Does Distance matter for Institutional Delivery in Rural India? An Instrumental Variable Approach

Kumar, Santosh and Dansereau, Emily and Murray, Chris

Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA

1 October 2012

Preliminary draft: Do not cite without author's permission

Does Distance matter for Institutional Delivery in Rural India? An Instrumental Variable Approach

Santosh Kumar
University of Washington

Emily Dansereau
University of Washington

Chris Murray
University of Washington

Keywords: In-facility delivery, access, distance, India.

Corresponding author: Santosh Kumar, Institute for Health Metrics and Evaluation, Department of Global Health, University of Washington, Seattle, WA, 98121, U.S.A (e-mail: skumar3@uw.edu).

ABSTRACT:

Skilled attendance at childbirth is crucial for decreasing maternal and neonatal mortality, yet many women in low- and middle-income countries deliver outside of health facilities, without skilled help. Distance to health facility is considered to be an important non-monetary barrier that impede utilization of health facilities. In this paper, we examine if access to health facilities affects institutional births in a resource-constrained country like India. We use *Two-Stage Residual Inclusion* (2SRI) and IV-Probit models to account for endogenous placement of health facilities. Our findings indicate that women living closer to health facilities have a higher probability of giving birth in health facility. An increase of one kilometer in the distance to the nearest health facility decreases the probability of institutional delivery by 4.4%. The results from policy simulation suggest that restricting the maximum distance to 5 kilometers would increase institutional delivery by 10%. Overall, our findings show that distance is an important barrier to service utilization and increasing the density of health facilities or improving transport infrastructure may be an important policy tool to improve facility-based delivery in developing countries.

Keywords: In-facility delivery, access, distance, India.

1 INTRODUCTION

Institutional delivery is often promoted to reduce the risk of maternal and neo-natal mortality. Maternal mortality has declined globally by 50% in the last two decades, however; the incidence of maternal deaths remains strikingly high in many African and South Asian countries, including India, which faces one of the greatest burdens of maternal and neonatal deaths (Hogan et al, 2010). These regions also have very low prevalence of skilled birth delivery. Identifying and quantifying the barriers to institutional delivery could be important for reducing maternal and neo-natal mortality.

Safe delivery is also hypothesized to improve health seeking behavior and health care practices related to postnatal care. Children born at health facility are more likely to be vaccinated and breastfed (Odiit and Amuge, 2003). Properly vaccinated and adequately breastfed children are less likely to malnourished and have better health. There is a growing literature on the interaction between early-life health and human capital accumulation (see Bleakley, 2010 for detail). Bleakley (2010) discusses that poor childhood health might depress the formation of human capital, which in turn could affect lifetime income either through schooling or labor-market productivity channel. Therefore, institutional delivery can also be thought as an investment in human capital and can play an important contributory role in the development process of an economy.

In India, both child mortality (especially neonatal mortality) and maternal mortality are high. In 2008, more than 63,000 maternal deaths and 1 million neonatal deaths occurred in India, representing 20% and 30% of the global burdens, respectively (WHO, 2010). Most of these deaths occurred because women delivered in risky environments lacking life-saving equipment, hygienic conditions, and supervision by skilled attendants. Per the latest estimates, close to 60% of births in rural parts of India still occur at home in the absence of skilled birth attendants (IIPS, 2010).

Previous literature has identified three main types of barriers to health care service utilization: (1) delay in deciding to seek care; (2) delay in reaching an adequate health care facility; and (3) delay in receiving adequate care at that facility. This is cited as the "three delays" model in the literature (Thaddeus and Maine, 1994). Delay 1 relates to lack of health related knowledge; delay 2 is caused by physical inaccessibility (*distance*) of facilities, lack of transportation, difficult terrain and high travel costs; and delay 3 occurs due to inadequate availability of equipment, drugs, and medical staff.

In this paper, we focus on delay 2, which is related to access and distance to health facilities. Distance is known to be one of the most important non-monetary barriers that impedes access to healthcare, especially in rural areas. Large geographic distances to a healthcare provider coupled with a lack of transportation facilities can adversely affect the utilization of health services and health outcomes (Sarma, 2009). Despite the importance of access on healthcare usage, very little evidence is available on the *causal* impacts of distance to health facility on healthcare utilization. The objective of this paper is to explore this relationship in a causal framework in a resource constrained country like, which relies heavily on a decentralized public health system.

The extent to which distance to health facilities may affect utilization of health services and health outcomes is an empirically open question and depends greatly on contextual factors. For instance, geographical distance may become irrelevant in a setting with high-quality health and transport infrastructures (Matthew et al., 2005). Some studies have shown that households are keen to travel longer distances for high-quality care, and there is some evidence of a trade-off between distance and quality of care (Collier, Dercon, and Mackinnon, 2002). Collier et al. (2002) finds that usage of health facilities is sensitive to the quality of health care and not just to distance.

However, these findings may not apply in a setting plagued with limited health services, inadequate transport infrastructure, and poor populations, such as rural India. In such setting physical access often presents a fundamental and insurmountable barrier to accessing adequate care at birth and therefore plays a central role in causing high maternal mortality

(Gething et al., 2012).

Furthermore, rural households are particularly deprived as they often lack efficient means of transportation. Even among women who do reach a facility, some are almost beyond help by the time they arrive (Ronsmans and Graham, 2006). Some research (including in India) has found that geographical access is more important than socioeconomic factors for the usage of maternal health services, particularly in rural areas with limited health services (Sawhney, 1993; Elo, 1992). Another study found that the distance to the nearest hospital was an important determinant of institutional delivery in rural India, although they identified wealth status as the most influential factor (Kesterton et al., 2010). While estimating the demand for outpatient care, Sarma (2009) found that distance to formal health care facilities negatively impacted the health care demand and the effect was modified by access to transport. In a similar study in Zambia, distance was found to have a negative impact on the probability of seeking professional care (Hjortsberg, 2003).

Against this background, the primary aim of this study is to assess the importance of the distance to the nearest health facility in determining the place of delivery (facility vs home), while also examining the influence of socio-economic status, mother's age, mother's education, mother's religion and region. We also take into account factors that modify the impact of distance, including access to roads and motorized vehicles.

Due to a lack of suitable data, very few studies that is nationally representative, have examined the effect of distance on maternity care after controlling for individual and household-level variables. The majority of health surveys do not contain information on the distance to health facilities. The recent Demographic and Health Surveys (DHS) have collected geographic coordinates, but the scrambled nature of GIS coordinates makes their use prone to errors and inconsistent. Fortunately, we have access to data that provides more reliable information on the distance from a village cluster to the nearest health facilities. Specifically, we use the District Level Household Survey (DLHS-3), implemented in 2007-2008 in all Indian districts, to test the hypothesis that better access to health facilities improves the probability of an in-facility delivery (IFD).

This study contributes to the existing literature on barriers to access to health care and utilization of health services. This is especially relevant in a resource-constrained setting where access to health care is not universal. This paper has four main strengths and differs from previous studies on these four dimensions. First, the data used in this study measure distance well, while very few other household data have distance measures. Second, the strength of this paper also lies in the large sample size. Most of the previous studies are either case studies or are focused on one particular region of a country. In contrast, our study uses a recent nationally representative survey and includes close to 200,000 births. The most similar national analysis for India was conducted using national family Health Survey (NFHS) 1 & 2 data collected in 1992 and 1998 respectively (Kesterton et al., 2010). The sample size in this study was less than 22,000 births, far less than the sample size of our study. Third, to the best our knowledge, this is the first study that tackles the endogeneity issues head on and provides a robust causal relationship between distance to facility and IFD. Finally, the study also contributes to the debate on relative importance of access versus quality of care. Policymakers in developing countries are deciding whether to spend resources in increasing the density of facilities or improving quality of care in existing facilities.

The remainder of this paper is organized as follows: section 2 presents the empirical methodology adopted in this paper, section 3 describes the data, section 4 presents the main results and finally section 5 concludes and discusses policy implications.

2 CONCEPTUAL FRAMEWORK

The health care demand model that we estimate in this study is similar to Borah (2006). Consider a representative consumer i who derives utility from consumption of medical and non-medical goods. The utility maximizing behavior of the individual is represented by the utility function U(C,H), where C is the consumption of non-medical goods and H is the expected level of improvement in health after receiving care. The usual assumptions are made about the utility function: $U_c > 0$, $U_{cc} < 0$, $U_h > 0$, $U_{hh} < 0$.

The model also assumes that an individual faces J alternative health care providers. However, for simplicity, we assume that the individual has two options: (a) seek modern care (J=1), or (b) use home-based treatment (J=2). The level of improvement in health (H) after receiving treatment from provider j is $H_j(X, Z_j)$, where X is observed attributes of the individual or household, and Z_j is a set of provider j attributes. For simplicity, assume that this consumer spends her income only on consumption of goods and health. Therefore, combining all these information, consumer i's utility maximization problem can be written as:

$$max_{C_{ij}, j \in (1,2)} U(C_{ij}, H_{ij}) \tag{1}$$

subject to the budget constraint:

$$C_{ij} + P_{ij} = Y_i; \; where \; P_{ij} = p_{ij} + d_{ij} \tag{2}$$

where Y is the income of an individual, C_{ij} is the amount spent on consumption goods, P_{ij} is the total amount spent on seeking care by provider j. P_{ij} is sum of direct (p_{ij}) and indirect cost (d_{ij}). d_{ij} may be the transport cost or foregone wages due to time spent on traveling to the provider j.

The consumer will choose provider j that gives her the highest expected utility. Giving birth in a health facility will certainly provide higher utility, however it involves cost as well (both direct and indirect). The consumer will choose to deliver in health facility only if the expected benefit of in-facility birth is higher than the total cost. One prediction of this model is that utility of consumer i from choosing provider j will be lower as the distance to provider j increases because greater distance implies increased travel time and that in turn results into higher transport cost & forgone wages (d_{ij}). The following section shows the empirical specification used to test this prediction.

3 EMPIRICAL FRAMEWORK

This section describes the econometric methodology used to investigate the model empirically, which is to estimate the causal effect of distance to closest health facility on IFD. Since, our main outcome variable, IFD, is binary, we first estimate the following probit model:

$$P(IFD_{ivd} = 1) = \phi(\alpha + \beta_1 DIST_v + \beta_2 X_{ivd} + \theta YOB_i + \mu_d + \epsilon_{ivd}) \quad (3)$$

where IFD_{ivd} is a binary variable indicating IFD by women i, DIST is distance to nearest health facility in kilometers, X_{ivd} is a vector that includes household and village-level variables as described below, YOB is year of birth dummies capturing the time trend, μ_d is a fixed effect unique to a district that captures the time-invariant differences across districts. Finally ϵ_{ivd} is the error term that captures the impact of all other unobserved variables that vary across individuals, villages, and districts and ϕ is the standard normal cumulative distribution. All models use survey weights to account for sample design and population weighting and standard errors are adjusted for clustering at the district level.

X_{ivd} represents characteristics of women, households and village, such as mother's age at birth; mother's education; whether the household belong to schedule caste/tribe; religion; household wealth quintile; whether the household possess motorcycle/car/truck; JSY payment, and finally whether the village is connected to an all weather road.[1] In-facility delivery includes births in private as well as public health facilities. Eq. (1) is estimated in STATA/SE 11 using the "probit" command and marginal effects are estimates using the "margins" command.

Distance to nearest health facility (DIST) in Eq. (1) is a continuous variable and is measured in kilometers. The village module of the DLHS-3 survey reports the distance

[1] Janani Surakhsa Yojana (JSY) is a conditional cash transfer scheme to promote birth in a health facility.

from the village to each type of facility in kilometers[2], and DIST is shortest distance to one of these six facilities. For some households, DIST could be distance to PHC, while for other households, it could be distance to DH. We also estimate a variant of model (1) that uses distance as a categorical rather than continuous variable. DIST is divided into three categories: less than 5 kilometers, between 5 and 10 kilometers, and greater than 10 kilometers, with less than 5 kilometers as the reference category.

Our preferred model is Probit, nevertheless for comparison we also estimate a Linear Probability Model (LPM) wherever possible. Previous studies have shown that if goal of the study is to estimate the average effect, which we often are interested in, LPM and Probit provides qualitatively similar results.[3]

Under the assumption that distance to the nearest health facility is purely exogenous, the probit estimates in equation (1) provides the causal estimate of access on IFD. However, this assumption is unlikely to be true as facility placement may not be plausibly random. Although many previous studies have assumed distance to be an exogenous variable, one could still argue that the placement of health facilities may be non-random. Health facilities may have been set up in areas that have poor health outcomes and high morbidity burdens. Endogeneity of DIST may also be due to measurement error and unobserved heterogeneity. There could be unobserved omitted variables that may affect the outcomes and placement of the health facilities. For example, a village with a larger, educated and political active population could attract health facilities and may also have better outcomes due to higher health knowledge and better health behavior of the population. This will bias the estimates in equation (1).

This discussion suggests that the exogeneity assumption for DIST in (1) may not be true. Therefore, we attempt to deal with these concerns by employing instrumental variable

[2]The village questionnaire reports distance to nearest government dispensary, primary health center (PHC), community health center (CHC), district hospital (DH), private clinics, and private nursing home.

[3]Angrist and Pischke (2009, Chapter 3) show that LPM is a good option for different kinds of limited dependent variables. Hellevik (2009) also makes a compelling case for choosing LPM over logit or probit.

estimation (IV) to consistently estimate the causal impacts of DIST on IFD. Instrumental variables (or instruments) are variables which (i) are correlated with the endogenous variable (distance to the nearest health facility) and (ii) are not correlated with the error term in the outcome equation. If these two conditions are satisfied, then one can identify and estimate a consistent estimate of causal effect of distance on IFD.

There are many variants of IV estimation for binary dependent variable, including two-stage residual inclusion (2SRI), IV-Probit, IV-LPM, and full information maximum likelihood (FIML) bivariate probit. Bhattacharya et al. (2006) reviewed the latter three estimators and showed that in a test on simulated data, the IV-Probit and IV-LPM estimators exhibited greater bias than the 2SRI and FIML estimator even in the presence of misspecification of the distribution of the error terms. Thus, the 2SRI estimator appears to be a good choice for this analysis, however as a robustness check, we also estimate IV-Probit and IV-LPM.

We estimate 2SRI model as suggested by Terza et al. (2008). The basic empirical model for our analysis is

$$DIST = \alpha + \beta_1 Z_v + \beta_2 * X_{ivd} + \theta YOB_i + \mu_d + \epsilon_{ivd} \tag{4}$$

$$IFD_{ivd} = \alpha + \beta_3 \widehat{D}IST_v + \beta_4 * X_{ivd} + \theta YOB_i + \mu_d + \epsilon_{ivd} \tag{5}$$

Eq. (2) is the first stage regression where the endogenous variable, DIST, is regressed on the instrument (Z) and other exogenous variables. In the second stage (Eq. 3), the outcome, IFD, is regressed on the predicted value of the endogenous variable, DIST, from the first stage (Eq. 2) along with other exogenous variables. For 2SRI, the endogenous DIST variable and the predicted residuals from the first-stage estimations are included in the second stage.[4]

X_{ivd} in Eq. (2) and (3) is the same set of controls used in Eq. (1) that are assumed to be exogenous. The parameters in Eq. (3) are identified uniquely by the assumption that the

[4]We use ”IVPROBIT and IVREG2” command, respectively, in STATA/SE 11 to estimate the IV-Probit and IV-LPM models.

instrument Z does not belong in Equation (3) i.e. assumption (ii) above. The excludability of the instrument Z from (3) is an assumption that is inherently un-testable since one can never observe the error term and so one has to rely on a priori reasoning to justify the assumption. Since, our IV model has one endogenous variable and one instrument, the model is exactly identified and there is no need for an over-identification test.[5]

The bias in the OLS estimates is not *a priori* established. For instance, an area with higher density of facilities (less distance) may also have better health behavior. In this case, the actual effect of distance will be underestimated. By contrast, if facilities were placed in areas with high disease burden, then the actual effect of distance will be overestimated. Consequently, it is not easy to correctly estimate the effect of distance on in-facility delivery by using OLS regression model.

Instrument: Finding good instruments for DIST is key in instrumental variables estimation. Valid instruments should affect DIST significantly, but only indirectly affect IFD through its direct effect on DIST. We chose instruments based on the prior studies in which DIST is an endogeneous variable. In development economics literature, a widely used instrument for distance to schools is distance to other village-level infrastructures. For example, while examining the effect of school supply constraints on educational outcomes in Ghana, Lavy (1996) used distance to a public telephone and post office as instruments for distance to middle school. Similar in spirit, Mukhopadhyay and Sahoo (2012) have constructed an index variable by Principal Component Analysis (PCA) that includes distance to the nearest telephone booth, police station, public distribution shop and bank and used this index to instrument the distance to nearest school.

Following this strand of literature, we use distance to non-health institutions of development as the instrument. Specifically, we construct an index variable using PCA that includes distance to the nearest town, distance to the district headquarter, distance to the nearest railway station, and distance to the nearest bus stop. The implicit assumption is that im-

[5] Angrist and Krueger (2001) is an accessible overview of the IV technique.

proving access to village facilities is correlated with the opening of health facilities but has no direct effect on IFD once we have controlled for other regressors.

4 Data

We use data from the third waves of the DLHS-3, which is a health survey covering family planning, maternal and child health, reproductive health of ever-married women and adolescent girls, and use of maternal and child health care services. The survey is representative at the district level. DLHS-3 was implemented during 2007 and 2008, interviewing 643,944 ever-married women between 15 and 49 years of age from all 611 districts and 34 states in India. Every woman was asked about her fertility history in the five years preceding the survey (since January 1, 2004). Out of the 643,944 ever-married women interviewed, 504,272 (78%) resided in rural areas and only 177,294 women had given birth in the last five years. After dropping observation with missing values, the final analytical sample comprises of 158,897 women. The household and women survey is integrated with the village and facility survey.

In the women module, for the last live/still births born during the three years preceding the survey, women were asked where (place) their children were born, who assisted during the delivery, characteristics of delivery, and any problems that they faced during the delivery. We used the information on place of the last delivery to construct our outcome variable. IFD is our main outcome variable that includes births either at public and private health facilities. The reference category is out-of-facility delivery which includes non-institutional births mainly at home. We exclude births in sub-centers, ayush clinics and non-governmental organization/trust clinics, as less than 1% of births were recorded in these facilities.

The main independent variable is the distance to the nearest public or private health facility. We use information on the distance to different types of health facilities collected in the village survey. The village module gathered information on distance to the nearest PHC, CHC, DH, private nursing home, and private doctors. The distance is measured from the village center and is reported in kilometers. The average distances to the nearest PHC, CHC and DH are is 9.06 kilometers, 17.66 kilometers and 32.57 kilometers, respectively. The average distances to the nearest private clinic and private hospital are 10.38 kilometers and 18.69 kilometers, respectively. The identifying assumption in our paper is that women visit the nearest health facility for delivery care, or that access to a nearby facility helps with referral to a higher-level facility.

Furthermore, we additionally control for potential confounders that may affect the outcome variable IFD. In India, utilization of health services varies greatly by social groups and caste, so we included household caste and religion as other explanatory variables. Previous research has also shown that mother's age and education are dominant predictors of facility delivery. To capture this, we add mother's age at birth of the last child and mother's education in our model. An asset-based wealth index is also included to capture the importance of financial resources in delivery care. Finally, we also include access to a drivable road and ownership of a car or motorcycle to capture the effect of transport barriers on IFD. Sampling weights are used in the regression models to account for sampling design.

5 RESULTS

5.1 Summary statistics

Table I provides the means and standard deviations for the variables used in the analysis. The institutional delivery rate is 36%, implying that 64% of births occur at home under risky and unsafe environments. The mean distance to the nearest health facility is 4.93 kilometers (this includes public as well as private facilities). About 58% of the women in rural India live within 5 kilometers of health facility and a quarter of women live between 5 to 9 kilometers. For 18% of women, the nearest health facility was located beyond 10 kilometers.

The IFD delivery rate is highest in the 0 to 4 kilometer category. About 42% of women living within 5 kilometer of health facility gave birth at the facility while the IFD rate was

32% and 26% for women living between 5 and 9 kilometers and more than 10 kilometers away from the nearest facility respectively. The average distance to PHC is 8.70 kilometers and the mean distance to district hospital is 34 kilometers.

The average age at birth was 25 years and about 34% of the women completed primary school, while only 7% of the women completed secondary school (more than 10 years of schooling). The majority of the women are Hindu and belong to disadvantaged social groups (SC/ST, 40%). Very few households had access to a car or motorcycle (2%), but the majority of the villages were connected with an all-weather drivable road (85%). About half of the sample (50%) belong to the bottom two wealth quintile categories implying that poor women constitute the major portion of the sample.

[Table I]

5.2 Regression results

Table III presents results from two models where distance to nearest health facility (DIST) is the main covariate of interest and probability of in-facility delivery (IFD) is the main outcome. The first model is the LPM and the second model is the standard probit model. In addition to the possible confounder variable, columns (2) and (4) include district fixed effects to account for time-invariant district characteristics.[6] Marginal effects (M.E.) are reported in columns (3) & (4). The coefficients for distance in all the four columns are negative and statistically significant, suggesting that distance to the nearest health facility is inversely associated with the probability of in-facility delivery.

As per the results in column (1), a one kilometer increase in the DIST results in a 0.5 percentage point reduction in the probability of IFD; the coefficient is statistically significant at the 1% level. Given that the mean IFD is 36%, this means a decrease in IFD by 1.4% (0.005/0.36). The coefficients on road access and ownership of a car/motorcycle are positive and statistically significant at 1% level (column 1).

[6] We do not report the coefficients for such variables to keep the presentation simple.

Inclusion of district fixed-effects reduces the coefficient by half, suggesting that district-level characteristics are also important in explaining the variation in IFD (column 2). The probability of IFD deceases to 0.2 percentage points in column (2). Columns (3) & (4) report marginal effects from the probit model and results are not very different from the results in first two columns in Table III. However, since our dependent variable is binary, our preferred specification is probit with district fixed-effect (column 4, Table III).

Our preferred model in column (4) shows that the probability of IFD decreases by 0.3 percentage points as the distance to health facility increases by one kilometer. This means a reduction in IFD by 0.8% at the average mean delivery rate of of 36%. Ownership of car of motorcycle and access to road have positive effect on IFD and the effects are statistically significant at the 1% level. Access to all-weather road increases the probability of IFD by 1.9 percentage points, while ownership of motorized transport increases the probability of IFD by 5.3 percentage points (column 4, Table III).

The point estimates in column 4 in Table III may seem small, however they are not. The mean distance to the nearest PHC, which is the lowest tier of health care in India, is close to nine kilometers (Table I). This implies that if a household lives nine kilometers from the nearest PHC, the IFD probability decreases by 2.7 percentage points (0.003*9=0.027). This translates into 7.5% decline in IFD (0.027/0.36) which is quite sizable. At the mean distance to the nearest health facility of 5km, this effect translates to 1.5 percentage points reduction in the probability of IFD. This result is consistent with the theoretical prediction of distance being an important constraint to seeking care and utilization of health services.

Heterogeneous effects Next, we examine if the effect of distance varies by the poverty status of the household. Poverty status of the household is measured by possession of the government-approved Below Poverty Line (BPL) card by the households. As stated above, financial constraints could limit seeking of delivery care as households may not have enough resources to spend on medical fees or transport. In this scenario, we should be able to see a bigger effect of distance on poor households compared to rich households. In Table

IV, we run model (1) separately on households that are below the poverty ine (BPL=1) and households that do noth belong to the BPL category (BPL=0). BPL is a government-recognized classification of poverty status and most government-sponsored social programs are run on the basis of the BPL classification.

Comparing the results in columns (1) and (2) in Table IV, it seems that distance does not have a differential effect on IFD depending on the poverty status of the household. Though the effects are statistically significant in each group, they are statistically not different across the poverty groups. The probability of IFD decreases by 0.3 percentage point for BPL women, while it is 0.2 percentage point for non-BPL women. The average IFD% among BPL women is 0.33 while it is 0.38 among non-BPL women. The difference in IFD across BPL and non-BPL women is 0.4%((0.003/0.33) - (0.002/0.38)) but the difference is not statistically significant (col 1 vs col 2).

Columns (3) and (4) explore the heterogeneous effect of distance by education of mother. Column (3) presents the result for women that have completed less than five years of schooling while column (4) shows the results for women with more than 5 years of schooling. We chose five years as the cut-off point because this is the primary level of schooling in India. The results suggest that effects are not statistically different across columns (3) & (4), though the baseline differential in the IFD rate across the groups is substantial. The IFD rate among primary schooled women is 58% while it is only 25% among women that are not primary-schooled.

[Table IV]

5.3 Instrumental variable results

The results presented so far show that distance is an important barrier for utilization of health care services. Women living farther away from health facilities have a lower probability of IFD. These results are consistent with the previous findings on the negative effects of distance on utilization of health services. However, for several reasons, these results may not be interpreted as the causal estimates of distance on IFD. For example, non-random placement of facility, unobserved heterogeneity, and measurement error may make the DIST variable endogenous, thereby rending the estimates biased. So we want to be sure that endogeneity in the key independent variable is not driving our results. To address this concern of endogeneity due to distance not being truly exogenous, we provide additional evidence on the causal effects in Table V by estimating an Instrumental Variable (IV) models. Specifically, we estimate Two-Stage Residual Inclusion (2SRI), IV-Probit, and IV-LPM. As stated above, we used DIST_INDEX as the instrument for the DIST variable.

First, we check if DIST is really an endogenous variable. To check this, we perform *Wu-Hausman F test* (Cameron and Trivedi, 2009). It should be noted that this is a test for the exogeneity of the regressors DIST and not for the exogeneity of the instrument. Based on this test, the null that DIST is exogenous is rejected implying that OLS is not consistent and IV approach would provide a consistent and efficient estimates of the parameter. The F-statistics is 638.63 with$P$-value < 0.0001.

Table V also reports the first-stage results. As stated above, the instrument is a summary index of distance to non-health infrastructures in the village. This index varies at the village level. Specifically, the index consists of distances to various non-health institutions of development. The first-stage results in Table V suggest that the instrument, DIST_INDEX, positively and significantly affects the availability of health facilities (DIST). This means that if the DIST_INDEX variable increases by one kilometer, the distance to nearest health facility increases by 1.34 kilometers. This satisfies the relevance condition of the IV, that is the instrument should be correlated with the endogenous variable. Note that the validity of the exclusion restriction cannot be tested since our model is exactly identified.

We also perform several weak-IV identification tests to make sure that the instrument does not suffer from weak-IV problem. Weak-IV test results are reported in column (1). The Cragg-Donald Wald F-Statistics is 10689.56 and Kleinbergen-Paap Wald rk F statistics is 238.34, rejecting the null hypothesis of weak instrument. These results suggest that the instrument is strongly related to the endogenous variable and the instrument does not suffer from weak-IV problem. The tests for weak-instrument-robust inference are also statistically different from zero, indicating that the estimated effects are robust to weak instrument problem if there is any.

Table V reports the 2SRI and the IV estimates from all the three models. The first column shows the results from LPM. As predicted, the simple LPM/Probit models have seriously underestimated the effect of distance on IFD. The 2SRI and IV estimates are about 5 times larger than the LPM/Probit estimates. For instance, according to LPM model, a one kilometer increase in DIST results in 1.7 percentage points reduction in the probability of IFD; the coefficient is statistically significant at the 1% level. Given that the mean IFD is 36%, this means a reduction of 4.7% at the sample mean (0.017/0.36). The coefficients on road access and ownership of car/motorcycle are positive, however the effect of connectivity to road loses statistical significance.

The second and third columns in Table V reports the IV-Probit and 2SRI results, respectively. The IV-Probit estimate in column (2) is negative and statistically significant. The results suggest that living farther from a health facility decreases the probability of IFD. The point estimate is 0.016, implying that an extra one kilometer of travel reduces the probability of IFD by 1.6 percentage points. Relative to mean of 0.36, the decline in IFD probability is 4.4%. Results are similar in our preferred specification of 2SRI model. The 2SRI results show that the probability of IFD is 1.6 percentage points lowers for every one kilometer of distance (column 3). Overall, the IV and 2SRI estimates suggest that distance is a significant burden on households for IFD. The magnitude of the effect in column 3 is large in the sense that households living 8.7 km away (mean distance to PHC) are 38% less likely to deliver in health facility (0.016*9/0.36).

As a robustness check, we re-constructed the DISTANCE_INDEX by including a few additional distance variables in the principal component model. For example, the new index additionally includes distance to telegraph service, distance to railway station, distance to phone, and distance to bank. We re-estimate the same specification with this new instrument. Results are essentially unchanged (not reported but available upon request). We came to similar conclusions, though the point estimate is slightly lower than the results presented in Table V.

[Table V]

5.4 Nonlinear effects of distance

In order to investigate the non-linearities in the effects of distance on IFD, we estimate models with distance as categorical variables. This specification would also be useful in capturing any threshold effect. Distance is included as dummies for households at different distances (less than 5 kilometers, 5 to 10 kilometers, and more than 10 kilometers, with less than 5 kilometers being the reference category). This specification ignores the concern for endogeneity since we do not have as many instruments as the number of endogenous variables.

Results are reported in Table VI. The first two columns report the results from the LPM estimation while columns (3) & (4) report marginal effects from probit model. Columns (1) and (3) consist of estimates without district fixed effects while columns (2) and (4) include district fixed effects. Results in column (1) shows that the probability of IFD decreases by 0.042 (4.2 percentage points) if the nearest health facility is between 5 and 9 kilometers, compared to the reference category of a facility within 5 kilometer. This translates to a 13% reduction in IFD (0.042/0.32). When the distance to the nearest health facility is greater than 9 kilometer, the probability of IFD reduces by 7.5 percentage points compared to the reference group, which translates to a reduction in IFD by 30%, relative to the mean IFD of 25% (column 1).

Coefficient estimates of distance are smaller in magnitude in columns (2) and (4), relative to results in columns (1) and (3). Including fixed-effects reduces the magnitude of the coefficients by more than half, suggesting that cross-district variations are also important in explaining the variation in IFD (column 2). With the district fixed-effect, the probability of IFD is 1.2 and 3.8 percentage points lower for the 5-9 kilometer and > 9 kilometer categories, respectively. We obtain extremely similar results using probit models. Coefficient estimate is 1.2 percentage points for 5-9 category, while it is 3.9 percentage points for > 9 kilometer category. Coefficient estimates for road access and motorized transport are positive and statistically significant. The magnitude of transport coefficient is 3.3 percentage points larger than the road access coefficient, indicating ownership of car or motorcycle is more important predictor than road connectivity. In summary, we obtain very similar results in LPM and probit models.

[Table VI]

To sum up, consistent with the theoretical prediction, we find that distance to the nearest health facility appears to impose a binding constraint on utilization of health services in rural parts of India.

5.5 Policy simulation

In this section, we discuss the results from the simulation exercise. We want to simulate the effect of increasing the density of facilities or decreasing the distance to the nearest health facility on IFD. The counterfactual question we would like to answer is "What would be the IFD rate if we all the sampled women live in 5 kilometers radius of the facility"? We attempt to answer this question by replacing greater than five kilometers value of DIST by five kilometers. So now the maximum value of DIST is 5 kilometer and the distance to closest facility lies between 0 to 5 kilometer.

The simulations are based on three steps. First, we estimate the baseline probabilities of IFD by averaging the individual predicted probability of IFD from the IV-LPM model. Second, simulated IFD probability for each women is predicted using the coefficients derived from the IV-LPM and altering the DIST variable and then the average IFD rate of the sample is calculated. Third, simulated policy effects are estimated by comparing the simulated probability values with the baseline probability values (simulated-baseline).

The results suggest that increasing the density of the health facilities in rural India, i.e. constructing more health facilities, would improve the IFD rate and the effects are sizeable. For instance, the predicted baseline probabilities of IFD is 36.27%. If we restrict the maximum DIST to be 5 km, the simulated probability based on IV-LPM coefficients is 39.72%, suggesting an increase of 3.45 percentage points and an approximate improvement of 10% over the current baseline probability.

6 Conclusions

This study shows that distance is a significant barriers to institutional delivery in India. We use both LPM and IV approaches to estimate the causal effects of distance to the closest health facility on delivery care. We find that distance to the closest health facility has a negative effect on the probability of institutional delivery. According to LPM, the probability of IFD decreases by 0.8% due to an increase in distance by one kilometer. However, the IV results suggest that LPM underestimates the true effects of distance on institutional delivery. Results from IV estimation show that the true causal impacts of distance on institutional delivery is 4.7%. We also find that access to road and transport infrastructure positively affects the delivery care.

To the best of our knowledge, this is one of a few studies that provides the causal impacts of distance on utilization of health services. Previous studies on the effect of distance on formal care in India are mainly from epidemiological literature and due to lack of appropriate econometric models, the causal interpretation can not be claimed in these studies. However, the significance of distance has been established in these studies as well. Many of these studies have demonstrated that access to health services within 5 kilometer of the village or higher density of health facility has statistically significant effect on institutional delivery, though the effect of access varies by state (Kumar et al. 1997; Stephenson & Tsui, 2002).

Previous evidence on the impacts of distance on health outcomes, such as child mortality has been mixed. For example, geographic access was not a significant barrier to infant and child mortality in Kenya (Moisi et al. 2010), or in Gambia (Rutherford et al. 2009). In contrast, many studies have found a strong association between child mortality and physical access to health care in Tanzania (Armstrong, Mrisho & Manzi, 2008), the Democratic Republic of Congo (Broeck, Eeckels & Massa, 1996) and Burkina Faso (Becher et al. 2004). The mixed evidence suggests that the effect of distance could depend on the country context, the density of health facilities and the outcome of interest.

Limitations: Although the reported measure of distance in this paper is an improvement over other studies, it still is an imperfect proxy for travel time and cost. Additionally, the distances are estimates based from the village center, while women may reside some distance from that location. Our controls for all-weather road access and vehicle ownership address this issue somewhat, but ideally these estimates should be household specific, and include information about the topography of the area and local travel costs. Our study also assumes that women deliver at the closest facility, or at least that this facility helped them access care through a referral or otherwise. Furthermore, future work should continue to address the possibility of a quality-distance trade-off by incorporating the quality of maternity care available at each facility. Due to unavailability of data, this study did not include any supply-side variable that may explain the variation in the quality of care.

Our study has important policy implications. Our study findings suggest that in countries like India, where geographic distance to health facilities is quite large in rural areas, geographic access to health care is a significant barriers to institutional delivery. The significant effect of distance suggest that increasing the density of health facilities and providers in rural areas may improve maternal and neo-natal, however, a comprehensive cost-effective analysis should be performed to demonstrate that benefits of institutional delivery outweigh the cost of building new facilities.[7] Another important finding is that reducing transport

[7]Conducting a cost-effectiveness analysis is beyond the scope of this paper.

barriers would also help reduce the inequity in geographic access to health facility and would likely improve institutional delivery coverage. Our findings will assist health policy makers in Indian and other resource-constrained countries to understand the likely impact of health infrastructure on improving health outcomes.

However, merely increasing the number of health facilities may not be sufficient to promote health service utilization as poor quality of care and high cost may inhibit the uptake of the services. The wealth gradient found in this study also emphasizes the importance of financial resources in household's choice of in-facility delivery. Therefore, in addition to increasing access, further improvements in institutional delivery can be met by improving quality-of-care and easing financial constraints. The Government of India has implemented a conditional cash transfer scheme, Janani Suraksha Yojana (JSY), to promote institutional delivery. This scheme has been successful in improving the in-facility delivery rate (Lim et al. 2010) to some extent. However, to meet the target of fifth MDG by 2015 and provide universal coverage, the Government of India would have to undertake a series of supply and demand-side interventions, ranging from increasing access to breaking cultural beliefs.

ACKNOWLEDGEMENTS

The authors wish to thank Anirban Basu, Emmanuela Gakidow, Mike Hanlon, Mehtabul Azam, and Edward Hoang for their helpful comments. The paper also has benefited from comments received at the IHME research seminar. The research was supported by funding from the Bill and Melinda Gates Foundation. We thank Kelsey Moore for project management. We are responsible for any errors that may remain. The authors bear sole responsibility for the content of this paper.

References

Angrist, Joshua D., and Alan B. Krueger. 2001. Instrumental variables and the search for identification: from supply and demand to natural experiments." *Journal of Economic Perspectives* **15**(4): 69Ű85.

Angrist JD and Pischke J-S. 2008. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

Armstrong Schellenberg JR, Mrisho M, Manzi F. 2008. Health and survival of young children in southern Tanzania. *BMC Public Health* **8**:194

Becher H, Muller O, Jahn A, Gbangou A, Kynast-Wolf G, Kouyate B. 2004. Risk factors of infant and child mortality in rural Burkina Faso. *Bull World Health Organ* **82**: 265-273.

Bhattacharya J, Goldman D, McCaffrey D. 2006. Estimating probit models with self-selected treatments. *Statistics in Medicine* 25: 389Ű413.

Bleakley H. 2010. Health, Human Capital, and Development, *Annual Reviews of Economics*, 2:283–310

Borah B. 2006. A mixed logit model of health care provider choice: analysis of NSS data for rural India. *Health Economics* **15**: 915Ű32..

Broeck J, Eeckels R, Massa G. 1996. Maternal determinants of child survival in a rural African community. *Int J Epidemiol* **25**: 998-1004.

Cameron, Colin A. and Pravin K. Trivedi. 2009. Microeconometrics Using Stata, Texas: Stata Press.

Collier, Paul, Dercon, Stefan, and John Mckinnon. 2002. Density versus quality in health care provision: using household data to make budgetary choices in ethiopia. *World Bank Econ Rev* **16**(3): 425-448.

Elo IT. 1992. Utilisation of maternal health-care services in Peru: the role of women's education. *Health Transit Rev* **2**(1): 49-69.

Gething, Peter W., Amoako Johnson, Fiifi et al. 2012. Geographical access to care at birth in Ghana: a barrier to safe motherhood. Southampton, GB, ESRC Centre for Population Change, 34pp. (Centre for Population Change Working Paper, 21).

Hellevik, O. 2009. Linear versus logistic regression when the dependent variable is a dichotomy. *Quality and Quantity* **43**:59Ű74

Hjortsberg, C. 2003. Why do the sick not utilise health care? The case of Zambia. *Health Economics* 12: 755Ű70.

Hogan MC, Foreman KJ, Naghavi M, et al. 2010. Maternal mortality for 181 countries, 1980-2008: a systematic analysis of progress towards Millennium Development Goal 5. *The Lancet* **375**: 1609-23.

International Institute for Population Sciences (IIPS), 2010. District Level Household and Facility Survey (DLHS-3), 2007-08, IIPS, Mumbai, India.

Kesterton AJ, Cleland J, Sloggett A & and Ronsmans C. 2010. Institutional delivery in rural India: the relative importance of accessibility and economic status. *BMC Pregnancy and Childbirth* **10**:30

Kumar RM, Singh M, Kaur M. 1997. Impact of health centre availability on utilisation of maternity care and pregnancy outcome in a rural area of Haryana. *J Indian Med Assoc* **95**(8):448-50.

Lavy, Victor. 1996. School supply constraints and children's educational outcomes in rural ghana. *Journal of Development Economics* **51**(2): 291Ű314.

Lim SS, Dandona L, Lalit, Hoisington JA, James SL, Hogan MC, Gakidou E. 2010. India's Janani Suraksha Yojana, a conditional cash transfer programme to increase births in health facilities: an impact evaluation. *The Lancet* Vol. 375, Issue 9730, Pages 2009-2023.

Matthews Z, Ramakrishna J et al. 2005. Birth-rights and rituals in rural south India: Care-seeking in the intra-partum period. *J Biosoc Sci* **37**(4): 385-411.

Moisi JC, Hellen Gatakaa, Abdisalan M Noor, Thomas N Williams. 2010. Geographic access to care is not a determinant of child mortality in a rural Kenyan setting with high health facility density. *BMC Public Health* **10**: 142.

Mukhopadhyay Abhiroop and Soham Sahoo. 2012. Does access to secondary education affect primary schooling? evidence from India. IZA DP No. 6507.

Odiit A, Amuge B. 2003. COMPARISON OF VACCINATION STATUS OF CHILDREN BORN IN HEALTH UNITS AND THOSE BORN AT HOME, *East African Medical Journal*, 80(1).

Ronsmans C, Graham WJ. 2006. On behalf of the lancet maternal survival series steering group: maternal mortality: who, when, where and why. *The Lancet* **368**: 1189-1200.

Rutherford ME, Dockerty JD, Jasseh M, Howie SR, Herbison P, Jeffries DJ, Leach M, Stevens W , Mulholland K, Adegbola RA, Hill PC. 2009. Access to health care and mortality of children under 5 years of age in the Gambia: a case-control study. *Bull World Health Organ* **87**: 216-224.

Sarma Sisira. 2009. Demand for outpatient healthcare: empirical findings from rural India. *Applied Health Economics and Health Policy*, Springer, **7**(4): 265-277.

Sawhney N. 1993. Management of family welfare programmes in Uttar Pradesh: infrastructure utilisation, quality of services, supervision and MIS. In Family Planning and MCH in Uttar Pradesh (A Review of Studies). Edited by Premi M. New Delhi: Indian Association for the Study of Population; 50-67.

Stephenson R, Tsui A. 2002. Contextual influences on reproductive wellness in Northern India. *Am J Public Health* **93**(11):1820-1829

Terza J, Basu A, Rathouz P. 2008. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics* 27(3): 531Ű543.

Thaddeus S, Maine D. 1994. Too far to walk: maternal mortality in context. *Soc Sci Med* **38**(8):1089-1110.

World Health Organization, UNFPA and the World Bank. 2010. Trends in maternal mortality: 1990 to 2008 estimates developed by WHO, UNICEF, UNFPA and the World Bank, Geneva: World Health Organization.

TABLE 1
Descriptive statistics, N=158,897

	Mean	SD
	(1)	(2)
In-facility delivery	0.36	0.48
In-facility delivery (0-4 km)	0.41	0.49
In-facility delivery (5-9 km)	0.32	0.47
In-facility delivery (> 9 km)	0.25	0.43
Distance to nearest PHC (km)	8.70	7.97
Distance to nearest block PHC (km)	11.70	10.63
Distance to nearest CHC(km)	17.64	15.46
Distance to nearest DH (km)	34.07	23.83
Distance to nearest private clinic (km)	10.06	12.49
Distance to nearest private hospital (km)	18.67	17.49
Distance to nearest health facility (km)	4.94	6.08
Women living in 0-4 km	0.58	0.49
Women living in 5-9 km	0.24	0.43
Women living in > 9 km	0.18	0.38
Scheduled Caste/Tribe (SC)	0.40	0.49
Other Backward Caste (OBC)	0.41	0.49
Below Poverty Line (BPL)	0.34	0.47
Hindu	0.79	0.41
Muslim	0.12	0.32
Mother's age at birth	25.08	5.36
Mother's education (primary=1)	0.34	0.47
Mother's education (secondary=1)	0.7	0.26
Wealth quintile (1, poorest)	0.25	0.43
Wealth quintile (2, poor)	0.25	0.43
Wealth quintile (3, middle)	0.22	0.42
Wealth quintile (4, rich)	0.19	0.39
Wealth quintile (5, richest)	0.10	0.30
JSY payment	0.10	0.31
All weather road in the village	0.85	0.36
Owns car/motorcycle	0.02	0.12

Notes: Means are reported for the continuous variables, and proportions are reported for categorical variables with the standard deviations. PHC refers to primary health centers, CHC refers to community health centers, and DH refers to district hospitals. Primary schooling means completion of more than five years of school while secondary schooling means completion of more than 10 years of school. Distance to nearest health facility is the shortest distance to one of six types of facilities.

TABLE 2

Effect of distance on institutional delivery in rural India

	In-facility births			
	LPM	LPM	Probit M.E.	Probit M.E.
	(1)	(2)	(3)	(4)
DIST (km)	-0.005***	-0.002***	-0.005***	-0.003***
	(0.0004)	(0.003)	(0.0004)	(0.0003)
Access to road	0.035***	0.019***	0.038***	0.019***
	(0.006)	(0.004)	(0.007)	(0.004)
Own motorized vehicle	0.068***	0.057***	0.066***	0.053***
	(0.011)	(0.010)	(0.011)	(0.010)
District fixed-effect	No	Yes	No	Yes
r2	0.27	0.36	0.22	0.32
N	158897	158897	158897	158897

Notes: Robust standard errors, clustered by district, are presented in parentheses. Columns (1) and (2) report coefficients from Linear Probability Model (LPM), while columns (3) and (4) report marginal effects (M.E.) from Probit model. All regressions adjust for caste, religion, mother's age at birth, mother's education, wealth quintile, JSY receipt, and year of birth dummies. Distance is continuous and is measured in kilometers. All results are adjusted for population weighting and survey design. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$,

TABLE 3
Probit estimates of institutional delivery - by poverty status and mother's education

	In-facility births			
	Below poverty line		Mother's education	
	No	Yes	< 5	> 5
	(1)	(2)	(3)	(4)
DIST (km)	-0.002***	-0.003***	-0.002***	-0.003***
	(0.0003)	(0.0005)	(0.0003)	(0.0005)
Access to road	0.021***	0.018***	0.022***	0.010
	(0.005)	(0.006)	(0.004)	(0.007)
Own motorized transport	0.058***	0.007***	0.021	0.073***
	0.011	(0.029)	(0.018)	(0.013)
District fixed-effect	Yes	Yes	Yes	Yes
N	104608	54289	104948	52288

Notes: Robust standard errors, clustered by district, are presented in parentheses. All columns report marginal effects (M.E.) from Probit model. All regressions adjust for caste, religion, mother's age at birth, mother's education, wealth quintile, JSY receipt, and year of birth dummies. Distance is continuous and is measured in kilometers. BPL is below poverty line. All results are adjusted for population weighting and survey design.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$,

TABLE 4
Marginal effects on institutional delivery from 2SRI, IV Probit and IV Linear Probability Model (LPM)

	In-facility births		
	IV-LPM	IV-Probit	2SRI
	(1)	(2)	(3)
First-stage	1.34***	1.34***	1.34***
	(0.087)	(0.87)	(0.087)
Distance to nearest health facility (km)	-0.016***	-0.016***	-0.017***
	(0.002)	(0.001)	(0.001)
Road access	0.001	0.004	0.005
	(0.002)	(0.005)	(0.002)
Own motorized vehicle	0.057***	0.054***	0.055***
	(0.010)	(0.009)	(0.010)
Weak IV identification test			
Cragg-Donald Wald F statistic	10689.56		
Kleibergen-Paap Wald rk F statistic	238.34		
Weak-instrument-robust inference			
Anderson-Rubin Wald test, F(1,588)	175.72(p<0.000)		
Stock-Wright LM S statistic, Chi-sq(1)	129.43(p<0.000)		
District fixed-effect	Yes	Yes	Yes
N	158897	158897	158897

Notes: Robust standard errors, clustered by district, are presented in parentheses. Columns (1) & (2) report predicted probabilities. In the IV-Probit and IV-LPM, distance to closest health facility was instrumented by PCA index of distances to non-health infrastructures (defined in text).
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$,

TABLE 5

Effect of distance dummies on institutional delivery in rural India

	In-facility births			
	LPM	LPM	Probit (M.E.)	Probit (M.E.)
	(1)	(2)	(3)	(4)
Distance (5-10 km)	-0.042***	-0.012***	-0.041***	-0.012***
	(0.006)	(0.004)	(0.006)	(0.003)
Distance (>10 km)	-0.075***	-0.038***	-0.078***	-0.039***
	(0.007)	(0.004)	(0.007)	(0.004)
Road access	0.037***	0.020***	0.039***	0.020***
	(0.006)	(0.004)	(0.007)	(0.004)
Own motorized vehicle	0.069***	0.057***	0.066***	0.053***
	(0.011)	(0.010)	(0.011)	(0.010)
District fixed-effect	No	Yes	No	Yes
N	165949	158897	158897	158897

Notes: Robust standard errors, clustered by district, are presented in parentheses. Columns (1) & (2) show results from Linear Probability Model (LPM) while columns (3) & (4) reports marginal effects from Probit. Distance is a categorical variable.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$,