

Preference for Boys, Family Size and Educational Attainment in India*

Adriana Kugler¹

Santosh Kumar²

December 2013

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 arising from 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.

¹ McCourt School of Public Policy, Georgetown University, Email: ak659@georgetown.edu.

² Department of Economics & International Business, Sam Houston State University, Email: skumar@shsu.edu

*We greatly thank Joshua Angrist, Gary Solon and seminar participants at 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 as well as it is 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-constrained households may face a quantity-quality trade-off in terms of their child-rearing decisions (Becker and Lewis 1973).³

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 interesting because Q-Q trade-offs likely affect additional margins of education and are likely to be stronger in resource-constrained households in developing countries. In addition, we are able to exploit the preference for male child 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

³ 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 health services. In 2006, 37% of adults and 19% of young people were illiterate and 31% of children had dropped out of school by the 5th grade. 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.⁴ 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 challenging because fertility decisions and investments in children are jointly determined and depend on 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 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.

⁴ These statistics on educational and health outcomes in India come from the World Development Indicators publication of the World Bank.

To address this issue, we employ Instrumental 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 charged with the care for elder parents and also have better opportunities to do well in the labor market, 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, in terms of primary and secondary school enrollment and completion. Also, there is little evidence that households who have a first-born who is a 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 selective abortions are likely to reduce the chances of having a first-born who is a girl. Furthermore, 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 third round of District Level Household Survey (DLHS-3) implemented in 2007-08 to examine the impact on educational outcomes and the National Family Health Survey (NFHS-3, 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, school progression and years of schooling. While our regressions control for a large number of children and parental characteristics, there could be unobservable factors related to both household size and children's investments. To address the endogeneity of family size we instrument family size by gender of the first born. First-stage results show that having a first-born who is a female is strongly positively correlated with family size and Anderson-Rubin and Stock-Wright tests of weak instruments are all rejected. The IV results show that an extra child in the family reduces schooling by 0.1 years and reduces the probability of ever attending or being enrolled in school by between 2 and 1 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. This may happen due to poor and dysfunctional public education system in India, but 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, the empirical strategy for the Q-Q trade-off analysis is presented. Section 4 centers on description of data. In Sections 5 and 6, we present average and heterogenous effects of family size on education results. Section 7 discusses the effects of family size on health, and the conclusion is 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, forced sterilization in India), 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 et al. (2005) use twins as an instrument for family size using Norwegian data and found no evidence that family size affects educational attainment of children in Norway, after controlling for birth order and other potential confounding variables. Similarly, Angrist, Lavy, & Schlossar (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 Public Use Microdata Sample from US and gender composition of first two child 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).⁵

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

⁵ 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.

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. 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 & 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 & Zhang (2009), who also rely on *twins* as an exogenous source of increase 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 identify 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 20th and 85th percentile. Using twinning as an instrument, Ponczek and Souza (2011) also report negative effects on educational outcomes in Brazil, while Agüero and Marks (2008) find evidence of Q-Q trade-off only for health indicators (weight-for-age and breastfeeding practice) but not for education.⁶ 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 & multiple births found a positive but insignificant relationship between family size and weight-to-height ratio.⁷

⁶ 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.

⁷ Our paper differs from Sarin (2004) as we include educational attainment as outcomes in addition to

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 % 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, we are the first study to focus on India since the original twinning study of the 1980s 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 scant literature on the effect of family size on child health in a Q-Q framework.

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 investigate the effect of family size on children's educational and health outcomes by using the sex of the first birth (SFB). We first estimate the following OLS and 2SLS regressions,

$$EDU = \beta_0 + \beta_1 * FamilySize_i + \beta_2 X_i + \mu_d + \varepsilon_i \quad (1)$$

where EDU is the educational attainment of the child as measured by probability of being literate, probability of ever attended school, probability of currently being enrolled, grade progression, years of schooling, and probability of primary school completion. We also estimate similar regressions for several child health measures.⁸ The variable *FamilySize* is the number of children under 21 years of age in the family; X is a vector of covariates, and ε 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 coefficient on family size provides the extent to which there is a Q-Q trade-off.

Since fertility behavior and decisions about children's education are likely jointly determined, the OLS estimates of equation (1) are subject to endogeneity bias and are unlikely to capture causal effects. The OLS estimates may be downwardly or upwardly biased depending on nature of endogeneity. For example, in a country such as 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

⁸ In addition, we also look at health outcomes. In particular, we look at height, weight, height-for-age, weight-for-age, weight-for-height, and whether a person is underweight, stunted, and wasted.

parents may have more children and also invest more in their children's education, thus generating downward biases.

To address the endogeneity of family size, we use the first child's sex as an instrument. We estimate the following two-stage least square model:

$$FamilySize = \alpha_0 + \alpha_1 * SFB_i + \alpha_2 X_i + \mu_d + \varepsilon_i \quad (2)$$

$$EDU = \alpha_0 + \alpha_1 * FamilySize_i + \alpha_2 X_i + \mu_d + \varepsilon_i \quad (3)$$

where *FamilySize* is the number of children as measured by household size, *EDU* is measured by primary school completion and completed years of schooling, *SFB* is the sex of the first birth and is the instrumental variable for *FamilySize*, and *X* denotes a vector of exogenous regressors. 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 family size 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 patriarchal 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 a daughter gets married, a family will prefer to have boys so they can receive a dowry when son 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., $\text{Corr}(\text{FamilySize}, e_i) = 0$. 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 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)).⁹ 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 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.

⁹ 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, Miller and Mouritsen 1999).

3. Data Description

For educational outcomes, we use data from the Indian District Level Household Survey (DLHS-3) 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.¹⁰ 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

¹⁰ 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.

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 different 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 addition, we examine the impact of family size on grade progression, where grade progression is measured as the ratio of the actual years of schooling over the years of schooling a child with that age should have achieved if making normal progress through school without any grade repetitions and without skipping of grades. Thus, if the ratio is one, grade progression is as expected. If the ratio is less than 1, then the person has lagged behind due to grade repetition; if the ratio is greater than 1, then the person must have progressed through schooling faster than normal by skipping grades.

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 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.¹¹

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 third round of National Family Health Surveys (NFHS-3) which was implemented in 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

¹¹ 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.

of the survey (Aguero and Marks, 2008). In particular the survey collects 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-effect, we include state fixed-effect as 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. A z-score of one indicates that child is one standard deviation below the median of reference population.¹² 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.

¹² This is prefixed by a positive sign (+) or a negative sign (-) depending on whether the child's actual weight is more than the median weight or less than the median weight.

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 exogenous. 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.

4. 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; (d) grade progression; and (e) years of schooling for those who ever attended school. Except for grade progression and years of completed schooling, all other variables are binary and are coded as 1 or 0. The main independent variable, household size, 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 and grade progression 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. Finally, an extra child in the family reduces grade progression by 0.011.

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. Column 1 of Table 5 shows that family size increases by 0.33 children when the first-born is a girl and the effect is significant at 1% level of significance.¹³ We also estimate the first-stage for various sub-samples (unbalanced vs. balanced sex ratio, and Hindu vs. Muslim and Christian) and the relation remains always positive and statistically significant.

The IV results presented in Table 6 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 the probability of ever attending schooling and being currently enrolled in

¹³ 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.

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 6 is 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 3.3% when relying on OLS estimates. The IV estimate on grade progression is positive, suggesting a bias towards a more negative effect in the OLS estimates. The coefficient for grade progression is 0.017. Since the average grade progression is 0.95, this means that an additional child in the family would increase the grade progression by 1.8%. This is puzzling at first, but it could happen if, for example, parents convince school administrators and teachers to place children in the same classroom/school when they have many children to avoid additional commutes to school.

Once we account for the endogeneity of family size (using IVs), 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 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 6 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 6 to assess that our second-stage results are robust to weak-instrument inference.

5. Heterogeneity in the Quantity-Quality Trade-off

5.1 . 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 7 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 5.6% among low caste children, by 2.5% 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 9.3% compared to 2.8% on average in the previous section. The effects of family size on current enrollment are greatest on middle caste children. By contrast, the impact of family size on grade progression is positive but largest on high caste families, perhaps because resources can be pooled, and older children who are successful in school can help younger siblings in school.

5.2. 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 having larger family size is mostly larger and statistically significant in rural compared to rural areas, suggesting the quantity-quality trade-off is more pronounced in rural India. The last 4 columns of Table 7 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, on years of schooling, and on grade progression 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, -0.107 for years of schooling and 0.018 for grade progression in rural areas (it is zero in urban areas). These coefficients imply that an extra child reduces literacy by 3.6%, reduces the likelihood of ever attending school by 2% and years of schooling by 3.5% 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.

Interestingly, bigger households increase grade progression in rural areas, perhaps because parents try to get their children into the same grades/schools to simplify their schedules and make them compatible with their work in agriculture. By contrast, family size affects the likelihood of current enrollment for both rural and urban households.

5.3. 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 8 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 6.3%, the likelihood of attending school and being currently enrolled by 45% and 50%, and years of schooling by 8.7% among households at the bottom quintiles of the wealth distribution. By contrast, there are no effects on grade progression when dividing by wealth quintile.

5.4. 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 8 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 4.8% 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.

5.5 Robustness Checks

The validity of our instrument rests on the assumption that the sex of the first-birth is a random event. In the previous section, we presented evidence in support of this key identifying assumption. Several studies have shown that sex-selective abortion is not prevalent at the first-parity in India (Bhalotra and Cochrane (2010), Ebenstein (2007), Jha et al. (2011), Portner (2010), and Rosenblum (2010)). These studies find that the sex-ratio at first birth lies within the biologically range of 1.03-1.07. Nevertheless, to deal with the potential concerns for selective abortions we exploit the documented variation in sex-ratio across states. It has been clearly documented that Southern Indian states have balanced sex-ratio (in the biological range of 1.03-1.07), while northern states have poor sex-ratio (fewer females per 1000 males). Census data from 2011 also documents that Muslims and Christians have more balanced sex-ratio than Hindus. In Table 9 & 10, we present the

heterogenous effects of family size on children's education by religion (Hindu vs Muslim/Christian) and for southern states with balanced sex-ratio. Balanced sex-ratio states are Karnataka, Tamil Nadu, Kerala, Andhra Pradesh, Odisha, Chattisgarh, Assam, Himachal Pradesh, Uttarakhand, Manipur, Meghalaya, Mizoram and Tripura. All of these states have more than 950 females per 1000 males. Table 9 presents the results by caste categories and by rural-urban. Results indicate that children's quality is negatively impacted by the family size even in these southern states which are potentially less affected by selective abortions. The trade-offs are more pronounced among low caste and rural households.

Table 10 presents the results by wealth quintiles and mother's education level. Results are qualitatively similar for these households as well. Households belonging to the lower wealth quintile and headed by illiterate mothers face greater trade-offs compared to richer and more educated households.

6. Effect of Family Size on Nutrition

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) discusses 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 Integrated Child Development Scheme (ICDS) and 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 a high fertility level 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 11 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 11 shows the estimates from OLS, while Panel B shows the results from 2SLS analysis. 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 11 show that children born in larger families are also more likely to be underweight and stunted, respectively.

Panel B shows the corresponding IV estimates. The IV analysis fails to detect any significant evidence of Q-Q trade-off in India. 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 explanation 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.

7. 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.

Our findings are also supportive of theoretical work by Galor and Moav (2002), who were the first to argue that the quantity-quality trade-off was decisive to economic advancement, not just from the onset of the demographic transition, but throughout human history.

Correctly estimating the causal effect of family size on child-quality outcomes is important for a developing country's public policy perspective. The majority of large families are poor, and our results suggest that family size has a direct impact on important outcomes for children. This discussion can better inform the public debate about how to understand and address poverty, education, and child labor in developing countries.

Our results suggest that policymakers in developing countries should invest more in education in areas and households for whom the trade-off is severe in order to mitigate the adverse impacts of larger family size.

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 Souza. 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 (3.45)	9.41 (3.34)	9.79 (3.55)
Gender of first child (female=1)	0.49 (0.49)		
Literate	0.82 (0.38)	0.81 (0.39)	0.83 (0.37)
Ever attended school	0.9 (0.30)	0.89 (0.31)	0.91 (0.29)
Still enrolled	0.95 (0.21)	0.95 (0.21)	0.95 (0.22)
Grade progression	0.95 (0.49)	0.96 (0.49)	0.94 (0.48)
Years of schooling	3.08 (2.92)	2.94 (2.85)	3.22 (2.98)
Mother's age	30.94 (3.36)	30.88 (3.34)	31.00 (3.37)
Father's age	36.48 (4.81)	36.42 (4.79)	36.54 (4.82)
Mother's years of schooling	2.99 (4.06)	3.05 (4.09)	2.93 (4.03)
Father's years of schooling	5.48 (4.74)	5.56 (4.76)	5.40 (4.72)
Family size	3.54 (1.33)	3.70 (1.33)	3.40 (1.31)
Rural	0.82 (0.39)	0.81 (0.39)	0.82 (0.39)
Low caste (SC & ST)	0.41 (0.49)	0.41 (0.49)	0.41 (0.49)
Middle caste (OBC)	0.39 (0.49)	0.39 (0.49)	0.39 (0.49)
Low wealth	0.49 (0.50)	0.48 (0.50)	0.49 (0.50)
Medium wealth	0.39 (0.49)	0.4 (0.49)	0.39 (0.49)
No. of observations	393,597	193,263	200,334
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 (17.01)	28.31 (17.04)	28.57 (16.99)
Gender of first child (female=1)	0.51 (0.49)		
Weight (Gram)	10251.72 (3123.52)	10142.56 (3061.12)	10363.79 (3182.72)
Height (Centimeter)	81.85 (13.34)	81.62 (13.27)	82.10 (13.42)
WAZ	-1.63 (1.16)	-1.64 (1.17)	-1.63 (1.16)
HAZ	-1.47 (1.50)	-1.46 (1.52)	-1.48 (1.48)
WfH	-0.92 (1.13)	-0.92 (1.12)	-0.94 (1.13)
Child is underweight	0.39 (0.49)	0.39 (0.49)	0.39 (0.49)
Child is stunted	0.35 (0.48)	0.35 (0.48)	0.35 (0.48)
Child is wasted	0.15 (0.36)	0.15 (0.36)	0.15 (0.36)
Family size	2.17 (0.42)	2.20 (0.44)	2.16 (0.39)
Rural	0.61 (0.49)	0.62 (0.49)	0.60 (0.49)
Low caste (SC & ST)	0.33 (0.47)	0.33 (0.47)	0.32 (0.47)
Middle caste (OBC)	0.33 (0.47)	0.33 (0.47)	0.33 (0.47)
Low wealth	0.29 (0.45)	0.29 (0.45)	0.29 (0.45)
Medium wealth	0.22 (0.41)	0.22 (0.41)	0.22 (0.41)
Mother's age	24.09 (3.75)	24.14 (3.75)	24.04 (3.74)
Father's age	29.38 (4.79)	29.46 (4.85)	29.30 (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 (0.004)	-0.007 (0.011)
Low wealth	-0.006 (0.006)	-0.016 (0.016)
Medium wealth	-0.004 (0.005)	-0.009 (0.013)
Religion (Hindu=1)	0.006 (0.004)	0.016 (0.011)
Scheduled caste/tribe (Yes=1)	0.003 (0.004)	0.009 (0.011)
Other backward caste	0.002 (0.004)	0.005 (0.010)
Mother is illiterate	-0.008* (0.004)	-0.02* (0.011)
Mother is primary schooled	-0.003 (0.004)	-0.007 (0.011)
Father Illiterate	-0.009** (0.004)	-0.023** (0.010)
Father is primary schooled	-0.011*** (0.004)	-0.029*** (0.010)
Mother's age	0.037*** (0.007)	0.093*** (0.017)
Father's age	0.002 (0.003)	0.004 (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	Grade progression
	(1)	(2)	(3)	(4)	(5)
Family size	-0.020*** (0.001)	-0.018*** (0.001)	-0.20*** (0.006)	-0.014*** (0.0006)	-0.011*** (0.0009)
Children's control	yes	yes	Yes	yes	Yes
Parents' controls	yes	yes	Yes	yes	Yes
District fixed-effect	yes	yes	Yes	yes	Yes
R-square	0.25	0.14	0.71	0.15	0.3
No of observations	393597	393510	393597	345985	243411

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. Grade progression is defined as (years of schooling/age-6) and excludes children with zero years of schooling.

Table 5: First-stage by Groups

	Full sample	Number of children			
		Unbalanced sex-ratio	Balanced sex-ratio	Hindu	Muslim/ Christian
		(1)	(2)	(3)	(4)
First girl (FG)	0.329*** (0.009)	0.246*** (0.008)	0.165*** (0.011)	0.243*** (0.007)	0.140*** (0.013)
N	393597	271386	122211	303166	90428

Notes: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust standard errors, clustered by district, are shown in parentheses. All regressions include household religion, household caste, household socio-economic status, mother's age & age square, father's age & age square, mother's education, father's education, and rural dummy. District fixed-effects are included in all models. The regression sample is restricted to mother's age between 20-35 years. Balanced sex-ratio states are southern states having more than 950 females per 1000 males, while unbalanced sex-ratio states are mainly northern states having less than 950 female per 1000 males. Hindu tends to have poor sex-ratio compared to Muslims and Christians.

Table 6: IV Estimates of the Effect of Family Size on Children's Educational Outcomes

	Instrument: First child is a girl (G)				
	Literate	Ever attended school	Years of schooling	Currently enrolled	Grade progression
	(1)	(2)	(3)	(4)	(5)
First Stage	0.219*** (0.007)	0.219*** (0.007)	0.219*** (0.007)	0.228*** (0.007)	0.245*** (0.008)
Family size	-0.028*** (0.007)	-0.018*** (0.006)	-0.081** (0.033)	-0.011*** (0.004)	0.017** (0.008)
Weak-Identification Tests					
Kleibergen-Paap Wald rk F-stat	814.61	814.61	421.51	932.80	851.93
P-value	0.00	0.00	0.00	0.00	0.00
Weak-Instrument-Robust – Inference					
Anderson-Rudin F	11.77	11.77	0.02	5.86	4.29
P-value	0.00	0.00	0.88	0.016	0.03
Stock-Wright S stat	11.36	11.36	0.02	5.79	4.23
P-value	0.00	0.00	0.88	0.016	0.03
Children's control	yes	yes	yes	yes	yes
Parents' controls	yes	yes	yes	yes	yes
District fixed-effect	yes	yes	yes	yes	yes
R-square	0.18	0.09	0.14	0.14	0.23
N	393597	393597	393597	345985	243411

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. Grade progression is defined as (years of schooling/age-6) and excludes children with zero years of schooling.

Table 7: OLS and 2SLS Estimates of the Effects of Family Size on Education by Caste and Residence

Dependent variables	Instrument: First child is a girl (FG)									
	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*** (0.002)	-0.046*** (0.012)	-0.019*** (0.001)	-0.021* (0.011)	-0.018*** (0.002)	-0.004 (0.013)	-0.020*** (0.002)	-0.030*** (0.007)	-0.018*** (0.002)	-0.023 (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*** (0.002)	-0.036*** (0.010)	-0.017*** (0.001)	-0.012 (0.009)	-0.016*** (0.002)	0.007 (0.010)	-0.018*** (0.002)	-0.018** (0.007)	-0.016*** (0.002)	-0.021 (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*** (0.009)	-0.162** (0.054)	-0.201*** (0.008)	-0.089 (0.050)	-0.210*** (0.013)	0.093 (0.064)	-0.197*** (0.006)	-0.107** (0.035)	-0.197*** (0.013)	-0.046 (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*** (0.001)	-0.005 (0.006)	-0.015*** (0.001)	-0.019*** (0.005)	-0.016*** (0.001)	-0.008 (0.007)	-0.014*** (0.0007)	-0.010** (0.004)	-0.016*** (0.001)	-0.026** (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
Grade progression	-0.009*** (0.002)	0.006 (0.013)	-0.011*** (0.002)	0.018 (0.012)	-0.012*** (0.002)	0.033* (0.016)	-0.011*** (0.001)	0.018* (0.008)	-0.009*** (0.002)	0.003 (0.023)
R-square	0.32	0.247	0.30	0.220	0.28	0.207	0.30	0.234	0.28	0.211
N	95732	95732	94793	94793	52886	52886	195436	195436	47975	47975
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, amenities and durables.

Table 8: 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 (G)											
	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***	-0.402***	-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	191211	191211	154262	154262	48124	48124	227697	227697	63815	63815	102085	102085
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
Grade progression	-0.007***	0.012	-0.014***	0.020	-0.0104***	0.018	-0.010***	0.001	-0.009***	0.055**	-0.008***	0.013
	(0.001)	(0.012)	(0.002)	(0.12)	(0.003)	(0.022)	(0.001)	(0.010)	(0.002)	(0.017)	(0.002)	(0.018)
R-square	0.31	0.24	0.29	0.23	0.26	0.20	0.30	0.23	0.31	0.21	0.27	0.21
N	104348	104348	103355	103355	25705	35705	128803	128803	42981	42981	71627	71627
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, amenities and durables.

Table 9: OLS and 2SLS Estimates of the Effects of Family Size on Education by Caste and Residence: Balanced Sex-Ratio Sample

Dependent variables	Instrument: First child is a girl (FG)									
	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	-0.012*** 0.002	-0.050* (0.022)	-0.013*** 0.002	-0.033 (0.020)	-0.015*** 0.003	-0.028 (0.025)	-0.015*** 0.002	-0.048** (0.015)	-0.005** 0.003	0.011 (0.032)
R-square	0.22	0.13	0.22	0.17	0.15	0.11	0.21	0.13	0.22	0.19
N	58371	58371	40043	40043	23797	23797	100517	100517	21694	21694
Ever in school	-0.009*** 0.003	-0.033 (0.017)	-0.010*** 0.002	-0.029 (0.018)	-0.012*** 0.002	-0.017 (0.019)	-0.011*** 0.002	-0.030* (0.012)	-0.003* 0.002	-0.024 (0.021)
R-square	0.17	0.06	0.09	0.05	0.08	0.04	0.15	0.05	0.09	0.05
N	58321	58321	40026	40026	23781	23781	100436	100436	21692	21692
Years of schooling	-0.173*** 0.013	-0.284* (0.115)	-0.171*** 0.015	0.034 (0.115)	-0.173*** 0.022	0.280 (0.162)	-0.192*** 0.011	-0.097 (0.077)	-0.131*** 0.018	-0.078 (0.154)
R-square	0.69	0.67	0.79	0.78	0.81	0.78	0.73	0.71	0.83	0.82
N	58371	58371	40043	40043	23797	23797	100517	100517	21694	21694
Currently enrolled	-0.009*** 0.001	-0.026* (0.013)	-0.019*** 0.002	-0.031* (0.015)	-0.012*** 0.002	0.016 (0.018)	-0.012*** 0.001	-0.017 (0.010)	-0.014*** 0.002	-0.037 (0.022)
R-square	0.17	0.15	0.18	0.16	0.15	0.108	0.17	0.15	0.15	0.12
N	50814	50814	36890	36890	22288	22288	89669	89669	20323	20323
Grade progression	-0.011*** 0.002	0.023 (0.029)	-0.005* 0.003	0.023 (0.027)	-0.012*** 0.004	0.029 (0.037)	-0.011*** 0.002	0.020 (0.019)	-0.006* 0.003	0.034 (0.047)
R-square	0.33	0.25	0.30	0.23	0.30	0.24	0.32	0.25	0.28	0.21
N	34964	34964	27486	27486	16243	16243	63675	63675	15018	15018
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 fixed-effect	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, amenities and durables. High sex-ratio states are Karnataka, Tamil Nadu, Kerala, Andhra Pradesh, Odisha, Chattisgarh, Assam, Himachal Pradesh, Uttarakhand, Manipur, Meghalaya, Mizoram, Tripura. Hindu tends to have higher sex-ratio than muslim and christian.

Table 10: OLS and 2SLS Estimates of the Effects of Family Size on Education by Household Wealth and Mother's Education: Balanced Sex-Ratio Sample

Dependent variables	Instrument: First child is a girl (G)											
	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	-0.019***	-0.118***	-0.007***	0.008	-0.004**	0.018	-0.022***	-0.077***	-0.002	-0.041	-0.002	0.023
	0.002	(0.029)	0.001	(0.014)	0.002	(0.028)	0.003	(0.021)	0.001	(0.023)	0.001	(0.021)
R-square	0.21	0.05	0.20	0.17	0.21	0.16	0.20	0.11	0.20	0.15	0.21	0.17
N	51056	51056	56403	56403	14752	14752	55409	55409	24495	24495	42307	42307
Ever in school	-0.016***	-0.072**	-0.005***	-0.002	-0.001	-0.016	-0.018***	-0.064***	-0.0008	-0.015	-0.0002	0.019
	0.002	(0.023)	0.001	(0.010)	0.001	(0.015)	0.002	(0.018)	0.001	(0.017)	0.001	(0.013)
R-square	0.15	0.03	0.06	0.04	0.05	0.02	0.15	0.03	0.09	0.05	0.07	0.04
N	51018	51018	56361	56361	14749	14749	55353	55353	24483	24483	42292	42292
Years of schooling	-0.187***	-0.371**	-0.151***	0.198*	-0.048*	-0.055	-0.216***	-0.411***	-0.147***	0.162	-0.105***	0.242**
	0.016	(0.132)	0.011	(0.079)	0.025	(0.125)	0.015	(0.116)	0.015	(0.130)	0.012	(0.092)
R-square	0.64	0.60	0.81	0.79	0.90	0.90	0.66	0.61	0.83	0.80	0.89	0.87
N	51056	51056	56403	56403	14752	14752	55409	55409	24495	24495	42307	42307
Currently enrolled	-0.014***	-0.033	-0.010***	-0.008	-0.002	-0.007	-0.017***	-0.024	-0.010***	-0.029	-0.007***	-0.007
	0.002	(0.017)	0.001	(0.011)	-0.0013	(0.011)	0.002	(0.015)	0.002	(0.018)	0.001	0.007
R-square	0.20	0.19	0.16	0.14	0.06	0.05	0.20	0.18	0.16	0.14	0.08	0.07
N	43222	43222	52636	52636	14134	14134	46933	46933	22779	22779	40280	40280
Grade progression	-0.011***	0.012	-0.009	0.051*	-0.002	-0.008	-0.010***	0.007	-0.006*	0.075*	-0.008***	0.024
	0.002	(0.029)	0.002	(0.024)	0.005	(0.046)	0.002	(0.025)	0.003	(0.038)	0.003	(0.035)
R-square	0.33	0.24	0.31	0.24	0.25	0.20	0.31	0.24	0.35	0.22	0.29	0.23
N	28948	28948	38792	38792	10951	10951	32659	32659	16566	16566	29468	29468
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, amenities and durables. High sex-ratio states are Karnataka, Tamil Nadu, Kerala, Andhra Pradesh, Odisha, Chattisgarh, Assam, Himachal Pradesh, Uttarakhand, Manipur, Meghalaya, Mizoram, Tripura. Hindu tends to have higher sex-ratio than muslim and christian.

Table 11 : OLS and 2SLS Estimates of the impact of Family Size on Child's Health Outcomes

	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*** (46.63)	-0.755*** (0.158)	-0.114*** (0.0295)	-0.182*** (0.0344)	-0.00350 (0.0270)	0.0371*** (0.0132)	0.0483*** (0.0119)	0.00962 (0.00915)
Panel B: IV Results								
First stage			0.026** (0.010)					
Family size	450.8 (1418)	7.364 (6.658)	-1.499 (1.162)	-0.858 (1.253)	-0.639 (1.252)	-0.0286 (0.417)	0.00435 (0.306)	0.143 (0.323)
N	10107	10113	10136	10136	10136	10136	10136	10136

F test of excluded instruments:
 $F(1, 28) = 5.73$
 Prob > F = 0.0236

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