Preference for Boys, Family Size and Educational Attainment in India

Adriana Kugler and Santosh Kumar*

March 2014

Abstract

Using data from nationally representative household surveys, we test whether Indian parents make trade-offs between the number of children and investments in education and health of their children. To address the endogeneity due to the joint determination of quantity and quality of children by parents, we instrument family size with the gender of the first child. Given a strong family preference for male children in India, parents tend to have more children if the first born is a female. Our IV results show that children from larger families have lower educational attainment, are more likely to be illiterate, and less likely to be enrolled in school, even after controlling for parents' characteristics and birth-order of children. The effects are larger for poorer and low-caste families and families with less educated mothers. On the other hand, we find no evidence of a trade-off for health outcomes.

JEL classification: J13, I21, N3, O10.

Keywords: Quantity-Quality Trade-off, Family size, Education, Health, India

---

* Kugler: McCourt School of Public Policy, Georgetown University, NBER, CEPR, CreAM and IZA. Email: ak659@georgetown.edu; Kumar: Department of Economics & International Business, Sam Houston State University, Email: skumar@shsu.edu

We gratefully thank Josh Angrist, Dean Karlan, Ganesh Seshan, Gary Solon, Dan Westbrook, and seminar participants at Univeristy of Georgetown-Qatar, University of Gottingen and Sam Houston State University for many helpful comments. We also wish to thank Nisha Sinha for excellent research assistance. An earlier version of this paper was circulated as "Testing the Children Quantity-Quality Trade-Off in India".

1. Introduction

High levels of population growth have long been considered a potential deterrent for economic growth and development because it reduces savings and generates overuse of scarce resources. By contrast, human capital accumulation is considered one of the main determinants of income growth. An educated and healthy population is essential for the production of more sophisticated goods and a key determinant of technical change in the economy. At the household level, family size and human capital also move in opposite directions; a larger family has less resources to devote to each child's education and health. That is, resource-contrained households may face a quantity-quality trade-off in terms of their child-rearing decisions (Becker and Lewis 1973).[1]

In this paper, we test the empirical validity of a children quantity-quality (Q-Q) trade-off in India. Testing the Q-Q trade-off in the Indian context is important because Q-Q trade-offs likely affect additional margins of education and are likely to be stronger in resource-contrained households in developing countries. In addition, we are able to exploit the preference for male children in India as a natural experiment to examine the causal effect of the quantity of children on the investments that parents make on their children. By contrast, in high income countries, the preference for gender balance among children is often used as an exogenous source of variation for family size.

Finally, the Indian context is important in its own right if one is trying to understand low human capital investments in one of the most populous countries in the world. Today, India is the second most populous country in the world with over 1.2 billion people and it

[1] Becker and Lewis (1973) coined this term and developed the original quantity-quality model.

is projected that by 2025 it will become the most populous country in the world. Population growth in India has been high over the last few decades and continues at 1.4%. Its fertility rate is still above replacement levels with 2.5 children born per woman. Yet, many children in India lack access to a good education and to quality health services. Per 2011 census, the literacy rate (age 7 and above) is 73% meaning that India is home to about 330 million illiterates. There is wide gender disparity in the literacy rate and the disparity is even more prnounced in rural areas. The school drop-out rate is alarming as over 80 million children fail to complete the full cycle of elementary school in 2011 (IHDR, 2011). Too many children in India still suffer from poor health outcomes. For example, among under 5-year old children, about 45% suffer from malnutrition and over 30% do not receive basic immunizations.[2] Thus, it is important to quantify the extent to which households make Q-Q trade-offs to understand if family planning or other policy initiatives that reduce the number of children could also encourage increased human capital investments by parents.

Empirical testing of the Q-Q trade-off is often challenging because fertility decisions and investments in children are jointly determined and depend on some unobserved common factors (Browning 1992; Haveman and Wolfe 1995). For example, more educated parents may choose to have fewer children, but the children of more educated parents are likely to receive more education themselves. Even if a researcher controls for observable parental characteristics, parental preferences and other unobservable household characteristics may affect both the number of children and investments in children. For example, if parents who are more concerned about future opportunities for their children may choose to have fewer children and spend more time

[2] These statistics health outcomes in India come from the World Development Indicators publication of the World Bank.

and educational resources to educate each child. Omitted variable bias of this type will tend to exaggerate the negative relation between family size and human capital investments.

To address this issue, we employ an Instrumenal Variable strategy (IV) and use the gender of the first child to instrument family size as measured by the number of children. A key identification assumption is that the gender of the first child is a good predictor of the number of children. The preference for male children in India means that if a household has a first born who is a girl, they will continue to have more children until they have a boy. Thus, in a son-biased society, the first child's sex should be a good predictor for the probability of having a second child or of the total number of children in the household. Son preferences are indeed widely documented in countries such as India, China, and Korea and deeply rooted in social, economic and cultural factors. India is a patrilineal society; which means that sons carry the family name and inherit the family's patrimony. They are responsible for the care of elder parents and also have better labor market opportunities, so parents prefer to have sons who can support them in older age. In addition, the dowry system in India provides another reason why parents may prefer boys over girls. Parents who have girls have to provide a generous dowry when their daughters marry, while parents of boys benefit from this tradition.

The other key identification assumption is that the gender of the first-born does not have an independent impact on the educational performance or health of subsequent children in the household. In fact, while parents may have a preference for having boys, boys and girls in India are nearly equal in numbers and in terms of primary and secondary school enrollment. Also, there is little evidence that households who have a first-born female are different in other ways from those who have a first-born male. One reason why the gender of the first-born may be related to educational performance for reasons other than family size is if sex-selective abortions are likely to reduce the chances of having a first-born who is a girl. However, India passed the Pre-natal Diagnostic Technique Act in 1996, which made fetal-sex determination illegal. In addition, Bhalotra and Cochrane (2010), Ebenstein (2007), Jha et al. (2011), Portner (2010), and Rosenblum (2010) found no evidence of selective abortions for the first-birth but only for subsequent pregnancies in India.

We use the District Level Household Survey (DLHS) from 2007-08 to examine the impact on educational outcomes and the National Family Health Survey (NFHS) from 2005-06 to examine the impact of family size on weight and height of young children. Our Ordinary Least Square (OLS) results show negative correlations between family size and literacy, school attendance, current enrollment, and years of schooling. While our regressions control for a large number of children and parental characteristics, there could be unobservable factors affecting both household size and children's investments. To address the endogeneity of family size we instruement family size by gender of the first born. First-stage results show that having a first-born female is strongly positively correlated with family size and Anderson-Rubin and Stock-Wright tests of weak instruments are rejected. The IV results show that an extra child in the family reduces schooling by 0.08 years and reduces the probability of ever attending or being enrolled in school by between 1 and 2 percentage points, respectively. Also, an additional sibling reduces the probability of being literate by 3 percentage points. We also find interesting heterogeneous effects. We find larger effects for illiterate, poor, and low-caste mothers. The impacts of an extra child in terms of reducing enrollment and attendance double and years of schooling increase fourfold for illiterate and poor mothers, suggesting much larger gains from reducing family size in disadvantaged households. On the other hand, our IV estimates of the impacts of family size on health outcomes, including, height, weight, height-for-age, weight-for-age, weight-for-height, malnutrition and stunting show no significant effects. The impact on education but not on health of children may be due to differences between the public education and health care systems in India. The public education system is poor and dysfunctional, while the Indian health care system is designed to provide basic health support even to poor families.

The rest of the paper proceeds as follows. In Section 2, we provide a review of the literature and highlight the contribution of our paper. In Section 3, we present the empirical strategy for the Q-Q trade-off analysis. In Section 4, we describe the data. In Sections 5 and 6, we present average and heterogenous effects of family size on education results. In Section 7, we discuss the effects of family size on health, and we conclude in Section 8.

2. Literature Review

Since Becker and Lewis developed the quantity-quality model, a number of studies have tried to quantify the magnitude of Q-Q trade off. These studies address the endogeneity of family size by taking advantage of exogenous variation in policy experiments (one-child policy in China), natural occurrences of twin birth, and sibling sex composition. While the original causal test of the Q-Q trade-off was conducted using data from India in the 1980s, there has been renewed attention on this topic in developed and developing countries.

A number of studies have used twins to study the Q-Q trade-off in high-income countries, including the U.S., France, Israel, the Netherlands, and Norway. Black, Devereux, and Salvanes (2005) use twins as an instrument for family size using Norwegian data and found no evidence that family size affects educational attainment of children, after controlling for birth order and other potential confounding variables. Similarly, Angrist, Lavy, and Schlosser (2010) use multiple births and same-sex siblings in families with two or more children as instruments for family size in Israel. They also found no significant relation between family size and schooling and employment, but they do find an effect on woman's early marriage. Similarly, Haan (2005) demonstrated that having more children in the family does not have a significant effect on the educational attainment of the oldest child either in the U.S. or in the Netherlands. Using the Public Use Microdata Sample from US and gender composition of the first two children as instrument, Conley and Glauber (2005) find that children living in larger families are less likely to attend private school and are more likely to fall behind in school. Similarly, Caceres-Delpiano (2006) finds that an additional younger sibling in the family reduces the likelihood that older children attend private school, reduces mother's participation in the labor market, and increases the likelihood of their parents' divorce. In contrast, the impact of family size on measures of child well-being such as highest grade completed and grade retention is weak and unclear. Juhn et al. (2012) instead use the National Longitudinal Survey of Youth (NLSY) data to examine the impact of family size on cognitive and non-cognitive abilities during youth as well as long-term outcomes for the U.S. This study shows that growing up with an additional sibling reduces a child's educational attainment by a third of a year and that a larger family size decreases labor market participation and family income and increases the likelihood of criminal behavior and teenage pregnancies. A similar study in France reports an evidence of trade-offs; the school performance of children from large families was worse than children with smaller families (Goux and Maurin, 2005).[3]

[3] Using marital fecundability- as measured by the time interval from the marriage to the first birth- as a source of exogenous variation in family size, Klemp and Weisdorf (2011) documents a large and significantly negative effect of family size on children's literacy in UK.

Small or no effects of family size on education and other human capital investments in developed countries may be due to the presence of a well-functioning public education system, which may substitute for private education and may still allow parents to provide a good education. By contrast, child labor practices and the absence of good public education may make this trade-off more pronounced in developing countries, where parental investments in education are a substantial part of a family's budget.

To the best of our knowledge, Rosenzweig and Wolpin (1980) were the first to examine the empirical validity of Becker-Lewis's Q-Q trade-off model in a developing country context. They exploited *twins* as an exogenous increase in the family size and found a weak negative effect of family size on educational attainment as well as on consumption durables for non-twin children. However, the study was based on a small non-representative sample of 1633 households that only included 25 households (HHs) with twins.

In recent years, the Q-Q literature has grown for developing countries. In China a number of studies have found mixed evidence of the Q-Q trade-off (Li et al., 2008; Qian, 2009; Rosenzweig and Zhang, 2009). Li et al. (2008) use data from the 1% sample of the 1990 Chinese Census and rely on twin births as an instrument. They found that larger family size reduces a child's education even after controlling for birth order effects, especially in rural China. In contrast, Qian (2009) uses a sample of households from China Health and Nutritional Survey and relies on the relaxation of the one-child policy's impact to look at the impact of family size on children's human capital. She finds a positive effect of family size on education, which she attributes to economies of scale. Rosenzweig and Zhang (2009), who also rely on *twins* as an exogenous source of incease in the number of children, find that an extra child significantly decreases schooling progress, expected college enrolment, grades in school and self-assessed health of all children in the family, thereby asserting that Q-Q- trade-off does exist in China. However, they argue that the use of *twins* as instrument generates upward biases on estimates of the Q-Q trade-off because of differences in birth weight, which change parent behavior and overall resource allocation.

Studies for other developing countries that mainly rely on the twinning experiment tend to show small or no effects. An exception of a study that does not rely on the twinning experiment is Jensen (2012), who uses infertility as an instrument to explore the causal effects of family size on a child's nutrition in India, and found significant results only for girls. Millimet and Wang (2011) use mixed sex composition to identity the impact of family size on health outcomes in Indonesia and find evidence of a Q-Q trade-off only in some families. They find statistically significant results on Q-Q trade-off only at the tails of the BMI distribution i.e. the 20^{th} and 85^{th} percentile. Using twinning as an instrument, Ponczek and Souzay (2011) also report negative effects on educational outcomes in Brazil, while Aguero and Marks (2008) find evidence of Q-Q trade-off only for health indicators (weight-for-age and breastfeeding practice) but not for education.[4] Additionally, Glick et al. (2007) use twinning at first birth and find that unplanned fertility increases the nutritional status and school enrollment of later-born children in Romania. Using Matlab Health and Socioeconomic Survey (MHSS) from Bangladesh, Peters et al. (2013) found little evidence of trade-off between child quantity and health.

Sarin (2004) and Lee (2009) are the only other studies, aside from ours, to use preference for boys as an instrument. Sarin (2004) after instrumenting the family size by sex of the first born and multiple births found a positive but insignificant relationship

[4] The study uses data from Demographic and Health Surveys in Latin America, namely, Bolivia (conducted in 1994 and 1998), Brazil (1996), Colombia (1995 and 2000), the Dominican Republic (1996), Guatemala (1998), Nicaragua (1998), and Peru (1996), and instrument the family size by a mother's infertility status.

between family size and weight-to-height ratio.[5]

Lee (2008) is another study that have used preference for boys as an instrument for family size. However, instead of using direct measures of educational attainment, Lee's study uses parents' monetary investment in education to measure child's quality. According to IV estimates, per-child investment for households with two children and three children are 74.6% and 57.6 % higher, respectively, compared to households with no children.

Our paper adds to the existing literature on Q-Q trade-off in a number of important ways. First, our study uses a credible and novel instrument in a developing country context combined with good measures of child quality. This is important, because most studies have relied on the twinning experiment, which Rosenzweig and Zhang (2009) indicate that *twins* are likely to generate biases. Second, while many studies have focused on China and other regions of the developing world, this is the first study to focus on the impact of family size on educational outcomes in India since the original twining study of the 1980s, which relied on a very small sample. Not only is India host to 17% of the world's population and important in its own right, but imperfect credit markets and lack of good public schools are likely to affect the extent and severity of the Q-Q trade-off. Third, only a handful of studies have examined the effect of family size on child health; most of the previous studies have focused on educational attainment/progress. By including health as an outcome, we contribute to the thin literature on the effect of family size on child health in a Q-Q framework.

[5] Our paper differs from Sarin (2004) as we include educational attainment as outcomes in addition to health, use rich set of controls with large sample size and covers time period when sex-selective abortions was minimal due to legal ban on fetal sex determination.

3. Empirical Framework

We estimate the effect of family size on children's educational and health outcomes using OLS and 2SLS (instrumental variable) regression analyses. We first estimate the following OLS model:

$$Y_{ihd} = \beta_0 + \beta_1 * FamilySize_{hd} + \beta_2 X_{ijd} + \mu_d + \epsilon_{ihd} \quad (1)$$

where Y is the educational and health outcomes of child *c* in household *h* residing in district *d*. The educational outcomes of the child are measured at the probability of being literate, probability of ever attending school, probability of currently being enrolled, and years of schooling. We also estimate equation (1) for several child health measures, including the height-for-age, weight-for-age, weight-for-height and whether the preson is underwieght, stunted and wasted. The variable *FamilySize* is the number of children under 21 years of age in the family; X is a vector of covariates, and € is an error term. The covariates include the following children as well as parental characteristics: age, gender, ethnic group, birth order and place of residence (rural vs. urban), and the age and education levels of the child's father and mother. The main coefficient of interest is β_1 which captures the existence of Q-Q trade-off. A negative value of β_1 would mean that a trade-off between the quantity and quality of children does exist in India.

β_1 will provide the causal impact of family size on child quality only if family size is exogenously determined. However, there are several factors that may render the family size variable endogenous and non-random. One such factor is the fact that fertility decisions and decisions about investment in child's quality are jointly determined. In this case, the OLS estimate of β_1 *in* equation (1) is subject to endogeneity bias and is unlikely to capture the causal effect of family size on child quality. The OLS estimates may be downwardly or upwardly biased depending on nature of endogeneity. For example, in a country like India, wealthier households may have fewer children and also invest more in their children's schooling, thus generating an upward bias in the Q-Q trade off. However, highly committed parents may have more children and also invest more in their children's education, thus generating downward biases.

In order to capture only exogeneous variation in family size, we rely on an Instrumental Variable (IV). The challenge with any instrument is to identify a variable that predicts *FamilySize* but is uncorrelated with the error term in equation (1). To address the endogeneity of family size, we use sex of the first-born (SFB) as an instrument and estimate the following two-stage least square model:

$$FamilySize_{hd} = \alpha_0 + \alpha_1 * SFB_{hd} + \alpha_2 X_{ihd} + \mu_d + u_{hd} \qquad (2)$$

$$Y_{ihd} = \pi_0 + \pi_1 * \widehat{FamilySize}_{hd} + \pi_2 X_{ihd} + \mu_d + v_{ihd} \qquad (3)$$

where *FamilySize* the number of children under 21 years of age in the family, *Y* is the outcome variables, *SFB* is the sex of the first birth and is the instrumental variable for *FamilySize*, and X denotes a vector of exogenous regressors at child and parent's level. The SFB is a dummy variable that equals 1 if the first-born is a female and 0 otherwise. We also include district fixed-effects to control for time-invariant differences across districts. Standard errors are clustered at the district level.

Equation (2) is the first-stage while equation (3) is the second-stage regression. The second stage regresses the outcomes on the predicted value of *FamilySize* from equation (2) and other exogenous variables. In addition, we estimate the 2SLS regressions for a number of sub-groups including: different castes, different quintiles, and different levels of educational attainment of the mother and for urban and rural sub-samples, separately.

A key condition for the gender of the first child to be a valid instrument is that family size be highly correlated with the gender of the first child, i.e., *Corr(SFB, FamilySize*)$\neq 0$. In India, there is a long-standing social and cultural norm which is the preference for sons over daughters. First, only sons carry the family name. More importantly, since India is a patriarchical society, sons inherit the family's patrimony. Parents also prefer male children as sons are supposed to provide financial support and care for their parents in old age. In addition, since males are more likely to join labor markets and earn higher wages, they further contribute to a family's preference for boys. In Indian tradition, daughters are married out and become part of another family. Because parents have to provide a dowry when daughters get married, a family will prefer to have boys so they can receive a dowry when their sons gets married. In this type of patrilineal familial system the gender of the first-born is likely to have important implications for family size. In particular, if the first born is a girl, parents are likely to want to continue having children until a son is born. In Section 4, we test for this empirically by estimating the first-stage relationship in equation (2).

The second key underlying assumption behind this identification strategy is that the sex of the first born is uncorrelated with educational or health outcomes other than through family size, i.e., *Corr(FamilySize, €*)=0. Since sex of the first child is determined by nature, this is considered a random event and uncorrelated with educational attainment and health. However, if parents have any control over births, and they make decisions over births depending on sex, the sex of the first birth will not be random. Therefore, the presence of sex-selective abortion may undermine the validity of the instrument because the access to ultrasound use and abortion services allows parents to choose the sex of their children. This may not be as big a concern in our study given that the Pre-natal Diagnostic Technique (PNDT) Act was passed in India in 1996 making fetal-sex determination illegal. In addition, many previous studies have shown that parents do not use sex-selective abortion for *first-borns* but only for second-borns in India; these studies find that the sex-ratio at first birth lies within the biologically range of 1.03-1.07 (Bhalotra and Cochrane (2010), Ebenstein (2007), Jha et al. (2011), Portner (2010), and Rosenblum (2010)).[6] Sociological studies also provide evidence that parents only have a strong preference for sons after this first birth (Patel, 2007). Using the same data as ours, Rosenblum (2010) reports lack of sex-selection abortion at the first-parity. About 36 percent of women report induced abortions at the second and third-parities. Given that sex-selection abortion became illegal in India in 1996, zero and positive reporting of induced abortion at first and high-order, respectively, further provides confidence that sex-selection at first-parity is not rampant and gender of first-born can be treated as exogenous. Additionally, using the first two rounds of the National Family and Health Survey, Retherford and Roy (2003) report little or no evidence of sex selection at the first-birth. Jha et al. (2011) use the National Family Health Survey and find no significant declines in the sex ratio for first-births or second-order births if the first-born was a son. By contrast, they find a decline of 0.52% in the sex ratio for second-order births when the first-born was a girl, and they note that these declines were much greater for

[6] In the absence of any interventions, the probability of having a son is approximately 0.512 and this probability is independent of genetic factors (Ben-Porath and Welch 1976; Jacobsen, Mïller and Mouritsen 1999).

mothers with 10 or more years of schooling and for wealthier households. Taken together, these studies provide evidence that sex of first-birth is indeed exogenous and random.

To further address the exogeneity of the instrument, we explore whether the instrument is correlated with other observable characteristics of the household to gauge whether the sex of first-birth can also be assumed to be uncorrelated with unobservables. In the next section, we report results of probability models of the sex of the first-birth on a vector of explanatory variables.

4. Data Description

For educational outcomes, we use data from the Indian District Level Household Survey (DLHS) collected in 2007-08. The sample is representative at the district level, the lowest tier of administration and policy-making in India. The DLHS covers 601 districts and on average draws a random sample of 1,000-1,500 households from each district.

The survey has four parts: a household questionnaire: a questionnaire for ever-married women (15-49 years); a questionnaire for unmarried women (15-24 years); and a module covering village and health facilities characteristics. Our study is based on the household questionnaire, that collected information on assets, number of marriages and deaths in the household since January 2004 and socio-economic characteristics of all members of the household. In particular, the survey collects the following information for each household member: age, gender, schooling attendance, literacy, and years of completed schooling. We identify individuals who are labeled sons/daughters, and estimate the family size by counting the number of sons/daughters in the household, and merge these data to the parents' information.

We restrict the sample in the following ways. First, we restrict the sample to individuals who are either parents (head of the household and spouse) or who are either sons/daughters of the head of the household.[7] Second, we restrict the sample to households with at least one child so that we can use gender of the first child as the instrument. Third, we restrict the sample to children of school going age who are 5 years of age or older but are under 21. We use 5 as the lower age bound because the household roster only collects education information of all individuals older than 4. In India, primary school (grades 1 to 5) begins at age 5 or 6 and ends at age 10 or 11, while high school is usually completed by age 18. However, due to deferred enrollment or grade repetition completion of either primary or secondary schooling might get delayed. We exclude mothers over 35 years of age to minimize the possibility that adult children may have already left the household. Finally, we exclude households with missing or unreliable information on any of the variables used in the analysis. This yields a sample of 393,597 children.

The main outcome variables this paper analyzes are several measures of educational attainment. The outcome measures are: an indicator of whether the person is literate or not; an indicator of whether the person ever attended school or not; an indicator of whether the person is currently enrolled in school or not, and years of schooling. In all the models we have described, we control for the following covariates: caste, religion, an asset-based standard of living index, mother's age, father's age, mother's education, and father's education. For the caste variable, we consider three groups: scheduled caste and scheduled tribe are combined together to constitute the low caste category (a group that is socially segregated and disadvantaged), and other backward classes (officially identified as socially and educationally backward) as middle caste, and the upper caste (comprising Brahmins and other higher castes that are privileged) as high caste. We consider 4 major religious

[7] We drop individuals who are son or daughter-in-law, grandchildren, parent, parent-in-law, brother/sister, brother or sister-in-law, niece or nephew, other relative, etc.

groups, Hindus, Muslims, Sikhs, and Christians. The DLHS data does not contain information on individual or household incomes. The survey asked a multitude of questions about the ownership of assets including ownership of a car, television, real state property, and other assets. The DLHS uses ownership of assets to create a standard of living index with three categories: low, middle and high.[8]

Table 1 reports the summary statistics of these individual and household variables. The average age of children in the sample is 10 years and the average years of schooling is 3.08. About 51 % of first born children are male. Fathers are older than mothers; the average age of mothers is 31 years and the average age of fathers is 36. As expected, mothers have less education than fathers. The average years of schooling for mothers and fathers are 3.0 and 5.5 years, respectively. The average family size is 3.54. Approximately 82% of children live in rural areas. About 41% of the children are classified as low caste and 20% high caste. Finally, around 49% of children have the lowest standard of living index, 39% a middle standard of living, and 12% the highest standard of living.

For health outcomes, we use the National Family Health Survey (NFHS) from 2005-06. NFHS is India's primary and only source of data on health and nutrition. Women between the age of 15-49 were sampled and interviewed to collect anthropometric data for their children who were under 5 years of age at the time of the survey. The survey collects anthropometric data for children present in the household at the time of the interview. Therefore, for analyzing health outcomes, we are constrained to use households where the oldest child is at most 5 years old at the time of the survey. In particular the survey collects

[8] By combining household amenities, assets and durables, the DLHS data computed a wealth index and divided into quintiles. The principle of factor loading to amenities, assets, and durables derived by factor analysis is used for the computation of the wealth index. Households are categorized from the poorest to the richest groups corresponding to the lowest to the highest quintiles.

information on height and weight of the children. Using these anthropometric information, the NFHS also reports height-for-age z-score, weight-for-age z-score, and weight-for-height z-score based on World Health Organization guidelines.

We analyze several health outcomes, such as weight in kilograms, height in centimeters, height-for-age z-score (haz), weight-for-age z-score (waz), and weight-for-height z-score (wfh). In addition, we examine the impact of family size on probability of underweight (waz < -2), probability of stunting (haz < -2), and probability of child being wasted (wfh < -2). We use the same set of controls as used in the analysis of educational outcomes. In terms of econometric specification, the only difference is that instead of district fixed-effects, we include state fixed-effects since NFHS does not provide district information unlike DLHS.

Our health sample is comprised of 10,090 children with non-missing data on height and weight. Table 2 provides summary statistics for the analytical sample. The average age of children in the sample is 28 months. The average weight-for-age z-score and height-for-age z-score are -1.63 and -1.47, respectively. Z-scores of minus one and one indicate that a child is one standard deviation below and above the median of reference population, respectively. About 39% of sampled children are *underweight* and 35% of them are *stunted*. The percentage of *wasted* children is 15%. A majority of the children live in rural areas (61%) and the average family size is 2.17.

Table 3 reports results of linear probability and probit models of the likelihood that the first-born is a girl on the characteristics reported in Table 1 to address whether the instrument is likely to be exogeneous. Table 3 shows that most of the explanatory variables are statistically insignificant. The exceptions are mother's age and father's schooling. The older the mother, which is given by nature, the higher the probability that the first-born will be a girl. The results also show that the less educated the father the less likely the first-born will be a girl, but the effects are very small. Furthermore, as shown in Table 1, about 51% of first-born are male indicating that the sex-ratio at first birth is in the biological range.

5. Effects of Family Size on Educational Attainment

The outcome variables in this study include a set of indicators of educational attainment or educational progress. These include: (a) literacy; (b) ever attended school; (c) school enrollment; and (d) years of schooling for those who ever attended school. Except for years of completed schooling, all other variables are binary and are coded as 1 or 0. The main independent variable, *FamilySize*, is continuous and is measured by the total number of 0-20 years old children in the family at the time of survey.

Table 4 reports OLS regression results which control for children's and parents' characteristics. Additionally, district fixed-effects are included to account for time-invariant district characteristics. Results in Table 4 indicate a significantly negative correlation between family size and children's education. Column 1 shows that an extra child in the family reduces the probability of literacy by 2 percentage points. Column 2 and 4 show that children in families with one additional child are 1.8 percentage points less likely to have ever attended school and the likelihood that they are currently enrolled in school is 1.4 percentage points lower. The results for years of schooling also indicate harmful effects of family size on child's education. For years of schooling, the point estimate is -0.2, suggesting that on average children in families with 5 more siblings will end up with a a year less of schooling.

Recognizing the limitation of interpreting the OLS estimates in Table 4 as causal, we then proceed to instrument the main endogenous variable, family size, by sex of the first-birth and estimate the same relationship using 2SLS. We first check for the relevance condition in Table 5. From the first-stage regression, it follows that the instrument is highly significant and has a positive effect on family size. The first row in Table 5 shows that family size increases by 0.22 children when the first-born is a girl and the effect is significant at 1% level of significance.[9]

The second row in table 5 presents result from the estimation of 2SLS model. The IV results presented in Table 5 show a negative and significant impact of family size on children's quality. The IV estimates for literacy is 2.8 percentage points, which corresponds to a decrease of 3.4 % of the probability of being literate. Similarly, the point estimates for ever being in school and current enrollment are negative and statistically significant, confirming that the detrimental effects of family size on children's education comes from not attending school or dropping out of school along the way. Columns 2 and 4 show that the probability of ever attending schooling and being currently enrolled in school drop by 1.8 and 1.1 percentage points when an additional sibling is added to the family. Consequently, years of schooling fall as well. However, the IV estimate in Column 4 of Table 5 is 2.5 times smaller than the OLS estimate in Table 4. The IV estimate indicates that an exogenous increase in household size of one extra child decreases the years of schooling by 2.6% compared to 6.5% when relying on OLS estimates.

Once we account for the endogeneity of family size (using IV), the coefficients are, for the most part, smaller (or less negative) compared to the ordinary least square estimate, with the exception of the literacy estimate. The IV estimates suggest that OLS coefficients

[9] The first stage coefficient is smaller than the twins first-stage of about 0.6 in the Angrist and Evans (1998) study. It is possible, that the birth of of a girl as the first-birth results in a smaller increase in family size due to the fact that Indian families are larger to begin with.

overestimate the true trade-off and are biased toward finding effects that are too large suggesting that unobservables which drive parents to have big families also drive parents to invest too little in their children.

Table 5 also reports the Kleibergen-Paap rk Wald test to detect if the instrument suffers from a weak-IV problem. The first-stage F-stat and Kleibergen-Paap rk Wald Stat, both are significant, suggesting that our analysis does not suffer from weak identification. We also provide Anderson-Rubin F-test Statistic and Stock-Wright S-statistic in Table 5 to assess that our second-stage results are robust to weak-instrument inference.

6. Heterogeneity in the Quantity-Quality Trade-off

6.1 . Caste Differences in the Q-Q Trade Off

Given the disadvantaged situation of lower castes in India, one may expect lower and even middle castes to have less access to good public schools and less access to markets than higher castes.

We capture the heterogeneity in the Q-Q trade off across different caste categories by estimating the regressions for low, middle and high castes separately. Results in Table 6 shows that once the family size is instrumented, the effects of family size on literacy are greatest for low caste children and then for middle caste children but not significant for high caste children. The impact of an extra child reduces the likelihood of being literate by 5.8% among low caste children, by 2.6% among middle caste children and has not impact on the literacy of high caste children. Similarly, the effect of family size on the likelihood of ever attending school and actual years of schooling is greatest for low caste individuals. For example, an extra sibling in the classroom reduces the years of schooling by 5.7%

compared to 2.6% on average in the previous section. The effects of family size on current enrollment are greatest on middle caste children.

6.2. Rural-Urban Differences in the Q-Q Trade Off

Given the lack of good public school in rural areas in India, we may expect for the Q-Q trade-off to be greater in rural than in urban areas. Indeed, there are large rural-urban gaps in educational attainment. For our sample children, the primary school completion rate is 35% in rural areas while it is 41% in urban areas.

Indeed, the impact of of having larger family size is mostly larger and statistically significant in rural compared to urban areas, suggesting the quantity-quality trade-off is more pronounced in rural India. The last 4 columns of Table 6 show the OLS and IV estimates for rural and urban households separately. The impact of family size on literacy, the likelihood of ever attending school, and on years of schooling is greater in rural than urban households, once the endogeneity of family size is taken into account by using sex of the first-born as an instrument. The trade-off coefficient is -0.3 for literacy, -0. 018 on ever attending school, and -0.107 for years of schooling in rural areas (it is zero in urban areas). These coefficients imply that an extra child reduces literacy by 3.7%, reduces the likelihood of ever attending school by 2% and years of schooling by 3.6% in rural households. This finding is similar to Li et al. (2008), who also report that trade-off was more evident in rural parts of China and was negligible in urban areas. Family size affects the likelihood of current enrollment for both rural and urban households.

6.3. Wealth and the Q-Q Trade Off

The degree of trade-off might also be different across household wealth levels. Wealthier households are less likely to be subject to credit constraints when making the choice between the number of children and the educational opportunities offered to each child. We divide the sample by household wealth levels to explore whether the trade-off differs among low, median, and high-wealth levels and run the IV model separately for these subsamples. The first six columns of Table 7 show the OLS and IV results for the bottom two quintiles, the third quintile and the top quintile. The effect of an extra child on literacy, the likelihood of going to school, years of schooling, and of being currently enrolled in school are all greatest for those in the bottom two quintiles. By contrast, there are no effects of an extra child on literacy, the likelihood of ever attending school and years of schooling in the top quintile. For those in the bottom two quintiles, the effects are substantial. An extra sibling reduces literacy by 6.9%, the likelihood of attending school and being currently enrolled by 4.8% and 1.9%, and years of schooling by 10.6% among households at the bottom quintiles of the wealth distribution.

6.4. Does Mother's Educational Attainment affect the Q-Q Trade Off?

Mothers play a key role in the household by making decisions in regards to expenditures and by providing a supportive environment for children. Less educated mothers will generally be worse positioned to provide support for children in their studies and may not be well informed about possible alternatives to support their children leading to a bigger Q-Q trade off.

The last four columns in Table 7 provide the coefficients of OLS and IV regressions for illiterate mothers and mothers with less than primary schooling. The IV results show that the effects of an extra child on educational attainment are greatest for children of illiterate mothers, such as the trade-off coefficient on the likelihood of being literate and ever attending school for children of illiterate moms. The likelihood of attending school falls by 4.6% when an additional sibling is added to a household headed by an illiterate mother but has no effect on households with mothers who have even some primary schooling. By contrast, the impact on current enrollment is similar independent of mothers' schooling. However, while literacy and years of schooling fall for children in both households headed by illiterate mother and mother with some primary schooling, the effects are much greater on the former.

All in all, thus, the Q-Q trade offs are more pronounced among lower caste, rural and poorer households, as well as households headed by less educated mothers probably because these households face greatest credit constraints, worse public school systems and are less able to compensate for bad schooling by educating their children at home or by private tutoring.

7. Effect of Family Size on Children's Health

Child malnourishment is another serious bottleneck that affects underdevelopment. There is a growing literature on the interaction between early-life health and human capital accumulation (see Bleakley, 2010 for detail). Bleakley (2010) argues that poor childhood health might depress the formation of human capital, which in turn can affect lifetime income either through schooling or labor-market productivity channel. About 48% of all children in India are malnourished, either moderately or severely underweight (IIPS, 2010). For many years, the Government of India has implemented several health programs including the Integrated Child Development Scheme (ICDS) and the National Rural Health Mission (NRHM) to improve the nutritional status of children. Despite this, malnourishment levels among Indian children are very high compared to other south-Asian countries. Therefore, it is quite plausible that high fertility rates also might be contributing to this problem. Given that health is another component of child quality, we test whether there is a trade-off between the number of children and the health of each child in this section.

We use the National Family Health Survey to examine the trade-off in child's health. Table 8 examines the effect of family size on a number of nutrition measures, including height, weight, height-for-age, weight-for-age, weight-for-height, and whether the person is underweight or stunted. Panel A in Table 8 shows the estimates from OLS, while Panel B shows the results from 2SLS model. The OLS results provide evidence in support of Q-Q trade-off. Results indicate that each additional child in the household lowers the average weight of the children by 177 grams and height by 0.8 cm. Columns 6 and 7 of Table 8 show that children born in larger families are also more likely to be underweight and stunted, repectively.

Panel B shows the corresponding IV estimates. The IV analysis fails to detect any significant evidence of Q-Q trade-off in health outcomes. None of the estimates are statistically significant at conventional levels of significance. This implies that once we control for the endogeneity of the family size, the Q-Q trade-off between family size and health outcomes disappear. One plausible exaplanation for insignificant effect of family size on nutrition could be that the health system in India, unlike the education system, provides support to families so that vast trade-offs for children may not be visible.

8. Conclusions and Discussions

Testing the theoretical trade-off between the quantity and quality of children has been on the research agenda for a long time, however, the empirical evidence supporting the prediction of the Beckerian model is limited. The empirical evidence has been mixed so far. A few studies find a negative effect of family size on the quality of children, measured by either education or health status (Rosenzweig and Wolpin, 1980). In contrast, others find no empirical support for the child quantity-quality trade-off (Black et al 2005; Haan 2005). A variety of instruments such as twinning, sex of first child, sex of first two child, infertility etc. are used to address the endogeneity concern.

In this paper, we have used data from households in India to test the empirical validity of child quantity-quality trade-off. A strong preference for sons over daughters in Indian societies allows us the use of a novel instrumental variable, namely sex of first birth, to test the Q-Q trade-off. Testing this model has important policy implications. From a policy point of view, it is important to know the extent to which a policy formulated to control population improves the human capital of the country and quality of the labor force. Not only the quantity of human capital plays a role, rather quality of human capital is equally important for economic development.

We find that the Beckerian theory of child quantity-quality trade-off holds in India. Family size has a significant negative causal impact on educational outcomes of children. After controlling for potential endogeneity, an additional child in the family reduces the literacy rate for all children by 3.4% and years of schooling by 2.6%, hence a strong support for Becker's trade-off hypothesis. The observed trade-off persists after including child and parent characteristics. We find non-uniformity in the existence of trade-off between rural and urban India. The negative relationship between family size and children's education is more pronounced among rural households who are severely budget-constrained. Urban children are less likely to face the trade-off.

The effect also differs by caste, mother's education level, and household wealth. For children belonging to low and middle caste, the trade-off is severe compared to high-caste children. More educated mothers are also able to mitigate the trade-off because the trade-off is only evident for illiterate mothers. Similarly, we observe a wealth-gradient in the trade-off across wealth groups; the trade-off is more pronounced in low-wealth households with an extra children reducing the years of schooling by as high as 0.3 years.

Quanitfying the causal estimate of family size on child quality is important from public policy perspective as well. Since the majority of large families in developing countries are poor, less educated, and resource-constrained, our findings can help us better understand why poverty persists and how people can be moved out of poverty. Improving access and uptake of family planning methods and public policies aimed at increasing awareness about the benefits of having a smaller family may help weaken the severity of the trade-off. Furthermore, policymakers in developing countries can supplement the family planning policies with more investment on education and health in regions and households for which the trade-off is severe in order to mitigate the adverse impacts of larger families.

References

Aguero, Jorge M, and Mindy S Marks. 2008. "A New Approach to Estimate the Effect of Family Size on Child Development." *Mimeo.*

Angrist, Joshua, and Williams Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *American Economic Review* 88 (3): 45-477.

Angrist, Joshua, Victor Lavy, and Analia Schlossar. 2010. "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." *Journal of Labor Economics* 28: 773-824.

Becker, G S, and H G Lewis. 1973. "On the Interaction between the Quantity and Quality of Children." *Journal of Political Economy* 81 (2): S279-88.

Ben-Porath, Yoram, and Finis Welch. 1976. "Do Sex Preferences Really Matter?" *Quarterly Journal of Economics* 90 (2): 285–307.

Bhalotra, Sonia, and Tom Cochrane. 2010. "Where Have All the Young Girls Gone? Identification of Sex Selection in India." *IZA Discussion Paper 5381*.

Black, Sandra, Paul Devereux, and Kjell Salvanes. 2005. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *Quarterly Journal of Economics* 120 (2): 669-700.

Bleakley, Hoyt. 2010. "Health, Human Capital, and Development." *Annual Reviews of Economics* 2: 283-310.

Browning, M. 1992. "Children and Household Economic Behavior." *Journal of Economic Literature* 30: 1434-75.

Caceres-Delpiano, Julio. 2006. "The Impacts of Family Size on Investment in Child Quality." *Journal of Human Resources* XLI (4): 738-754.

Conley, Dalton, and Rebecca Glauber. 2006. "Parental Educational Investment and Children's Academic Risk: Estimates of the Effects of Sibship Size and Birth Order from Exogenous Variation in Fertility." *Journal of Human Resources* 41 (4): 722-737.

Ebenstein, A Y. 2007. "Fertility Decision and Sex Selection in Asia: Analysis and Policy." *UC Berkeley, Mimeo.*

Galor, Oded, and O Moav. 2002. "Natural Selection and the Origin of Economic Growth." *Quarterly Journal of. Economics* 117: 1133-1191.

Glick, P, A Marini, and D E Sahn. 2007. "Estimating the Consequences of Unintended Fertility for Child Health and Education in Romania: An Analysis using Twins Data." *Oxford Bulletin of Economics & Statistics* 69: 667-691.

Goux, Dominique, and Eric Maurin. 2005. "The effect of overcrowded housing on children's performance at school." *Journal of Public Economics* 89 (5-6): 797-819.

Haan, Monique De. 2010. "Birth Order, Family Size and Educational Attainment." *Economics of Education Review* 29 (4): 576–588.

Haveman, R, and B Wolfe. 1995. "The Determinants of Children's attainment: A Review of Methods and Findings." *Journal of Economic Literature* 33: 1829-78.

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

Jacobsen, R, H Moller, and A Mouritsen. 1999. "Natural Variation in the Human Sex Ratio." *Human Reproduction* 14 (12): 3120-3125.

Jensen, Robert. 2012. "Another Mouth to Feed? The Effects of (In)Fertility on Malnutrition." *CESifo Economic Studies* 58 (2): 322-47.
Jha, Prabhat, Maya Kesler, Rajesh Kumar, Faujdar Ram , and Usha Ram. 2011. "Trends in Selective Abortions in India: Analysis of Nationally Representative Birth Histories from 1990 to 2005 and Census Data from 1991 to 2011." *The Lancet* 377 (9781): 1921-1928.
Juhn, Chinhui, Yona Rubinstein, and Andrew Zuppann. 2013. "The Quantity-Quality Tradeoff and the Formation of Cognitive and Non-cognitive Skills." *University of Houston, Mimeo.*
Klemp, Marc, and Jacob Weisdorf. 2011. "The Child Quantity-Quality Trade-Off During the Industrial Revolution in England." *University of Copenhagen, DP # 11-16.*
Lee, Jungmin. 2008. "Sibling Size and Investment in Children's Education: An Asian Instrument." *Journal of Population Economics* 21 (4): 855-875.
Li, Hongbin, Junsen Zhang, and Yi Zhu. 2008. "The Quantity-Quality Trade-Off of Children in a Developing Country: Identification using Twins." *Demography* 45 (1): 223-243.
Millimet, Daniel L, and Lee Wang. 2011. "Is the Quantity-Quality Trade-Off a Trade-Off for All, None, or Some?" *Economic Development & Cultural Change* 60 (1): 155-195.
Patel, Tulsi. 2007. *Sex-Selective Abortion in India.* New Delhi, India: Sage Publications.
Peters, Christina, Daniel I Rees, and Rey Hernández-Julián. 2013. "The Trade-off between Family Size and Child Health in Rural Bangladesh." *Eastern Economic Journal* 40: 71-95.
Ponczek, Vladimir, and Andre P Souzay. 2012. "New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country." *Journal of Human Resources* 47 (1): 64-106.
Portner, Claus. 2010. "Sex Selective Abortions, Fertility and Birth Spacing." *University of Washington, WP # 2010-4.*
Qian, Nancy. 2009. "Quantity-Quality and the One Child Policy: The Only Child Disadvantage in School Enrollment in China." *NBER Working Paper No. 14973.*
Retherford, Robert D, and T K Roy. 2003. "Factors Affecting Sex-Selective Abortion in India and 17 Major States." *NFHS Subject Report 21, IIPS, Mumbai, India.*
Rosenblum, Dan. 2013. "Economic Incentives for Sex-Selective Abortion in India." *Dalhousie University, Mimeo.*
Rosenzweig, M R, and J Zhang. 2009. "Do Population Control Policies Induce More Human Capital Investment? Twins, Birthweight, and China's "One Child" Policy." *Review of Economic Studies* 76 (3): 1149-1174.
Rosenzweig, M R, and K I Wolpin. 1980. "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment." *Econometrica* 48 (1): 227-240.
Sarin, Ankur. 2004. "Are Children from Smaller Families Healthier?: Examining the Causal Effects of Family Size on Child Welfare." *University of Chicago, PhD Thesis.*
The World Bank. 2013. "World Development Indicators." Washington, DC.

Table 1: Descriptive Statistics of the Sample

	All	First-born girl	First-born boy
Child Age (5-20 years old)	9.60	9.41	9.79
	(3.45)	(3.34)	(3.55)
Gender of first child (female=1)	0.49		
	(0.49)		
Literate	0.82	0.81	0.83
	(0.38)	(0.39)	(0.37)
Ever attended school	0.9	0.89	0.91
	(0.30)	(0.31)	(0.29)
Still enrolled	0.95	0.95	0.95
	(0.21)	(0.21)	(0.22)
Years of schooling	3.08	2.94	3.22
	(2.92)	(2.85)	(2.98)
Mother’s age	30.94	30.88	31.00
	(3.36)	(3.34)	(3.37)
Father’s age	36.48	36.42	36.54
	(4.81)	(4.79)	(4.82)
Mother’s years of schooling	2.99	3.05	2.93
	(4.06)	(4.09)	(4.03)
Father’s years of schooling	5.48	5.56	5.40
	(4.74)	(4.76)	(4.72)
Family size	3.54	3.70	3.40
	(1.33)	(1.33)	(1.31)
Rural	0.82	0.81	0.82
	(0.39)	(0.39)	(0.39)
Low caste (SC & ST)	0.41	0.41	0.41
	(0.49)	(0.49)	(0.49)
Middle caste (OBC)	0.39	0.39	0.39
	(0.49)	(0.49)	(0.49)
Low wealth	0.49	0.48	0.49
	(0.50)	(0.50)	(0.50)
Medium wealth	0.39	0.4	0.39
	(0.49)	(0.49)	(0.49)
No. of observations	393,597	193263	200334
No. of districts	601		

Notes: Standard deviations are shown in parentheses. All sampled children were 5-20 years old at the time of survey (2007-08). The analytical sample is restricted to 20-35 years old mother.

Table 2: Descriptive Statistics of the Health Sample

	All	First-born girl	First-born boy
Child Age (months)	28.0	28.31	28.57
	(17.01)	(17.04)	(16.99)
Gender of first child (female=1)	0.51		
	(0.49)		
Weight (Gram)	10251.72	10142.56	10363.79
	(3123.52)	(3061.12)	(3182.72)
Height (Centimeter)	81.85	81.62	82.10
	(13.34)	(13.27)	(13.42)
WAZ	-1.63	-1.64	-1.63
	(1.16)	(1.17)	(1.16)
HAZ	-1.47	-1.46	-1.48
	(1.50)	(1.52)	(1.48)
WfH	-0.92	-0.92	-0.94
	(1.13)	(1.12)	(1.13)
Child is underweight	0.39	0.39	0.39
	(0.49)	(0.49)	(0.49)
Child is stunted	0.35	0.35	0.35
	(0.48)	(0.48)	(0.48)
Child is wasted	0.15	0.15	0.15
	(0.36)	(0.36)	(0.36)
Family size	2.17	2.20	2.16
	(0.42)	(0.44)	(0.39)
Rural	0.61	0.62	0.60
	(0.49)	(0.49)	(0.49)
Low caste (SC & ST)	0.33	0.33	0.32
	(0.47)	(0.47)	(0.47)
Middle caste (OBC)	0.33	0.33	0.33
	(0.47)	(0.47)	(0.47)
Low wealth	0.29	0.29	0.29
	(0.45)	(0.45)	(0.45)
Medium wealth	0.22	0.22	0.22
	(0.41)	(0.41)	(0.41)
Mother's age	24.09	24.14	24.04
	(3.75)	(3.75)	(3.74)
Father's age	29.38	29.46	29.30
	(4.79)	(4.85)	(4.73)
N	10090	5111	4979
No. of states	29		

Notes: Standard deviations appear in the parentheses. Under age-6 sample from NFHS-3 (2005-06).

Table 3: Regression of Gender of First Born on Control Variables

	Dependent variable: First-born is a girl	
	LPM	Probit
	(1)	(2)
Rural	-0.003	-0.007
	(0.004)	(0.011)
Low wealth	-0.006	-0.016
	(0.006)	(0.016)
Medium wealth	-0.004	-0.009
	(0.005)	(0.013)
Religion (Hindu=1)	0.006	0.016
	(0.004)	(0.011)
Scheduled caste/tribe (Yes=1)	0.003	0.009
	(0.004)	(0.011)
Other backward caste	0.002	0.005
	(0.004)	(0.010)
Mother is illiterate	-0.008*	-0.02*
	(0.004)	(0.011)
Mother is primary schooled	-0.003	-0.007
	(0.004)	(0.011)
Father Illiterate	-0.009**	-0.023**
	(0.004)	(0.010)
Father is primary schooled	-0.011***	-0.029***
	(0.004)	(0.010)
Mother's age	0.037***	0.093***
	(0.007)	(0.017)
Father's age	0.002	0.004
	(0.003)	(0.008)

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard errors, clustered by district, are shown in parentheses. All models include district fixed-effects. Column 2 reports marginal effects from the probit model.

Table 4: OLS Estimates of the Effect of Family Size on Education

	Literacy	Ever attended school	Years of schooling	Currently enrolled
	(1)	(2)	(3)	(4)
Family size	-0.020***	-0.018***	-0.20***	-0.014***
	(0.001)	(0.001)	(0.006)	(0.0006)
Children's control	yes	yes	Yes	yes
Parents' controls	yes	yes	Yes	yes
District fixed-effect	yes	yes	Yes	yes
R-square	0.25	0.14	0.71	0.15
No of observations	393597	393510	393597	345985

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard errors, clustered by district, are shown in parentheses. Children's controls include age, age squared, gender and birth order. Parents' control includes education levels of father and mother, household religion, household caste, rural, and household socioeconomic status. Family size is total number of 0-20 years old children in the family at the time of the survey.

Table 5: IV Estimates of the Effect of Family Size on Children's Educational Outcomes

	Instrument: First child is a girl			
	Literate	Ever attended school	Years of schooling	Currently enrolled
	(1)	(2)	(3)	(4)
First Stage	0.219***	0.219***	0.219***	0.227***
	(0.007)	(0.007)	(0.007)	(0.007)
Family size	-0.028***	-0.018***	-0.081**	-0.011***
	(0.007)	(0.006)	(0.033)	(0.004)
Weak-Identification Tests				
Kleibergen-Paap Wald rk F-stat	814.61	814.61	421.51	932.80
P-value	0.00	0.00	0.00	0.00
Weak-Instrument-Robust – Inference				
Anderson-Rubin F	11.77	11.77	0.02	5.86
P-value	0.00	0.00	0.88	0.016
Stock-Wright S stat	11.36	11.36	0.02	5.79
P-value	0.00	0.00	0.88	0.016
Children's control	yes	yes	yes	yes
Parents' controls	yes	yes	yes	yes
District fixed-effect	yes	yes	yes	yes
R-square	0.18	0.09	0.14	0.14
N	393597	393597	393597	345985

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age, square, gender, birth order, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of 0-20 years old children in the family at the time of the survey.

Table 6: OLS and 2SLS Estimates of the Effects of Family Size on Education by Caste and Residence

Dependent variables	Instrument: First child is a girl									
	Low caste		Middle caste		High caste		Rural		Urban	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Literate	-0.021***	-0.046***	-0.019***	-0.021*	-0.018***	-0.004	-0.020***	-0.030***	-0.018***	-0.023
	(0.002)	(0.012)	(0.001)	(0.011)	(0.002)	(0.013)	(0.002)	(0.007)	(0.002)	(0.017)
R-square	0.25	0.16	0.25	0.19	0.23	0.17	0.25	0.18	0.25	0.19
N	161380	161380	153015	153015	79202	79202	321140	321140	72457	72457
Ever in school	-0.019***	-0.036***	-0.017***	-0.012	-0.016***	0.007	-0.018***	-0.018**	-0.016***	-0.021
	(0.002)	(0.010)	(0.001)	(0.009)	(0.002)	(0.010)	(0.002)	(0.007)	(0.002)	(0.014)
R-square	0.16	0.074	0.14	0.088	0.14	0.073	0.14	0.084	0.16	0.097
N	161380	161380	153015	153015	79202	79202	321140	321140	72457	72457
Years of schooling	-0.181***	-0.162**	-0.201***	-0.089	-0.210***	0.093	-0.197***	-0.107**	-0.197***	-0.046
	(0.009)	(0.054)	(0.008)	(0.050)	(0.013)	(0.064)	(0.006)	(0.035)	(0.013)	(0.086)
R-square	0.67	0.637	0.71	0.683	0.78	0.758	0.69	0.663	0.78	0.765
N	161380	161380	153015	153015	79202	79202	321140	321140	72457	72457
Currently enrolled	-0.011***	-0.005	-0.015***	-0.019***	-0.016***	-0.008	-0.014***	-0.010**	-0.016***	-0.026**
	(0.001)	(0.006)	(0.001)	(0.005)	(0.001)	(0.007)	(0.0007)	(0.004)	(0.001)	(0.009)
R-square	0.16	0.140	0.16	0.147	0.15	0.131	0.16	(0.004)	0.15	0.131
N	138272	138272	135014	135014	72699	72699	279847	279847	66138	66138
Children's control	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Parents' controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
District F.E.	yes	yes	yes	Yes	yes	yes	yes	yes	yes	yes

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age, square, gender, birth order, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of 0-20 years old children in the family at the time of the survey. Low caste is scheduled caste (SC) and scheduled tribe(ST) households while middle caste is other backward caste (OBC) category. Poor is households in bottom two quintiles based wealth index constructed from assets, amenitites and durables.

Table 7: OLS and 2SLS Estimates of the Effects of Family Size on Education by Household Wealth and Mother's Education

Dependent variables	Instrument: First child is a girl											
	Household wealth						Mother's education					
	Bottom two quintile		Third quintile		Top quintile		Illiterate		Less than primary		Primary & above	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Literate	-0.024***	-0.052***	-0.016***	-0.017*	-0.007***	-0.007	-0.026***	-0.046***	-0.008***	-0.027*	-0.005***	-0.0004
	(0.002)	(0.012)	(0.001)	(0.008)	(0.002)	(0.013)	(0.001)	(0.010)	(0.002)	(0.012)	(0.001)	(0.011)
R-square	0.23	0.15	0.23	0.20	0.21	0.08	0.23	0.16	0.25	0.20	0.23	0.19
N	191211	191211	154262	154262	48124	48124	227697	227697	63815	63815	102085	102085
Ever in school	-0.023***	-0.040***	-0.013***	-0.009	-0.003***	0.003	-0.024***	-0.039***	-0.005**	-0.005	-0.003***	0.006
	(0.002)	(0.011)	(0.001)	(0.006)	(0.0009)	(0.007)	(0.001)	(0.009)	(0.001)	(0.008)	(0.0008)	(0.006)
R-square	0.14	0.08	0.10	0.07	0.05	0.03	0.14	0.08	0.09	0.07	0.08	0.05
N	191170	191170	154220	154220	48120	48120	227638	227638	63803	63803	102069	102069
Years of schooling	-0.177***	-0.261***	-0.177***	-0.004	-0.098***	0.113	-0.205***	-0.295***	-0.134***	0.104*	-0.102***	0.107**
	(0.008)	(0.052)	(0.008)	(0.039)	(0.012)	(0.068)	(0.008)	(0.046)	(0.009)	(0.057)	(0.007)	(0.050)
R-square	0.60	0.56	0.77	0.76	0.88	0.87	0.63	0.58	0.81)	0.78	0.87	0.86
N	191211	191211	154262	154262	48124	48124	227697	227697	63815	63815	102085	102085
Currently enrolled	-0.013***	-0.018***	-0.013***	-0.007	-0.008***	-0.013**	-0.015***	-0.018***	-0.011***	-0.020***	-0.008***	-0.005
	(0.0009)	(0.006)	(0.0009)	(0.005)	(0.001)	(0.006)	(0.0008)	(0.005)	(0.001)	(0.007)	(0.0008)	(0.004)
R-square	0.18	0.16	0.16	0.14	0.08	0.06	0.18	0.17	0.15	0.13	0.08	0.06
N	157962	157962	142067	142067	46226	46226	189169	189169	59323	59323	97493	97493
Children's control	yes	Yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Parents' controls	yes	Yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
District F.E.	yes	Yes	yes	yes	Yes	yes	yes	yes	yes	yes	yes	yes

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard error, clustered by district, are shown in parentheses. Children's controls include age, age, square, gender, birth order, religion, caste, SES and rural dummies. Parent controls include age, age square, and education levels of father and mother. Family size is total number of 0-20 years old children in the family at the time of the survey. Low caste is scheduled caste (SC) and scheduled tribe(ST) households while middle caste is other backward caste (OBC) category. Poor is households in bottom two quintiles based wealth index constructed from assets, amenitites and durables.

Table 8: OLS and 2SLS Estimates of the impact of Family Size on Children's Health

	Weight (gram)	Height (cm)	Weight-for-age z-score	Height-for-age z-score	Weight-for-height z-score	Underweight (waz <-2)	Stunting (haz <-2)	Wasting (WfH <-2)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: OLS Results								
Family size	-177.0***	-0.755***	-0.114***	-0.182***	-0.00350	0.0371***	0.0483***	0.00962
	(46.63)	(0.158)	(0.0295)	(0.0344)	(0.0270)	(0.0132)	(0.0119)	(0.00915)
Panel B: IV Results								
							F test of excluded instruments: F(1, 28) = 5.73	
First stage				**0.026****			Prob > F = 0.0236	
				(0.010)				
Family size	450.8	7.364	-1.499	-0.858	-0.639	-0.0286	0.00435	0.143
	(1418)	(6.658)	(1.162)	(1.253)	(1.252)	(0.417)	(0.306)	(0.323)
N	10107	10113	10136	10136	10136	10136	10136	10136

Notes: Family size is the number of 0-59 months old children in the family at the time of the survey. All models include child's age, birth order, birth size, gender, religion, caste of the household, rural dummy, mother's education, father's education, mother's age, father's age, socio-economic status of the household and state fixed-effects. Standard errors clustered by state are reported in parentheses. Data source: NFHS