[7]For evidence of 'adverse' selection into prenatal care, see Joyce (1994); Rous, Jewell and Brown (2004).

comes are more likely to choose an institutional delivery or because women are only taken to a health facility when complications arise. Titaley, Dibley and Roberts (2012) for example find a significantly *increased* risk of early neonatal death for deliveries taking place in public hospitals in Indonesia compared to unattended births at home. Ronsmans et al. (2009) find that mortality rates are twice as high among those who gave birth with a health professional (mostly midwives) than among those who did not.

Next we turn our attention to the economic literature on child health production. This literature has examined the effects of birth inputs (mostly prenatal care and maternal behaviors such as smoking and illicit drug use) on various child health outcomes – usually birth weight (see for example Rosenzweig and Schultz, 1983; Grossman and Joyce, 1990; Noonan et al., 2007; Habibov and Fan, 2011). We briefly discuss two papers that are particularly relevant for this study, Maitra (2004) and Panis and Lillard (1994). Both papers jointly model the demand for hospital delivery and child mortality within a system of simultaneous equations; Panis and Lillard (1994) using data from Malaysia and Maitra (2004) using data from India. Child mortality is modeled using a proportional hazard model in which the child is at risk of dying from the time of birth until the time of the survey.[8] They include a mother-specific heterogeneity term that is then allowed to be correlated across equations. Both papers find a strong effect of institutional delivery on child mortality.

Omitted variable bias remains a problem in these papers if there are inputs that affect child health that are not included in the mortality equation. Children born in a hospital for example, may be more likely than children born at home to receive postnatal care. If these variables are not included in the set of explanatory variables, then the effect of hospital birth on child mortality will be biased upwards. To give another example, household income *at the time of pregnancy/delivery* is likely to affect consumption of child health inputs (Bhalotra, 2010) and may also affect child health (for example through maternal nutrition). Failing to include precise measures of household income at the time of birth will lead to biased coefficients. Identification in these papers

[8]Panis and Lillard (1994) allows for fetal deaths i.e. miscarriages and abortions.

also relies on strong assumptions. In particular, the model assumes that once mother-specific heterogeneity is controlled for, the remaining variation in the error term is random. Mother-specific heterogeneity implies selection at the level of the woman – in other words, certain women are more likely to select into facility deliveries, but selection may be birth-specific. As Bharadwaj and Nelson (2010) showed, households expecting a son were more likely to invest in child health inputs. Systematic preferences for certain kinds of births may lead to overestimates of the treatment effect because of the induced correlation between facility deliveries and other possibly unobserved child health inputs.

This paper is also related to the growing literature estimating the returns to medical treatment. Recent examples of this literature include Almond et al. (2010), Almond and Joseph J. Doyle (2011), and Evans and Garthwaite (2009). Adhvaryu and Nyshadham (2011) is the only developing country paper that we are aware of. Lastly we note the recent literature estimating reduced form impacts of programs promoting institutional deliveries. Lim et al. (2010) evaluates the impact of the JSY, a conditional cash transfer program in India that incentivizes women to give birth in a health facility, on neonatal mortality. They find statistically significant reductions of between 2.3 and 2.4 neonatal deaths per 1000 live births. These findings have however been questioned by Mazumdar, Mills and Powell-Jackson (2011) who use different methods and, in contrast to Lim et al. (2010), find no effect on neonatal mortality.[9]

This paper makes a significant contribution to the existing literature by estimating the structural parameter of interest i.e. the causal effect of facility deliveries on infant mortality. A major strength of this paper is that our identification is transparent and relies on a shock to institutional births provided by the randomized rollout of a government program. This allows us to account for the endogeneity of facility deliveries. We also avoid the problem of correlation between hospital births and possibly unobserved postnatal child health inputs by restricting our attention to neonatal mortality.

[9]They conclude that quality of care in the facilities was a key barrier to mortality reductions.

3 Pay-for-Performance in Rwanda

Performance-based financing (PBF) began in Rwanda in 2001 when international NGOs started pilot schemes in two provinces (Cyangugu in 2001, and Butare in 2002).[10] Based on the success of these initiatives (Soeters, Musango and Meesen, 2005), the Ministry of Health with funding from USAID decided to scale up PBF and expand it to all health facilities in Rwanda. Due to capacity and financial constraints, it was not feasible to implement the program across all districts in the country at the same time. For reasons of equity and fairness, it was determined that districts would be assigned to Phase I (rollout starting in 2006), or Phase II (rollout starting in 2008) using a lottery. Areas with no existing performance based contracting operations were paired based on similar characteristics and a coin flip was used to decide which districts would receive the program in Phase I, and which would wait for Phase II.[11]

District boundaries were redefined in late 2005 as part of a government decentralization effort. This resulted in some evaluation districts being combined with districts with existing P4P pilot programs. As Basinga et al. (2011) describe, because P4P schemes could not be removed from health facilities in which they had already been implemented, the government enrolled all health facilities in newly formed districts that had existing pilot programs schemes into the first phase of the rollout.[12] The final evaluation sample consisted of twelve Phase I districts and seven Phase II districts.

The PBF program incentivized provision of fourteen different services (see Table A.1). Bonus payments were paid out to health facilities quarterly according to the following payment formula:

[10] A third scheme was begun in 2005 in Kigali-Ngali, Kabgayi, and Kigali Ville by Belgian Technical Cooperation (BTC), a development cooperation agency.

[11] Random allocation was thought to be a more transparent and equitable mechanism than other methods of allocation.

[12] Based on personal email communication from Paulin Basinga, one of the lead investigators on the original evaluation, this affected three districts. As part of our robustness checks, we drop these districts from the sample. This does not affect the results.

$$Payment_{it} = Q_{it} * \sum_{j} P_j U_{ijt}$$

P is the per unit payment for each incentivized service, U is the number of units provided of each service and i indexes the facility, j indexes the service, and t indexes time. Q is a quality index ranging from 0 to 1. Weights are assigned to various quality indicators including cleanliness, availability of drugs and supplies, etc. The index is described in detail in Rusa et al. (2009). Notice from Table A.1 that the highest per unit payments were for facility deliveries and for emergency transfers to hospital for obstetric care (\$4.59). Hospital referrals for at risk pregnancies were assigned a value of \$1.83.

To separate the income effect of the bonus payment from its incentive effect, budgets for health facilities in control districts (Phase II districts prior to 2008) were increased by an amount equivalent to the average of payouts to treated facilities on a quarterly basis. Health facilities had full discretion in spending incentive payments. Qualitative evidence suggests that incentive payments were primarily used by health facilities to increase staff salaries (Basinga et al., 2011).

Given the new schedule of payments, theory predicts that health providers will reallocate effort in order to equalize the returns at the margin across all the services offered in the clinic. One implication of this is that providers would reallocate effort towards higher margin services (see Eggleston, 2005; Dumont et al., 2008). Consistent with this Basinga et al. (2011) reported that "providers not only encouraged women to deliver in the facility during prenatal care, but some also commissioned community health workers to conduct outreach in the community to find pregnant women to deliver in the facility". One can think about this as a direct effect of the PBF program.

The PBF might also raise the rate of facility deliveries indirectly through its effect on improving quality. Notice that the payment formula rewarded overall quality through Q. To the extent therefore that quality enters into the demand function for facility births, women would be more likely to give birth in a health facility following the introduction of the PBF. An additional channel through which the PBF might affect rates of facility deliveries might be through decreased provider absence as a result of increased monitoring by district health officials (Duflo and Hanna, 2005; Banerjee and Duflo, 2006).[13]

4 Data

We use data from the Rwandan Demographic and Health Surveys (RDHS). The RDHS is a nationally representative household survey conducted approximately every five years. Households are selected in two stages: first, villages (also known as clusters or enumeration areas) are selected with probability proportional to the village size. A household listing is then conducted in each village and households are systematically selected from the household list.

We merged data from the 2005 and 2010 waves of the RDHS. This otherwise straightforward exercise was complicated by the fact that the 2005 RDHS contained province IDs but not district IDs.[14] To match clusters in 2005 to the appropriate district in 2010, we assigned each village/cluster in the 2005 data to a district in 2010 using geo-positioning data. Recent DHS surveys collect GPS coordinates for each village or cluster. These GPS readings are generally accurate up to 15-20 meters (ICF Macro, 2011). We matched each village/cluster in 2005 to a district in 2010 by calculating pairwise distances between each cluster in 2005 and all clusters in 2010 and then assigning each cluster to the district with the lowest sum of squared distances to cluster i. To put this formally, cluster i in 2005 was assigned to the district, k in 2010 that minimized

$$\min[A_1, A_2, ..., A_K]$$

where $A_k = \sum_j (C_i - C_j^k)$ and $C_i - C_j^k$ is the spherical distance between cluster i in 2005 and cluster j in district k in the 2010 DHS. The final study sample

[13]District supervisors were required to make quarterly visits to each facility to collect data on quality indicators. These visits were unannounced.

[14]In 2005, Rwanda was divided into 12 provinces. Following the decentralization, the 12 provinces were aggregated into 5 provinces.

consists of 11,153 births between 2000 and 2010. See distribution in Figure 1.

The RDHS collects detailed information about all pregnancies and birth within the five years preceding the survey date. All women between the ages of 15 and 49 who were permanent residents of the household or visitors present in the household on the night before the survey were eligible to be interviewed. For all births within the preceding five years, women are asked where the birth took place. Women can select from one of the following options: at home (either the respondent's home or some other home); in a health facility (either government or private), or elsewhere. We recoded this as a binary variable FACILITY equal to 1 if the birth took place in a health facility.

Details for each birth are collected including the sex of the child, the date of birth, birth order, whether it was a single or multiple birth, and survival status. For babies who died, information is collected about the age of death. Babies that were stillborn are coded as dying on the day of birth. For deaths within the first month, women are asked to provide the specific day of death if known. We have two mortality variables: $NEOMORT$ is equal to 1 if the baby died within the first month (between Day 0 and Day 30), and $EARLYMORT$, which is equal to 1 if the child died within the first 7 days. An early neonatal death is thought to more accurately reflect the quality of care received by the mother during childbirth (Ngoc et al., 2006).

For the most recent live birth, women are asked if they used prenatal care and how many prenatal care visits they attended. We constructed a binary variable equal to 1 if the woman made 4 or more prenatal visits over the course of the pregnancy. Information about the quality of prenatal care is also collected including whether the respondent was weighed, whether their blood pressure was taken, whether they were told about pregnancy danger signs and whether they received iron supplements. We use these prenatal care variables to test the validity of our instruments.

In addition to detailed birth information, the RHDS also collects information about respondent characteristics including age, religion, highest level of schooling completed, marital status, and relationship to the household head, and household characteristics including access to electricity, source of drinking water, type of toilet facilities, and type of roofing and flooring materials. These along with information on asset ownership are used in constructing a wealth index using principal component analysis.

4.1 Summary Statistics

Baseline characteristics for treated and control districts and the p-value from a test of differences in means are presented in Table 1. The baseline neonatal mortality rate is about 3.8 percent in the overall sample. About 26 percent of births took place in a health facility prior to implementation of the P4P program. Given the correlation between institutional deliveries and skilled birth attendance, it is not surprising to see that a nearly identical fraction of births were attended by a doctor or nurse. 12.4% of births had four or more prenatal visits during the course of the pregnancy. Average educational attainment is low in this sample, only 16.7% of women completed at least primary school. 88% of women in the sample are married, and their average age at the time of their first birth is 20.8 years. Only 2.8% of households reported having electricity, 15.5% and 49.5% of households respectively reported owning a bicycle or radio. The covariates are balanced across treatment and control districts.

5 Empirical Model

The aim of this paper is to identify the causal effect of a hospital birth on newborn mortality. The underlying structural model we wish to estimate is:

$$Mortality_i = \alpha + \beta HospBirth_i + X_{ij}\delta + \epsilon_i \quad (1)$$

$Mortality_i$ denotes the probability that infant i will die within 30 days of birth, $HospBirth$ is a dummy for whether the birth took place in a health facility, X_{ij} is a vector of exogenous birth characteristics, and ϵ is the error term. To obtain consistent estimates of β, $HospBirth$ has to be uncorrelated with the error term i.e. Cov($HospBirth$, ϵ)=0 , but this is unlikely to be true in practice.

As we have discussed, unobserved heterogeneity between childbearing women will induce a correlation between hospital births and the disturbance term in (1). Hospital births may be correlated with use of other birth inputs that have beneficial effects on infant health implying that estimating (1) with OLS will lead to estimates of β that are biased upwards.

We have cited evidence that women from richer households are more likely to give birth in a hospital (Gwatkin et al., 2000). Such women may be better nourished and therefore more likely to have healthier babies. It is also possible that women who choose hospital births are adversely selected. Women who anticipate a bad outcome based on private information about their risk may be more likely to choose a hospital delivery leading to estimates of β that are biased downwards. Note that accounting for mother-specific heterogeneity only partially solves this problem. Unobserved heterogeneity may also be birth-specific. There is an idiosyncratic component to delivery complications, and if women are only taken to hospital when they develop complications, this will induce a correlation between hospital births and the error term.

To identify a causal effect of hospital births, the ideal study would be an experiment in which some women were randomly assigned to deliver at home and others to deliver in a health facility. We are not aware of any Institutional Review Board that would approve such a study. An alternative is to find an exogenous variable (an instrument) that affects a woman's choice of delivery facility. We can then identify the treatment effect of a hospital birth by comparing health outcomes for infants born to mothers *that were induced by the exogenous instrument to give birth in a hospital* (but who would not otherwise), to the outcomes for mothers who were not exposed to the instrument. This is the approach we take in this paper.

The rollout of the government P4P program in Rwanda provides a plausibly exogenous shock to hospital births that allows us to recover the structural parameter of interest. This approach is similar in spirit to that of Evans and Lien (2005) who exploit the variation in access to prenatal care created by a bus strike in Pennsylvania to estimate the effect of prenatal care on birth outcomes. In the terminology of Angrist, Imbens and Rubin (1996), what we estimate is the Local Average Treatment Effect or $LATE$.

For the instrument to be valid, it must only affect newborn mortality through its effect on hospital deliveries. This rules out using cash transfers (or cash transfer programs) as an instrument because the additional income may be used to purchase market goods that have a direct effect on birth outcomes. Even if the cash is only paid out after the delivery has occurred, households may increase consumption in anticipation of this future income. Instrument validity also requires a sufficiently strong correlation with hospital births. Conditional on having a valid instrument, one can then recover consistent estimates of using two stage least squares. In the first stage we estimate the following model:

$$HospBirth_{ijt} = \pi_0 + \pi_1 Treat_{ijt} + X'_{ijt}\pi_2 + \theta_j + \eta_t + \nu_{ijt} \tag{2}$$

$HospBirth$ is equal to 1 if birth i in district j in year t took place in a hospital. $Treat$ is a dummy variable indicating whether the birth is in a treated district, X'_{ijt} is a vector of birth-level, mother-level, and household characteristics. We include controls for the sex of the baby, birth order, the time interval between the index birth and the last birth, whether it was a singleton or a multiple birth, and the age of the mother at birth. We also include birth month fixed effects and additional controls for mother's schooling, marital status, religion, relationship to the household head, and household wealth (measured using the standard wealth index). θ_j and η_t are district and birth year fixed effects respectively. ν is the disturbance term.

In the second stage we estimate a modified version of (1) where the endogenous variable $HospBirth$ is replaced with predicted values derived from (2). The model we estimate is:

$$Mortality_{ijt} = \gamma_0 + \gamma_1 Hosp\hat{B}irth_{ijt} + X'_{ijt}\gamma_2 + \theta_j + \eta_t + \mu_{ijt} \tag{3}$$

We include the same set of control variables used in (2). Conditional on the exclusion restrictions being valid, γ_1 is identified and has a causal interpretation. In the next section we examine a few potential threats to instrument validity.

5.1 Instrument validity

As we have discussed, for the instrument to be validly excluded from equation (1), it must have no effect on newborn mortality other than through increasing the probability of a facility birth. Given that the exogenous demand shifter is the pay-for-performance program, there is a concern that the P4P program may have also affected demand for other inputs that may have independent effects on infant health. An obvious worry is prenatal care.

There is now a critical mass of studies (mostly from developed countries) demonstrating beneficial effects of prenatal care on infant health outcomes (see for example Evans and Lien, 2005; Jewell, 2007; Rous, Jewell and Brown, 2004). From Table A.1 it is evident that the P4P program incentivized the delivery of prenatal care. Health facilities received incentive payments of 37 cents for each woman who completed at least 4 prenatal visits. If the P4P program also increased the demand for prenatal care, this would invalidate it as an instrument. In Table 2, Panel A we report results from a version of (2) where the outcome is prenatal care. We have two outcomes: (1) the total number of prenatal visits over the course of the pregnancy, and (2) the probability of completing at least four prenatal visits over the course of the pregnancy. We find no effects of the P4P program on either outcome.

It is possible that the P4P program did not influence quantity, but instead raised the *quality* of the prenatal visit. Notice that clinics received 46 cents for each pregnant woman that received a tetanus shot, and a similar amount for each pregnant woman that received malaria prophylaxis. Given the low baseline levels of tetanus vaccination and malaria prophylaxis during pregnancy, 20.3% and 5.2% respectively, clinics could increase their payments at relatively low cost by prescribing tetanus shots and malaria prophylaxis to all women who came in for a prenatal visit. Getting women to complete four prenatal visits on the other hand plausibly requires more effort and the marginal incentive is lower (37 cents). If the quality of each individual visit is higher, this might have a cumulative effect over the course of the pregnancy on the health of the newborn infant. There is much more limited evidence supporting the importance of prenatal care quality on infant health, nevertheless we explore this possibility in Table 3. We report results from a model in which we regress various indicators of prenatal care quality on the treatment dummy and additional controls. We find no effect on the probability of receiving a tetanus shot or malaria prophylaxis. We also find no effect of the treatment on a woman's probability of having her weight, weight (or blood pressure – not shown) taken during a prenatal visit.

Finally we address the possibility that the program increased postnatal child visits. Clinics received 18 cents for each preventive child visit and the same amount for each curative visit. Given that infections are responsible for up to 36 percent of neonatal deaths globally, and nearly half of all neonatal deaths in high mortality countries (Lawn, Cousens and Zupan, 2005), if the P4P program induced women to take sick infants to the clinic for treatment, this would pose a problem for instrument validity. In Table 2 Panel B we regress the following outcomes (i) probability of a postnatal child visit within the first week, and (ii) probability of a postnatal child visit within the first month, on the treatment dummy and additional controls. We find no effect on either outcome. In the models that follow, we report effects on neonatal mortality (deaths within the first month) separately from effects on early neonatal mortality (death within the first week). Taken together these results increase our confidence in the validity of the instrument.

6 Results

6.1 First-stage Results

A graphical treatment of the first stage is shown in Figures 2 and 3. In Figure 2, we averaged the rate of institutional deliveries over each phase separately for Phase I and Phase II districts: Phase 0 corresponds to the pre-implementation period (March 2000 – May 2006), Phase 1 of the rollout is from June 2006 to May 2008, and Phase 2 is from June 2008 to December 2010. As expected, baseline rates of institutional deliveries are similar in Phase I and II districts (from Table 1, the p-value from a test of means in 0.62). In the Phase 1 period, institutional deliveries grow at a faster rate in the treated districts (Phase I districts) and the lines diverge, but by Phase 2, when the previous control districts also become treated, the lines start to converge as the rate of growth slows down in the Phase I districts and Phase II districts catch up. In Figure 2, we plot means for Phase I and II districts by birth year.

In Table 4, we show the first stage regression results. In Column 1 no additional controls are included save for district and birth year dummies, in column 2 we include controls for a set of pre-determined birth characteristics, including the baby's gender, birth order, whether it was a singleton or multiple birth, the age of the mother and her partner at time of birth, and time interval since the last pregnancy. In column 3 we add in controls for respondent and household characteristics measured at the time of the survey. These include dummies for the mother's level of schooling, her marital status, religion and relationship to the household head, a dummy for whether she reports being solely responsible for making decisions about her health, household wealth quintile dummies, an urban/rural indicator, and interview month fixed effects. All the reported standard errors are clustered at the district level to allow for arbitrary correlation within district.

The results are consistent across all the specifications. We find a strong positive effect of the P4P program on institutional births. On average the P4P program increased institutional births by between 9 and 10 percentage points (about a 36 percent increase compared to baseline levels). In Panel B of Table 4, as a robustness check, we restrict the sample to only Phase I of the experiment. We obtain similar results. The coefficients on the control variables (omitted to save space) have the expected signs and magnitudes. We find for example that women in wealthier households are more likely to give birth in a health facility, rural households are less likely to give birth in a health facility and maternal schooling is strongly predictive of an institutional birth. We do not find any effect of maternal age at the time of birth or of birth order.

Looking at the coefficient on the year dummies (and also from Figures 2 and 3), it is clear that there is a strong time trend in the rate of institutional deliveries – facility births were also increasing in control districts. As we mentioned in Section 3, budgets for health facilities in control districts were increased by an amount equivalent to the average of the incentive payout to treated facilities to allow evaluators distinguish the income effect from the incentive effect of the P4P program. To the extent that these additional payments were invested in staff salaries or in improving health facility quality, this might help to explain the increase in control districts.

In the first stage we tried alternative specifications in which we interacted the treatment dummy with maternal schooling, household wealth, the gender of the baby, and with the urban/rural dummy to create multiple excluded instruments but the simple specification with only one instrument resulted in the strongest first stage. Results from these alternative specifications are not shown but are available on request.

6.2 Second-stage Results

In Table 5, we present the IV estimates. In Panel A, the dependent variable is neonatal mortality (deaths within the 1st 30 days), while in Panel B the dependent variable is early neonatal mortality (deaths within the first week). As before, Column 1 includes no additional controls save for district and birth year dummies, in column 2 we include controls for a set of pre-determined birth characteristics, including the baby's gender, birth order, whether it was a singleton or multiple birth, the age of the mother and her partner at time of birth, and time interval since the last pregnancy. In column 3 we add in controls for respondent and household characteristics measured at the time of the survey. These include dummies for the mother's level of schooling, her marital status, religion and relationship to the household head, a dummy for whether she reports being solely responsible for making decisions about her health, household wealth quintile dummies, an urban/rural indicator, and interview month fixed effects. We report Kleibergen-Paap F-statistics for the test of weak instruments rather than Cragg-Donald F-statistics because the errors are non-i.i.d. We also report Shea's partial R-squared.

The IV estimates are close to zero and are not statistically significant in any of the specifications. In the models where neonatal mortality is the dependent variable, the coefficient in fact reverses sign in two of the three specifications. Restricting the sample to include only Phase I and dropping the three districts that were affected by the redistricting exercise does not change the results or our conclusions (results available on request). As is clear from Table 5, we have strong instruments; the Kleibergen-Paap F-statistics are above 20 in all the specifications.

In Table 6 for comparison, we report results from an OLS model. Consistent with our earlier discussion, we find evidence of adverse selection into facility deliveries. A 'naive' OLS regression shows that babies born in hospital are *more* likely to die within the first 7 days of birth. On average, a hospital birth is associated with a 1-percentage point *increase* in the probability of dying within the fist week and in two of the three specifications, the coefficient is statistically significant. Unless hospitals worsen birth outcomes, which seem unlikely, this is evidence of adverse selection. Notice that controlling for an extensive set of birth, mother, and household level characteristics (as the epidemiologic studies are wont to do) does not reduce the selection bias, suggesting that selection is driven by unobserved differences in risk between hospital and home births. In Table 7 we show results from reduced form models in which mortality is regressed on the treatment dummy. Not surprisingly, the coefficients are also close to zero and statistically insignificant.

To put this results in the context of the literature, our results diverge from Panis and Lillard (1994) and Maitra (2004). We have already discussed the possibility that their estimates might be biased upwards. In addition, because we use data from a different region (Africa compared to Asia) and from a much more recent time period, the patterns of health care utilization and the quality of obstetric care are likely to be different.

The treatment effect estimated in this paper applies only to the population of 'compliers' i.e. those induced by the P4P program to give birth in a hospital. This is however likely to be the relevant $LATE$ for policy makers given the growing use of incentive programs to bring pregnant women into health facilities. If compliers are relatively low risk i.e. if these are women who would have healthy babies regardless of where they deliver, then the treatment effect of a hospital birth for this population is likely to be small.

Our results offer support to some of the more recent literature that find no effects of institutional delivery programs on neonatal mortality (see Mazumdar, Mills and Powell-Jackson, 2011). See also Titaley, Dibley and Roberts (2012) who find no effect of place of delivery on neonatal mortality despite finding a beneficial effect of prenatal care, and Hatt et al. (2009) who finds little evidence that the improvements in neonatal survival in Indonesia between 1986 and 2002 had anything to do with increased skilled birth attendance. Jehan et al. (2009) in a longitudinal follow-up study of 1,369 pregnant women in Pakistan from 20 weeks of gestation through delivery, finds high rates of neonatal mortality despite the fact that three-quarters of the women gave birth in a health facility. 80 percent of infants who died were born in a hospital or maternity clinic, and 85 percent of the infants who died received medical treatment in a hospital. Much of this recent literature concludes that quality of health professionals and health facilities are a key constraint to improvement in health outcomes.

As we mentioned in the literature review section, this paper is also related to the nascent literature that estimates the returns to formal health care usage in developing countries. Adhvaryu and Nyshadham (2011) find a strong effect of health facility care on malaria using a similar 2SLS approach in contrast to our findings here. However malaria and delivery complications such as birth asphysxia, eclampsia or obstructed labor are drawn from different ends of the severity distribution. Uncomplicated malaria is relatively easy to recognize and manage, the same cannot be said for birth complications. In a previous study in which we assessed provider competence across more than 200 primary health care clinics in Nigeria using clinical vignettes, we found that while three-quarters of surveyed providers were able to correctly diagnose uncomplicated child malaria, less than 20 percent of those providers were able to correctly diagnose a case of Pelvic Inflammatory Disease (PID) (Okeke, 2012). See also Wall et al. (2009) who in service assessments carried out in six African countries, found that 72 %–93 % of birth attendants were not trained in resuscitation, and basic resuscitation equipment such as bag-and-mask was missing from 53 %–84 % of facilities.

7 Conclusion

Current global health policies emphasize institutional deliveries as a pathway to achieving MDG targets of reductions in maternal and child mortality. Given this emphasis, a growing number of countries are implementing programs to incentivize women to give birth in a health facility. Conditional cash transfers (CCT) that reward women for giving birth in a hospital are becoming increasingly popular. As we however argued in the introduction, whether these programs pass a cost-benefit test depend crucially on the causal effect of hospital births on mortality. Estimates of this parameter are however hard to come by. While in the abstract, better care at the time of delivery should improve health outcomes, studies from many developing countries continue to show that care provided in a health facility does not necessarily equate to higher quality care. In this paper we have attempted to generate causal estimates of the effect of hospital births on newborn mortality. We find no statistically significant effects of facility births on neonatal mortality in any of our specifications.

There are several caveats we need to mention: first we were unable to study effects on maternal mortality because while the DHS collects data on maternal deaths (defined as deaths that occurred during pregnancy, childbirth, or within two months after the birth or termination of a pregnancy), the way the data is collected makes it impossible to determine whether the death occurred in that district or in a different district. In addition the DHS does not ask where the birth took place (hospital or home). It is therefore possible that while we find no beneficial effects for infants, there might be beneficial health impacts for women either through decreased mortality or decreased incidence of maternal complications. Secondly, null effects for neonatal mortality on average do not mean that there are no subgroups for whom facility deliveries are beneficial.

Neonatal mortality can be reduced through a range of institutional and community-based approaches, pre-, peri-, and post-partum (Darmstadt et al., 2005) and a fixation on increasing rates of institutional deliveries without addressing other key supply-side constraints is unlikely to result in the large gains that have been posited. Prevention of intrapartum-related deaths requires not only recognition of obstetric complications, but functioning referral and transport systems, and timely access to comprehensive care, many of which are lacking in developing countries. Attempting to increase institutional deliveries without addressing some of these other constraints is therefore likely to result in very little improvements in child health outcomes.

References

Adhvaryu, Achyuta, and Anant Nyshadham. 2011. "Healthcare Choices, Information and Health Outcomes." Yale University Working Paper.

Almond, Douglas, and Jr. Joseph J. Doyle. 2011. "After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays." *American Economic Review: Economic Policy*, 3(3): 1–34.

Almond, Douglas, Jr. Joseph J. Doyle, Amanda E. Kowalski, and Heidi Williams. 2010. "Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns." *Quarterly Journal of Economics*, 125(2): 591–634.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, 91(434): 444–455.

Banerjee, A., and E. Duflo. 2006. "Addressing absence." *Journal of Economic Perspectives*, 20(1): 117.

Banerjee, Abhijit, Angus Deaton, and Esther Duflo. 2004. "Health Care Delivery in Rural Rajasthan." *Economic and Political Weekly*, 39(9): 944–49.

Bartlett, A., M.E. Paz de Bocaletti, and M.A. Bocaletti. 1993. "Reducing perinatal mortality in developing countries: high risk or improved labour management?" *Health Policy and Planning*, 8: 360–368.

Basinga, Paulin, Paul J Gertler, Agnes Binagwaho, Agnes LB Soucat, Jennifer Sturdy, and Christel MJ Vermeersch. 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.

Bhalotra, Sonia. 2010. "Fatal fluctuations? Cyclicality in infant mortality in India." *Journal of Development Economics*, 93(1): 7–19.

Bharadwaj, Prashant, and Leah K. Nelson. 2010. "Discrimination Begins In The Womb: Evidence Of Sex-Selective Prenatal Investments." Mimeo, University of California, San Diego.

Campbell, Oona MR, and Wendy J Graham. 2006. "Strategies for reducing maternal mortality: getting on with what works." *The Lancet*, 368(9543): 1284–1299.

Chalumeau, M, B Salanave, M-H Bouvier-Colle, L de Bernis, A Prua, and G Bréart. 2000. "Risk factors for perinatal mortality in West Africa: a population-based study of 20 326 pregnancies." *Acta Pædiatrica*, 89(9): 1115–1121.

Chaudhury, Nazmul, Jeffrey S. Hammer, Michael Kremer, Karthik Muralidharan, and Halsey F. Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *Journal of Economic Perspectives*, 20(1): 91–116.

Darmstadt, Gary L, Zulfiqar A Bhutta, Simon Cousens, Taghreed Adam, Neff Walker, and Luc de Bernis. 2005. “Evidence-based, cost-effective interventions: how many newborn babies can we save?” *The Lancet*, 365(9463): 977 – 988.

Das, Jishnu, Jeffrey Hammer, and Kenneth Leonard. 2008. “The Quality of Medical Advice in Low-Income Countries.” *Journal of Economic Perspectives*, 22(2): 93–114.

Duflo, Esther, and Rema Hanna. 2005. “Monitoring Works: Getting Teachers to Come to School.” National Bureau of Economic Research Working Paper 11880.

Dumont, Etienne, Bernard Fortin, Nicolas Jacquemet, and Bruce Shearer. 2008. “Physicians’ multitasking and incentives: Empirical evidence from a natural experiment.” *Journal of Health Economics*, 27(6): 1436–1450.

Eggleston, Karen. 2005. “Multitasking and mixed systems for provider payment.” *Journal of Health Economics*, 24(1): 211–223.

Evans, William N., and Craig L. Garthwaite. 2009. “Estimating Heterogeneity in the Benefits of Medical Treatment Intensity.” National Bureau of Economic Research Working Paper 15309.

Evans, William N., and Diana S. Lien. 2005. “The benefits of prenatal care: evidence from the PAT bus strike.” *Journal of Econometrics*, 125(1-2): 207–239.

Filippi, Véronique, Carine Ronsmans, Oona MR Campbell, Wendy J Graham, Anne Mills, Jo Borghi, Marjorie Koblinsky, and David Osrin. 2006. “Maternal health in poor countries: the broader context and a call for action.” *The Lancet*, 368(9546): 1535–1541.

Grossman, Michael, and Theodore Joyce. 1990. "Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City." *Journal of Political Economy*, 98(5): 983–1007.

Gwatkin, D., S. Rustein, K. Johnson, R. Pande, and A. Wagstaff. 2000. "Socioeconomic differences in health, nutrition, and population – 45 countries." Washington, DC: World Bank.

Habibov, Nazim N., and Lida Fan. 2011. "Does prenatal healthcare improve child birthweight outcomes in Azerbaijan? Results of the national Demographic and Health Survey." *Economics & Human Biology*, 9(1): 56–65.

Harvey, Steven A., Yudy C.W. Blandón, Affette McCaw-Binns, Ivette Sandino, Luis Urbina, César Rodríguez, Ivonne Gómez, Patricio Ayabaca, Sabou Djibrina, and the Nicaraguan maternal and neonatal health quality improvement group. 2007. "Are skilled birth attendants really skilled? A measurement method, some disturbing results and a potential way forward." *Bulletin of the World Health Organization*, 85(10): 783–790.

Hatt, L., C. Stanton, C. Ronsmans, K. Makowiecka, and A. Adisasmita. 2009. "Did professional attendance at home births improve early neonatal survival in Indonesia?" *Health Policy and Planning*, 24(4): 270–278.

Hofmeyr, G. Justus, Rachel A. Haws, Staffan Bergström, Anne CC Lee, Pius Okong, Gary L. Darmstadt, Luke C. Mullany, Eh Kalu Shwe Oo, and Joy E. Lawn. 2009. "Obstetric care in low-resource settings: What, who, and how to overcome challenges to scale up?" *International Journal of Gynecology & Obstetrics*, 107, Supplement(0): S21–S45.

Houweling, Tanja AJ, Carine Ronsmans, Oona MR Campbell, and Anton E Kunst. 2007. "Huge poor-rich inequalities in maternity care: an international comparative study of maternity and child care in developing countries." *Bulletin of the World Health Organization*, 85: 745 – 754.

ICF Macro. 2011. "Incorporating Geographic Information into Demographic and Health Surveys: A Field Guide to GPS Data Collection." Calverton, Maryland, USA: ICF Macro.

Jehan, Imtiaz, Hillary Harris, Sohail Salat, Amna Zeb, Naushaba Mobeen, Omrana Pasha, Elizabeth M. McClure, Janet Moore, Linda L. Wright, and Robert L Goldenberg. 2009. "Neonatal mortality, risk factors and causes: a prospective population-based cohort study in urban Pakistan." *Bulletin of the World Health Organization*, 87(2): 130–138.

Jewell, R. Todd. 2007. "Prenatal care and birthweight production: evidence from South America." *Applied Economics*, 39(4): 415–426.

Joyce, Theodore. 1994. "Self-Selection, Prenatal Care, and Birthweight among Blacks, Whites, and Hispanics in New York City." *The Journal of Human Resources*, 29(3): 762–794.

Klemick, Heather, Kenneth Leonard, and Melkiory Masatu. 2009. "Defining Access to Health Care: Evidence on the Importance of Quality and Distance in Rural Tanzania." *American Journal of Agricultural Economics*, 91(2): 347–358.

Kusiako, T., C. Ronsmans, and L. Van der Paal. 2000. "Perinatal mortality attributable to complications of childbirth in Matlab, Bangladesh." *Bulletin of the World Health Organization*, 78(5): 621–627.

Lawn, Joy E., Anne CC Lee, Mary Kinney, Lynn Sibley, Wally A. Carlo, Vinod K. Paul, Robert Pattinson, and Gary L. Darmstadt. 2009. "Two million intrapartum-related stillbirths and neonatal deaths: Where, why, and what can be done?" *International Journal of Gynecology & Obstetrics*, 107, Supplement(0): S5–S19.

Lawn, Joy E, Simon Cousens, and Jelka Zupan. 2005. "4 million neonatal deaths: When? Where? Why?" *The Lancet*, 365(9462): 891 – 900.

Lee, Anne, Simon Cousens, Gary Darmstadt, Hannah Blencowe, Robert Pattinson, Neil Moran, G Hofmeyr, Rachel Haws, Shereen Bhutta, and Joy Lawn. 2011. "Care during labor and birth for the prevention of intrapartum-related neonatal deaths: a systematic review and Delphi estimation of mortality effect." *BMC Public Health*, 11(Suppl 3): S10.

Lim, Stephen S, Lalit Dandona, Joseph A Hoisington, Spencer L James, Margaret C Hogan, and Emmanuela Gakidou. 2010. "India's Janani Suraksha Yojana, a conditional cash transfer programme to increase births in health facilities: an impact evaluation." *The Lancet*, 375(9730): 2009–2023.

Maitra, Pushkar. 2004. "Parental bargaining, health inputs and child mortality in India." *Journal of Health Economics*, 23(2): 259–291.

Mazumdar, Sumit, Anne Mills, and Timothy Powell-Jackson. 2011. "Financial Incentives in Health: New Evidence from India's Janani Suraksha Yojana." SSRN Working Paper. Available at SSRN: http://ssrn.com/abstract=1935442.

Ngoc, Nhu Thi Nguyen, Mario Merialdi, Hany Abdel-Aleem, Guillermo Carroli, Manorama Purwar, Nelly Zavaleta, Liana Campódonico, Mohamed M. Ali, Justus G. Hofmeyr, Matthews Mathai, Ornella Lincetto, and José Villar. 2006. "Causes of stillbirths and early neonatal deaths: data from 7993 pregnancies in six developing countries." *Bulletin of the World Health Organization*, 84(9): 699–705.

Noonan, Kelly, Nancy E. Reichman, Hope Corman, and Dhaval Dave. 2007. "Prenatal drug use and the production of infant health." *Health Economics*, 16(4): 361–384.

Okeke, Edward. 2012. "First Do No Harm: Provider Quality as Constraints in Health Care Delivery." Working Paper.

Panis, Constantijn W. A., and Lee A. Lillard. 1994. "Health inputs and child mortality: Malaysia." *Journal of Health Economics*, 13(4): 455–489.

Powell-Jackson, Timothy, Joanna Morrison, Suresh Tiwari, Basu Neupane, and Anthony Costello. 2009. "The experiences of districts in implementing a national incentive programme to promote safe delivery in Nepal." *BMC Health Services Research*, 9(1).

Ronsmans, Carine, Mahbub Elahi Chowdhury, Nurul Alam, Marge Koblinsky, and Shams El Arifeen. 2008. "Trends in stillbirths, early and late neonatal mortality in rural Bangladesh: the role of public health interventions." *Paediatric and Perinatal Epidemiology*, 22(3): 269–279.

Ronsmans, C., S. Scott, S.N. Qomariyah, E. Achadi, D. Braunholtz, T. Marshall, E. Pambudi, K.H. Witten, and W.J. Graham. 2009. "Professional assistance during birth and maternal mortality in two Indonesian districts." *Bulletin of the World Health Organization*, 87: 416–423.

Rosenzweig, Mark R., and T. Paul Schultz. 1983. "Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight." *The Journal of Political Economy*, 91(5): 723–746.

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

Rusa, Louis, Miriam Schneidman, Gyuri Fritsche, and Laurent Musango. 2009. "Performance Incentives for Global Health: Potential and Pitfalls." , ed. Rena Eichler, Ruth Levine and the Performance-Based Incentives Working Group, Chapter Rwanda: Performance-Based Financing in the Public Sector, 189–214. Brookings Institution Press, Washington DC.

Soeters, Robert, Laurent Musango, and Bruno Meesen. 2005. "Comparison of two output based schemes in Butare and Cyangugu provinces with two control provinces in Rwanda." Global Partnership on Output-Based Aid (GPOBA), the World Bank.

Titaley, Christiana R, Michael J Dibley, and Christine L Roberts. 2012. "Type of delivery attendant, place of delivery and risk of early neonatal mortality: analyses of the 1994–2007 Indonesia Demographic and Health Surveys." *Health Policy and Planning*, 27(5): 405–16.

Wall, Stephen N., Anne CC Lee, Susan Niermeyer, Mike English, William J. Keenan, Wally Carlo, Zulfiqar A. Bhutta, Abhay Bang, Indira Narayanan, Iwan Ariawan, and Joy E. Lawn. 2009. "Neonatal resuscitation in low-resource settings: What, who, and how to overcome challenges to scale up?" *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics*, 107: S47–S64.

Walraven, Gijs, and Andrew Weeks. 1999. "The role of (traditional) birth attendants with midwifery skills in the reduction of maternal mortality." *Tropical Medicine and International Health*, 4(8): 527–529.

Weiner, Renay, Carine Ronsmans, Ed Dorman, Hilton Jilo, Anne Muhoro, and Caroline Shulman. 2003. "Labour complications remain the most important risk factors for perinatal mortality in rural Kenya." *Bulletin of the World Health Organization*, 81: 561 – 566.

WHO. 2006. *Working Together for Health - The World Health Report 2006.* World Health Organization.

WHO. 2012. "World Health Statistics 2012." World Health Organization, Switzerland, WHO Press, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland.

World Health Organization. 1999. "Reduction of maternal mortality. A joint WHO/UNFPA/UNICEF/World Bank statement." World Health Organization, Geneva.

You, Danzhen, Tessa Wardlaw, Peter Salama, and Gareth Jones. 2010. "Levels and trends in under-5 mortality, 1990–2008." *The Lancet*, 375(9709): 100–103.

Figure 1: Birth Sample Distribution

	Pre-Implementation (2000-2006)	Phase I (2006-2008)	Phase II (2008-2010)
Control Districts	2,521	878	1,037
Treated Districts	3,495	1,489	1,733
Total	6,016	2,367	2,770

Notes: Figures reported are the number of births

Figure 2: Trends in Facility Deliveries by Rollout Phase

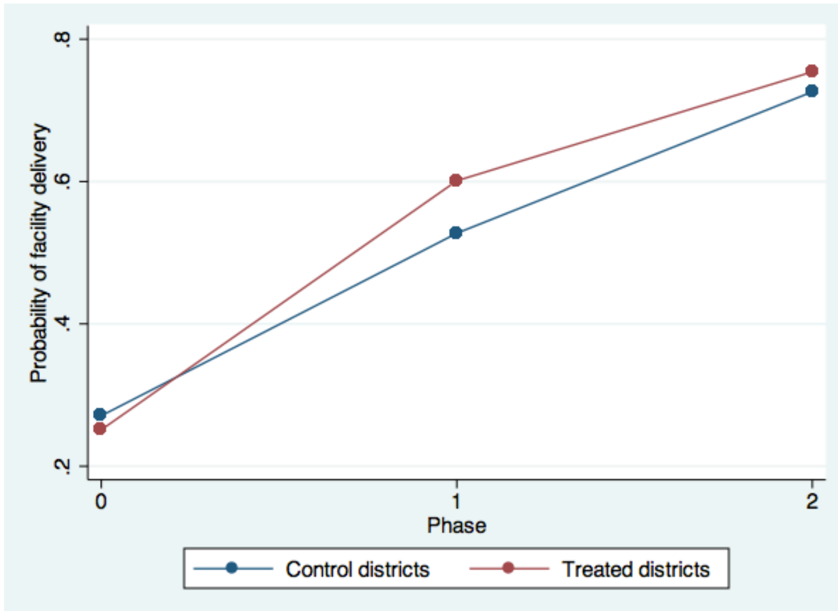

Figure 3: Trends in Facility Deliveries by Birth Year (2000-2010)

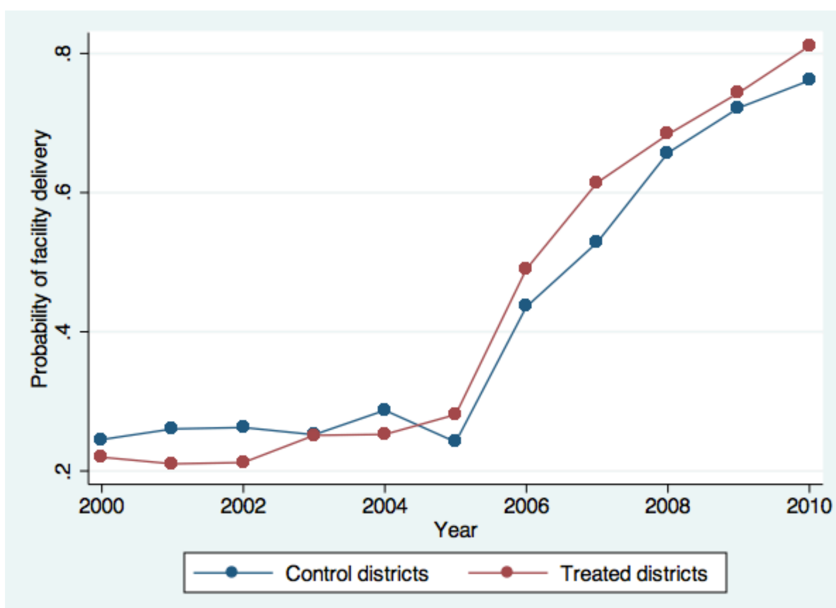

Table 1: Summary Statistics and Balance Test

	Total			Treated Districts		Control Districts			
	N	Mean	SD	Mean	SD	Mean	SD	Difference	p-value
Neonatal Mortality	6016	0.038	0.192	0.038	0.192	0.038	0.192	0.000	0.982
Institutional Deliveries	6016	0.259	0.438	0.251	0.434	0.271	0.445	-0.020	0.622
Skilled Birth Attendance	6015	0.263	0.440	0.256	0.436	0.274	0.446	-0.018	0.648
At least 4 prenatal visits	3559	0.124	0.330	0.137	0.344	0.108	0.310	0.029	0.217
Religion is catholic	6032	0.399	0.490	0.410	0.492	0.383	0.486	0.027	0.542
Mother completed Primary School	6032	0.167	0.373	0.169	0.375	0.163	0.370	0.006	0.784
Father completed Primary School	6032	0.245	0.430	0.230	0.421	0.266	0.442	-0.037	0.139
Married	6032	0.879	0.326	0.877	0.328	0.882	0.322	-0.005	0.730
Age at first birth	6032	20.783	3.221	20.648	3.119	20.970	3.349	-0.323	0.305
Household size	6032	5.647	1.978	5.626	2.005	5.675	1.938	-0.050	0.642
Household has electricity	6032	0.028	0.165	0.027	0.163	0.029	0.168	-0.002	0.894
Owns bicycle	6032	0.155	0.362	0.159	0.366	0.148	0.356	0.011	0.869
Owns radio	6032	0.495	0.500	0.492	0.500	0.500	0.500	-0.008	0.846

Notes: Standard errors used to calculate p-values are clustered at district level.

Table 2: Was there an effect of the P4P Program on Pre- and Post-natal visits?

	At least 4 prenatal visits			Total number of prenatal visits		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.0371	0.0385	0.0314	-0.00327	-0.000814	-0.0352
	(0.0289)	(0.0278)	(0.0277)	(0.0838)	(0.0813)	(0.0790)
Constant	0.122***	-0.427	-0.381	2.451***	0.965	2.822**
	(0.0219)	(0.299)	(0.340)	(0.0802)	(0.951)	(1.135)
Includes district & year FE	Yes	Yes	Yes	Yes	Yes	Yes
Includes birth characteristics	No	Yes	Yes	No	Yes	Yes
Includes other controls	No	No	Yes	No	No	Yes
Observations	7,323	7,323	7,323	7,323	7,323	7,323
R-squared	0.064	0.083	0.097	0.039	0.055	0.070

	Postnatal visit in the 1st month			Postnatal visit in the 1st week		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.000827	0.000354	-0.00124	0.00337	0.00414	0.00351
	(0.00801)	(0.00778)	(0.00682)	(0.00648)	(0.00628)	(0.00539)
Constant	0.0545***	0.131	0.103	0.0429***	0.0503	0.0327
	(0.0127)	(0.0768)	(0.0774)	(0.0105)	(0.0357)	(0.0431)
Includes district & year FE	Yes	Yes	Yes	Yes	Yes	Yes
Includes birth characteristics	No	Yes	Yes	No	Yes	Yes
Includes other controls	No	No	Yes	No	No	Yes
Observations	11,176	11,176	11,176	11,176	11,176	11,176
R-squared	0.020	0.029	0.036	0.017	0.025	0.032

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

Table 3: Was there an effect of the P4P Program on Quality of Prenatal care?

	Woman's Weight Measured			Woman's Height Measured		
	(1)	(2)	(3)	(1)	(2)	(3)
Treatment	0.00755	0.00884	0.00443	0.0106	0.00994	0.0119
	(0.00960)	(0.0101)	(0.00842)	(0.0151)	(0.0149)	(0.0133)
Constant	-0.00687	-0.344***	0.0395	-0.00898	-0.325***	0.264
	(0.0120)	(0.101)	(0.147)	(0.0170)	(0.108)	(0.166)
Includes district & year FE	Yes	Yes	Yes	Yes	Yes	Yes
Includes birth characteristics	No	Yes	Yes	No	Yes	Yes
Includes other controls	No	No	Yes	No	No	Yes
Observations	11,176	11,176	11,176	11,176	11,176	11,176
R-squared	0.543	0.561	0.590	0.279	0.293	0.312

	Tetanus Vaccination			Malaria Prophylaxis		
	(1)	(2)	(3)	(1)	(2)	(3)
Treatment	-0.00185	-0.00560	-0.00739	0.0360	0.0365	0.0336
	(0.0160)	(0.0162)	(0.0149)	(0.0251)	(0.0238)	(0.0232)
Constant	0.217***	0.863***	0.832***	0.912***	1.227***	0.853***
	(0.0225)	(0.248)	(0.257)	(0.0255)	(0.0836)	(0.239)
Includes district & year FE	Yes	Yes	Yes	Yes	Yes	Yes
Includes birth characteristics	No	Yes	Yes	No	Yes	Yes
Includes other controls	No	No	Yes	No	No	Yes
Observations	11,176	11,176	11,176	11,176	11,176	11,176
R-squared	0.047	0.272	0.281	0.548	0.569	0.576

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

Table 4: First Stage – What was the effect of the P4P Program on Institutional Births?

	Total Sample			Phase I Sample		
	(1)	(2)	(3)	(1)	(2)	(3)
Treatment	0.0927***	0.0996***	0.0925***	0.106***	0.115***	0.112***
	(0.0228)	(0.0244)	(0.0188)	(0.0288)	(0.0309)	(0.0204)
Year = 2001	-0.00309	0.00500	0.0128	-0.00301	0.00525	0.0131
	(0.0210)	(0.0195)	(0.0178)	(0.0209)	(0.0193)	(0.0178)
Year = 2002	-0.00314	0.00947	0.0192	-0.00328	0.00890	0.0188
	(0.0212)	(0.0184)	(0.0200)	(0.0212)	(0.0184)	(0.0202)
Year = 2003	0.0202	0.0272	0.0381*	0.0209	0.0282	0.0384*
	(0.0229)	(0.0211)	(0.0212)	(0.0227)	(0.0209)	(0.0213)
Year = 2004	0.0321	0.0402**	0.0569***	0.0327	0.0395**	0.0561***
	(0.0209)	(0.0187)	(0.0194)	(0.0206)	(0.0184)	(0.0194)
Year = 2005	0.0326	0.0385	0.0508**	0.0336	0.0383	0.0496**
	(0.0290)	(0.0304)	(0.0234)	(0.0291)	(0.0298)	(0.0233)
Year = 2006	0.199***	0.191***	0.103*	0.196***	0.188***	0.0870
	(0.0317)	(0.0342)	(0.0507)	(0.0327)	(0.0356)	(0.0513)
Year = 2007	0.290***	0.284***	0.204***	0.284***	0.275***	0.181***
	(0.0308)	(0.0328)	(0.0521)	(0.0335)	(0.0352)	(0.0542)
Year = 2008	0.360***	0.343***	0.269***	0.365***	0.349***	0.264***
	(0.0294)	(0.0335)	(0.0496)	(0.0341)	(0.0385)	(0.0534)
Year = 2009	0.406***	0.380***	0.311***			
	(0.0373)	(0.0417)	(0.0596)			
Year = 2010	0.466***	0.431***	0.360***			
	(0.0386)	(0.0411)	(0.0591)			
Constant	0.235***	0.691***	1.011***	0.233***	-0.0514	0.335
	(0.0188)	(0.233)	(0.262)	(0.0180)	(0.251)	(0.340)
Includes district & year FE	Yes	Yes	Yes	Yes	Yes	Yes
Includes birth characteristics	No	Yes	Yes	No	Yes	Yes
Includes other controls	No	No	Yes	No	No	Yes
Observations	11,153	11,153	11,153	8,383	8,383	8,383
R-squared	0.186	0.245	0.300	0.101	0.164	0.239
F-test of excluded instrument	16.56	16.68	24.12	13.58	13.85	30.15
P-value	0.0007	0.0007	0.0001	0.0017	0.0016	0.0000

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

Table 5: Effect of Institutional Births on Neonatal Mortality – 2SLS Estimates

	Neonatal Mortality			Early Neonatal Mortality		
	(1)	(2)	(3)	(1)	(2)	(3)
Institutional Birth	-0.00177	0.00217	0.00375	-0.000858	-0.00147	-0.00257
	(0.0854)	(0.0746)	(0.0800)	(0.0730)	(0.0605)	(0.0666)
Includes district & year FE	Yes	Yes	Yes	Yes	Yes	Yes
Includes birth characteristics	No	Yes	Yes	No	Yes	Yes
Includes other controls	No	No	Yes	No	No	Yes
Observations	11,153	11,153	11,153	11,153	11,153	11,153
Shea Partial R^2	0.00233	0.00287	0.00261	0.00233	0.00287	0.00261
Kleibergen-Paap F-statistic	20.80	25.80	22.99	20.80	25.80	22.99

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

Table 6: Effect of Institutional Births on Neonatal Mortality – OLS Estimates

	Neonatal Mortality			Early Neonatal Mortality		
	(1)	(2)	(3)	(1)	(2)	(3)
Institutional Birth	0.00800	0.00556	0.00755	0.0103**	0.00805	0.00989*
	(0.00495)	(0.00513)	(0.00539)	(0.00449)	(0.00480)	(0.00484)
Constant	0.0312***	0.190**	0.0654	0.0249***	0.184***	0.105
	(0.00594)	(0.0666)	(0.0896)	(0.00676)	(0.0595)	(0.0695)
Includes district & year FE	Yes	Yes	Yes	Yes	Yes	Yes
Includes birth characteristics	No	Yes	Yes	No	Yes	Yes
Includes other controls	No	No	Yes	No	No	Yes
Observations	11,153	11,153	11,153	11,153	11,153	11,153
R-squared	0.002	0.032	0.036	0.002	0.030	0.035

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

Table 7: Effect of Institutional Births on Neonatal Mortality – Reduced Form Estimates

	Neonatal Mortality			Early Neonatal Mortality		
	(1)	(2)	(3)	(1)	(2)	(3)
Institutional Birth	-0.000164	0.000216	0.000347	0.00147	0.00148	0.00129
	(0.00816)	(0.00765)	(0.00764)	(0.00700)	(0.00619)	(0.00627)
Constant	0.0231**	0.185**	0.0761	0.0164**	0.138***	0.0750
	(0.00860)	(0.0678)	(0.0908)	(0.00744)	(0.0287)	(0.0476)
Includes district & year FE	Yes	Yes	Yes	Yes	Yes	Yes
Includes birth characteristics	No	Yes	Yes	No	Yes	Yes
Includes other controls	No	No	Yes	No	No	Yes
Observations	11,153	11,153	11,153	11,158	11,158	11,158
R-squared	0.001	0.031	0.035	0.001	0.029	0.034

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

Appendix

Table A.1: Schedule of Payments for services incentivized by the P4P Program

Indicator	**Unit Payment (USD)**
Curative care visit	0.18
First prenatal care visit	0.09
Completion of 4 prenatal visits	0.37
First time family planning visit	1.83
Contraceptive resupply visit	0.18
Delivery in the facility	4.59
Child preventive care visits (0 - 59 months)	0.18
Tetanus vaccine received during prenatal care	0.46
Malaria prophylaxis received during prenatal care	0.46
At risk pregnancies referred to hospital for delivery	1.83
Emergency transfers to hospital for obstetric care	4.59
Completed child vaccinations	0.92
Malnourished children referred for treatment	1.83
Other emergency referrals	1.83