Prenatal Care and Infant Health: Elementary, Dear Watson? Accounting for Selection Bias in Nonexperimental Data

Satvika Chalasani *

21 September 2007

This paper reexamines the relationship between prenatal care and infant health using cross-sectional data from India. Given the dearth of experimental evidence on the nature and extent of the effects of prenatal care on birthweight and infant mortality, we have mostly drawn conclusions from observational data. These results often suffer from the nagging problem of selection bias: women that seek a certain level of prenatal care may be systematically different in a way that also affects their birth outcomes. I employ (i)a sibling-difference model, and (ii)a propensity score model to address this unobserved heterogeneity. Neither of these approaches has been used before in the literature on prenatal care. The use of data from a non-Western setting is also a first for the prenatal care-selection literature. The huge investments made in prenatal care worldwide as part of the effort to improve maternal and child health make this a question worth revisiting and answering right.

* Doctoral Candidate, Dual Degree Program in Sociology and Demography, The Pennsylvania State University, University Park, Pennsylvania 16802. satvika@pop.psu.edu

The author gratefully acknowledges support from the interdisciplinary Quantitative Social Science Initiative at Penn State.

Introduction

In this paper, I take a fresh look at the relationship between the use of prenatal care and subsequent infant survival using pooled cross-sectional data from the National Family Health Survey of India, 1992-93 and 1998-99[1]. Although there exists widespread support for the provision and usage of prenatal care, surprisingly to many, the actual evidence regarding the extent of its beneficial effects is less than conclusive. Most of the positive effects of prenatal care on birthweight documented in the medical literature are confined to full-term infants. When it comes to preterm infants however, there is very little evidence that prenatal care can (i)aid in the early identification and delaying of such births, or (ii) improve their birthweight (Alexander and Korenbrot 1995). Due to ethical considerations, there have been no randomized evaluations involving the comparison of women who received prenatal care versus those who did not. Instead there are a handful of studies that compare outcomes for women randomly assigned to select kinds of prenatal care – individual versus group care (Ickovics, Kershaw, Westdahl, Magriples, Massey, Reynolds, and Rising 2007), comprehensive medical, nutritional, educational care versus basic care, nurse midwifery, nurse home visitation, provision of psychosocial support (Alexander and Korenbrot 1995), etc. These studies have at best shown improvements in birthweight for very specific subpopulations.

Nonrandomized, observational studies while more common, almost always suffer from potentially severe self-selection bias. Simple multivariate analysis with a set of controls is inadequate for addressing the bias in the utilization of prenatal care that has long been a concern of some researchers investigating this topic. There are two sources of potential bias stemming from either favorable or adverse selection (Joyce 1994). Favorable selection implies that women who would have anyhow (even without the prenatal care) had relatively healthier babies are also more likely to seek prenatal care. A good example is the more intensive utilization of reproductive health care by women of higher socioeconomic status (SES) in developing country settings (Koenig and Shepherd 2001).

[1] I anticipate that the latest round of data will be released before the end of 2007. If so this will add substantially to the current sample given that the number of women interviewed in 2005-06 is larger than either of the first two surveys.

This would result in an overestimation of the positive effects of prenatal care if SES were not well-accounted for in the statistical analyses. Adverse selection implies that women who are more likely to have complications at birth (and anticipate them) also seek prenatal care more aggressively. This would result in an underestimation of the benefits that prenatal care truly has because this group has a particularly problematic set of infant outcomes. This adverse selection could be based on some underlying health attributes not just on income and education which are more easily observable by the outsider collecting the data. The mutual association of prenatal care usage and infant health outcomes with some unmeasured, potential true cause leads us to ask the central question of this paper: *how does one model this unobserved heterogeneity?* There are just a few studies that attempt to address this problem but they are not without limitations. The next section discusses their approaches.

Estimation Strategies in the Literature

In general, there have been three approaches to the issue of unobserved heterogeneity in the usage of prenatal care.

1. *Instrumental variable estimates*. Rosenzweig and Schultz (1983) were the first to employ a two-stage least squares approach to estimating the effects of prenatal care on birthweight - the 2SLS is essentially a special case of an instrumental variable model. The underlying theoretical motivation was the conceptual distinction that they made between the demand for health inputs, and the health production function itself. The health production function is estimated in the second stage of their regression model, where essentially y_i=birthweight and x_{ij}=prenatal care, mother's smoking behavior, birth order of the child, and mother's age at birth (and some other exogenous controls). But before that, in the first stage of their regression model they estimate separate demand equations for the four health inputs, where each is treated as endogenous. The right-hand side variables are: the price of inputs, education, and income. These three function as instruments because they are assumed to not affect birthweight except through the four health input behaviors.

Their results are revealing but definitely merit further investigation. Comparing OLS estimates to 2SLS, a significantly negative effect of delay in prenatal care on birthweight emerges with the latter specification, suggesting severe adverse selection in the usage of prenatal care. However, when using birthweight standardized for gestation length as a dependent variable, the beneficial effects of prenatal care disappear regardless of whether heterogeneity is accounted for. By the authors' own admission, their estimates could be sensitive to the omission of relevant behavioral variables. Secondly, they also concede that the instruments for prices of inputs may not be orthogonal to the health endowment itself. Additionally, I contend that the assumption that income operates only though their limited set of behavioral inputs is a very strong one. It is a fact that the 2SLS approach places a high premium on effective instruments and plausible identification restrictions (Joyce 1994).

Corman, Joyce and Grossman (1987) corroborate these results for black women but their study (although they include a different set of covariates in both equations) suffers from some of the same problems as the Rosenzweig and Schultz model. Warner (1998) using the same 2SLS approach concludes "Multicollinearity in the complex specifications and weak identification of the birthweight equation prevent strong conclusions as to whether the marginal returns to prenatal care are constant, decreasing, or increasing."

2. *Switching regression model.* Joyce (1994) uses an endogenous switching regression model that is also a two-stage estimation (but distinct from 2SLS). First, parameter estimates are obtained from an equation with prenatal care as the dependent variable. Then these estimates are used to construct correction factors (inverse mills ratios) that are included as independent variables in a regression with birthweight as the dependent variable. Selection bias is present if we reject the null that the coefficient on the correction factor is different from zero. Similar to the above studies, he finds that estimates of the effects of prenatal care on birthweight

understate the effects of prenatal care if the (adverse) selection is not explicitly modeled. But Joyce treats only prenatal care as an endogenous input. It has been argued by the previously-mentioned authors that there are several other factors in the birthweight equation that need to be treated as endogenous. Also, his rejection of the overidentification restriction among blacks suggests that the correction factors do not fully account for the adverse selection. Again, two-stage estimation procedures produce estimates inferior to OLS when the first stage equations have low explanatory power.

3. *Aggregate-level fixed effects*. Frank *et al* (1992) use time series data on counties to estimate a fixed effects model that controls for unobserved differences in the health endowments of women across counties. What this does is to remove the effects of any underlying, unchanging variables that simultaneously affect birthweight as well as the initiation of prenatal care, variables on which geographically clustered women may share the same values. The authors' final estimates of the impact of prenatal care on birthweight are significantly lower (although still positive) than those found in the above studies accounting for selection. They conclude that the expansion of early initiation of prenatal care will have only a small impact on reducing the risk of low birthweight. The big problem with this study is the classic ecological fallacy. One cannot make individual-level conclusions from aggregate-level data.. For example, suppose that there was some income growth over the 9-year study period, and that this growth disproportionately benefited higher SES individuals. So the distribution is not captured in an average income measure like the one Frank *et al* use. If the women who would have had fairly healthy babies anyway could now afford to seek more preventive prenatal care, then this presents a case of favorable not adverse selection as their study concludes.

As discussed above, all the current strategies are limited in one way or the other. I propose to improve upon them along certain dimensions (details in the next section).

Furthermore, none of the above studies uses data from poor countries where a different sort of selection may be in operation.

Methods

This paper employs a ***sibling-difference model*** to address the unobserved heterogeneity that yields biased and inconsistent estimates of the coefficient on prenatal care, and hampers our ability to make causal inferences regarding its effects on infant health. Household fixed effects models leverage the difference in outcomes within sibling pairs to remove the confounding effects of any family-specific endowments, whether genetic or environmental. Let j (j = 1, …n) denote the jth household. Let i (i = 0, 1) denote the ith sibling within the household where i = 0 represents the child that did not receive prenatal care, and i=1 represents the child that did. Also included in the regression equation are measured covariates (x_{ij}) that are potentially related to the outcome and our treatment variable (see the section on data and measures for a discussion). The additional effect of receiving prenatal care is captured by the parameter τ. Household-level variables both observed and unobserved are captured by α_j .This gives us the sibling-specific equations:

$$y_{0j} = \beta' x_{0j} + \alpha_j + \varepsilon_{0j} \qquad (1)$$

$$y_{1j} = \beta' x_{1j} + \tau + \alpha_j + \varepsilon_{1j} \qquad (2)$$

When we difference the two equations, the household-level factors that are stable across siblings drop out and yield the following:

$$y_{1j} - y_{0j} = \beta'(x_{1j} - x_{0j}) + \tau + (\varepsilon_{1j} - \varepsilon_{0j}) \qquad (3)$$

Equation (3) tells us that the difference in birthweight or mortality experience between siblings is a result of their differential prenatal care, conditional on any other differences between siblings and all possible household-level unobserved endowments.

This sibling-difference approach is not free of assumptions but I argue that this approach is still better than ignoring the heterogeneity problem. The first assumption is that the above-mentioned family-specific endowments are constant across siblings. It is possible to think of situations where this is not true. For one, there may be a progressive

weakening of the mother's body as she has more children such that siblings have different rates of fetal growth. But including the birth order variable in the regression model should indirectly capture any such differences. Also, there may be changes in the family's economic situation over time. However, given that this sample consists of first and second births, 90% of which are less than 4 years apart in age, the probability that a household's income has changed drastically over such a short period is rather low. Sibling-difference models generally also assume that there are no idiosyncratic child-level endowments that affect both the treatment and outcome variables. This assumption does not pose a problem in this study because the treatment here occurs before the birth of the child itself which makes it unlikely that it affects the prenatal care of the child when it was in the mother's womb. It is possible that parents may know the gender of the fetus and consequently change their behavior but again, the control for child's gender should address this unlikely event. In sum, I contend that the assumptions regarding family endowments and child-specific endowments are certainly weaker than the assumption that those unobserved factors are uncorrelated with the seeking of prenatal care, and that they exert no influence on birth outcomes.

While the sibling-difference model is the main strategy for overcoming the selection problem, I propose using a second strategy, ***propensity score matching***, to bolster results from the first. Matching estimators are receiving renewed attention in the sociological literature. The fundamental logic behind matching is simple. We want to know the average treatment effect for the treated. But in order to arrive at that, we have to know (i) their outcome following treatment, AND (ii) the counterfactual: what would have been their outcome had they not been treated. Using the mean outcome of the untreated as a proxy for (ii) is incorrect because the treated and untreated were probably different to begin with, different in ways that are potentially correlated with the outcome. Randomized assignment eliminates that selection problem but in nonexperimental data, we have to find some way of coming up with a treatment and comparison group that were similar in their pretreatment profiles. Matching individuals on a whole vector of observed covariates leads to a dimensionality problem. According to Rosenbaum and Rubin (1983) matching treated and untreated based on a propensity score is sufficient to remove bias

due to observed covariates. The propensity score is then the conditional probability of assignment to a particular treatment given a vector of observed covariates.

In the first stage, I generate a propensity score by estimating a probit

$$\Pr\{D_i = 1|X_i\} = \Phi(h(X_i))$$

where D_i indicates whether or not a woman received prenatal care (or adequate versus inadequate care), Φ is the normal cumulative distribution function, and $h(X_i)$ is our set of observed covariates. I then employ the nearest neighbor matching algorithm to come up with a comparable group of controls for the treated subjects such that the balancing hypothesis is satisfied (Rosenbaum and Rubin 1983). An additional criterion that needs to be met is that of the common support: individuals with the same values on observed covariates should have a positive probability of being both treated and untreated. Once these two requirements are satisfied, we are ready to approximate the above-mentioned counterfactual. The average treatment effect for the treated is then simply the mean difference in outcomes between the treated and the matched controls over the region of common support, weighted by the propensity score distribution of participants (Caliendo and Kopeinig 2005).

The propensity score approach is not without limitations. The biggest issue is the conditional independence assumption, which implies that selection is based on observed traits of individuals, and that there are no unobserved variables simultaneously influencing treatment and outcome. Given this, the big question is *why would we even use this method if the whole point is to deal with unobserved heterogeneity*? Well the answer to that question is twofold. Firstly, I argue that when observations are matched on a relatively rich but incomplete set of observed covariates, it raises the probability that they are matched on unobserved traits as well. For this not to be true, the observed and unobserved endowments should be working in opposite directions, and it is hard to present good examples of that.

Secondly, while there exist formal ways of testing for heterogeneity in the fixed effects model using the Hausman test, the use of propensity scores allows us to determine the extent of the hidden bias using the concept of Rosenbaum bounds (DiPrete and Gangl 2004). If the conditional independence assumption is met, the odds of undergoing treatment should be the same for two individuals that have the same values on an observed covariate. This must mean that any difference between them in the odds of participation is given by

$$\exp[\gamma(v_i - v_j)]$$

where v_i and v_j are unobserved covariates, and γ is the coefficient on the difference between those values for individuals *i* and *j*. To determine the sensitivity of the treatment effect to a change in γ or in $(v_i - v_j)$, we examine the bounds on the odds ratio for participation that lie between $1/e^{\gamma}$ and e^{γ}. e^{γ} is then a measure of how far our model is from one that is free of hidden bias (Caliendo and Kopeinig 2005).

In sum, I will be presenting results from these *two approaches*: fixed effects and propensity scores, for *three specifications of the treatment*: receipt, timing, and adequacy of prenatal care, and for *three separate outcomes*: birthweight, size of infant at the time of birth, and neonatal mortality (see discussion in next section). Ordinary least squares regression will be used for birthweight, ordered logit for infant size, and the discrete-time logistic hazard model for neonatal mortality.

Data and Measures

Data for this analysis come from the Demographic and Health Survey series, known as the National Family Health Survey in India. The NFHS is the largest survey in India specifically targeted at measuring population health. Within each state, a two-stage stratified random sampling design is used in rural areas: first villages, then households. In urban areas, a three-stage design is employed: cities/towns, followed by urban blocks, and then households. Three survey instruments were used, a village questionnaire, a household questionnaire and a women's questionnaire(International Institute for

Population Sciences 1995). Relevant to this study are the data tabulated from responses to the women's and household questionnaire. One of the fundamental aims of these surveys was to obtain reliable state-level estimates of the parameters of interest (and then separately for rural and urban areas within states), so target sample sizes were determined accordingly. This meant that ultimately the national sample size was unusually large. In 1992-93,

> Interviews were conducted with a nationally representative sample of 88,562 households and 89,777 ever-married women in the age group 13-49, from 24 states and the then National Capital Territory of Delhi (http://www.nfhsindia.org/nfhs1.html).

In 1998-99,

> The survey covers a representative sample of about 91,000 ever-married women age 15-49 from 26 states in India who were covered in two phases, the first starting in November 1998 and the second in March 1999 (http://www.nfhsindia.org/nfhs2.html).

As mentioned earlier, I propose using three different outcome variables: birthweight, size of infant at the time of birth, and infant mortality. A brief discussion of each follows.

1. Birthweight in India is not widely recorded. In fact, 70% of births to women in the three years preceding the survey had not been weighed (International Institute for Population Sciences and ORC Macro 2000). The women that do not report on birthweight are more likely to be poor, from rural areas, and to have had a home delivery. This already implies some sort of favorable selection in the sample because mothers who report birthweight are more likely to have access to delivery facilities where prenatal care could also have been sought. However, remember that the modeling strategy is explicitly aimed at minimizing such bias. There is also the issue of recall error in birthweight even though the survey only asks about the woman's births in the past three years, usually her two most recent children. However, it has been argued that when using first-differencing, measurement error in the dependent variable would lead to imprecise yet unbiased estimates of the coefficients of interest (Hamermesh 1989). Still, birthweight in this survey seems to be plagued by some pretty serious problems.

2. Mothers were also asked about the size of each baby at birth (large, average, small, or very small). There is a fair degree of consonance between reported size and birthweight. Again, this is a somewhat crude measure but it gives us one way of triangulating our findings from models using the other outcomes.

3. Arguably the most reliable measure of infant health in the data is the mortality measure. The NFHS mortality data have been subject to rigorous quality assessment. In both surveys, it was found that early infant deaths had not been severely underreported. As for misreporting of age at death (particularly heaping at certain ages), it was found that the infant mortality rate was unlikely to be underestimated by more than 1-2% ((International Institute for Population Sciences 1995; International Institute for Population Sciences and ORC Macro 2000). I will be analyzing the effects of prenatal care on the odds of neonatal death (<28 days). While I could also analyze postneonatal and early childhood death (certainly it is increasingly common for researchers in various fields to take a lifecourse approach and talk about the long-term impacts of fetal growth and perinatal health), there are a whole host of other factors that begin to come into play after the first few days of an infant's existence. And these factors interfere with our aim of understanding of the direct effects of prenatal care.

Fixed effects models are very demanding in terms of sample size, and this is one of the big motivations to use the data from India. Only those sibling pairs are retained where the value of the dependent variable differs for siblings (if the difference between their outcomes is zero they do not contribute to the likelihood function). So while the sample for birthweight and infant size of approximately 1500 births is the smallest one in our set of analyses, it is considerably larger than the usual samples for fixed effects analyses. For neonatal mortality, the sample is certainly much larger (yet to be determined but it will be in the several thousands) because there is virtually no missingness in the mortality data.

Propensity score models are somewhat different in their data requirements. In theory, one wants a large number of controls relative to the number treated so that high score participants can be matched to high score nonparticipants instead of being matched to low score nonparticipants. This isn't quite the case in these data with about 2/3^{rds} of the women receiving care while only a 1/3rd did not. The solution to this is to use matching with replacement which increases the variance of the estimator but also has the desirable property of improving the average quality of matching and decreasing bias(Smith and Todd 2005). Furthermore, the sheer size of the dataset will help to minimize this problem: the odds of finding an exact match increase asymptotically, and this is by no means a small sample.

Exogenous variables in both the fixed effects and propensity scores approach include maternal age, birth order, gender of child, urban or rural residence, household's position in the wealth distribution (measured by a complex index based on household's ownership of key consumer durables and housing quality), mother's education, father's education, media exposure, religion, and caste. Additionally, when neonatal mortality is the outcome, I include nutrition, immunization, and illness histories of the child as covariates.

Receipt of prenatal care is first modeled as a simple 1/0 indicator considering that almost a third of the women in the sample did not receive any prenatal care at all. In the next set of analyses, prenatal care is modeled by trimester of initiation, a widely used measure in the literature (Kotelchuck 1994). In the last set of analyses, I use the Kessner/Institute of Medicine adequacy of prenatal care index which is a combination of both the timing of initiation, and the total number of visits.

Relevance

Given the mammoth investments that governments worldwide make in prenatal care, the importance of knowing the true returns to this investment cannot be overstated. While

the results that will be presented in this paper are by no means the final word on this issue, they would serve to bolster the case of *one of two camps*. If it turns out that we have been underestimating its benefits, then it is all the more reason for public outlays to favor prenatal care and promote it as a *preventive public health intervention*. If on the other hand, the benefits of prenatal care have been overstated in the past, then perhaps governments should invest more in *curative medical technologies*. Indeed this has been the story in the U.S. for the past few decades, with most of the decline in neonatal mortality coming not from improvements in birthweight but in birthweight-specific mortality (Alexander, Tompkins, Allen, and Hulsey 1999). This points more towards the importance of high-risk obstetric and neonatal care rather than prenatal care. However, medical technologies are clearly very expensive, and it is possible that even small returns on investments in prenatal care are much more cost-effective. Moreover, in resource-poor settings (which are almost always at a different stage in the epidemiological transition), the etiologies of low birthweight and infant mortality may differ substantially from those in richer parts of the world. So what is true for the U.S. may certainly not be true for India.

Either way, more accurate estimates of the benefits of prenatal care are of immediate relevance to all societies but especially to low-income societies. The use of data from a developing country like India marks a first for the literature on this topic. Notwithstanding the limitations of this study, I submit that the usage of two new estimation strategies in addressing a major substantive issue makes this an important contribution.

References

Alexander, G. R., M. E. Tompkins, M. C. Allen, and T. C. Hulsey. 1999. "Trends and Racial Differences in Birth Weight and Related Survival." *Maternal and Child Health Journal* 3:71-9.

Alexander, Greg R. and Carol C. Korenbrot. 1995. "The Role of Prenatal Care in Preventing Low Birth Weight." *The Future of Children* 5:103-120.

Caliendo, Marco and Sabine Kopeinig. 2005. "Some Practical Guidance for the Implementation of Propensity Score Matching." in *IZA Discussion Papers*. Germany.

DiPrete, Thomas and Markus Gangl. 2004. "Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments " *Sociological Methodology* 34:271-310.

Hamermesh, Daniel S. 1989. "Why Do Individual-Effects Models Perform So Poorly? The Case of Academic Salaries." *Southern Economic Journal* 56:39-45.

Ickovics, J. R., T. S. Kershaw, C. Westdahl, U. Magriples, Z. Massey, H. Reynolds, and S. S. Rising. 2007. "Group Prenatal Care and Perinatal Outcomes: A Randomized Controlled Trial." *Obstet Gynecol* 110:330-9.

International Institute for Population Sciences. 1995. "National Family Health Survey (Maternal and Child Health and Family Planning), India 1992-93." Bombay, India: IIPS.

International Institute for Population Sciences and ORC Macro. 2000. "National Family Health Survey (Nfhs-2), 1998-99." Mumbai, India: IIPS.

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

Koenig, Michael and Mary Shepherd. 2001. "Alternative Study Designs for Research on Women's Gynaecological Morbidity in Developing Countries." *Reproductive Health Matters* 9:165-175.

Kotelchuck, M. 1994. "An Evaluation of the Kessner Adequacy of Prenatal Care Index and a Proposed Adequacy of Prenatal Care Utilization Index." *Am J Public Health* 84:1414-20.

Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70:41-55.

Smith, J. and P. Todd. 2005. "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics* 125:305-353.