CHILD GENDER AND PARENTAL INVESTMENTS IN INDIA: ARE BOYS AND GIRLS TREATED DIFFERENTLY?*

Silvia Helena Barcellos (Rand)

Leandro Carvalho (Rand)

Adriana Lleras-Muney (UCLA and NBER)

June 30, 2010

Abstract: There is considerable debate in the literature as to whether boys and girls are treated differently in India. But son-biased stopping rules imply that previous estimates are likely to be biased. We propose a novel identification strategy to properly identify the effects of child gender on parental investments. Using data from a time use survey we document gender differences in childcare time which have not been studied before in developing countries. We find that boys receive on average 10% more time than girls. They are also more likely to be breastfed for longer, given vaccinations and vitamin supplementation.

I. Introduction

Throughout the Western world, boys have higher mortality rates than girls (United Nations Secretariat 1988). This pattern is, however, reversed in a number of South and East Asian countries, in which mortality rates are higher for girls than for boys (Chen, Huq and D'Souza 1981; Arnold, Choe and Roy 1998). As a result Sen (1990) computes there are approximately 100 million "missing women" in these countries, where it has also been documented that

* silvia@rand.org; carvalho@rand.org; alleras@econ.ucla.edu.We are grateful to seminar participants at Northwestern University, RAND, UC Irvine, UCLA, UIC, and UIUC for their suggestions and especially to Anne Case, Taryn Dinkelman, Henry Farber, Bo Honoré, Seema Jayachandran, Robert Jensen, David Lee, Christina Paxson and Sam Schulhofer-Wohl for their comments.

families prefer having sons to having daughters (Pande and Astone 2007). The proximate causes for this imbalance are hotly debated and include the differential treatment of girls in terms of different parental investments (Khanna et al 2003, Jensen 2005, Jayachandran and Kuziemko 2009).

There is, however, considerable debate in the literature as to whether boys and girls are indeed treated differently and thus die at greater rates. While some papers find evidence that boys receive more nutrition (Chen, Huq and D'Souza, 1981; Das Gupta, 1987), more healthcare (Basu 1989, Ganatra and Hirve, 1994) and are more likely to be vaccinated (Borooah 2004) than girls, other papers find no evidence of differential treatment. For example review studies by Sommerfelt and Arnold (1998) and Marcoux (2002) find no differences in anthropometric measures, and Deaton (2003) reports that vaccination rates are identical for boys and girls in India. Most notably Deaton (1997) reviews studies that use the adult goods method and states that there is no evidence that parents spend more on boys than on girls. Duflo (2005) concludes that"[e]ven in the countries where the preference for boys is strongest, it is hard to find evidence that girls receive less care than boys under normal circumstances."[1]

Previous work however assumes that boys and girls live in families with similar characteristics, both in terms of observables and unobservables. But this assumption is incorrect if families have a preference for sons and follow male-biased stopping rules of childbearing, which appears to be the case in India.[2] As a consequence, empirical estimates of discrimination are biased.

In this paper we develop a novel empirical strategy that addresses these issues. Our empirical strategy relies on the observation that—in the absence of sex-selective abortion—the child sex is random at birth. If the child sex is random, then families that just had a boy are identical to families that just had a girl in terms of predetermined characteristics. Therefore, any differences we observe in terms of parental inputs can be attributed to the sex of the newborn. However, over time a correlation will develop between the youngest child's gender and family characteristics, because the families that had a girl are less likely to stop having children. To overcome this problem, we restrict our sample to families with children that are still "young

[1] There is evidence however that households do discriminate in bad times (Rose 2000, Miguel 2005).
[2] For example families with fewer boys have shorter birth intervals, are more likely to want more children and to continue having children, and are less likely to use contraception (see Clark 2000 for a review).

enough" and haven't had the opportunity to have other children. The data suggest that families with boys and girls zero to 12 months of age (and possibly a bit older) look identical in terms of observables—we use them to study whether boys are given more inputs than girls.

The second contribution of this paper is to use our identification strategy to investigate whether boys and girls are treated differently in terms of an important but not frequently studied type of investment in children: *childcare time*. Starting with Becker (1991), economists have recognized the importance of time as an input into the "child production function." In Becker's framework, the "quality" of children is increasing in two kinds of parental investments: money and time. Time is particularly important to the extent that it is complementary to many other inputs that go into the child production function. For example, feeding children requires both the food to be cooked as well as the time to cook and feed them. However no estimates of gender differences in parental time allocation exist for developing countries.[3] Using data from the Indian Time Use Survey which collected information on how individuals allocate their time, we investigate whether families spend more time on childcare when the baby is a boy instead of a girl. Using the Indian Demographic and Health Surveys we also study other measures of parental investments that have been studied in the literature before, such as vaccinations.

The results indicate that families treat boys and girls differently. Households with a boy under two spend roughly 20 minutes more per day—or 10% more time—with childcare than households with a girl under two . This difference appears for different kinds of childcare, including supervision and physical care. The effect is larger for households with only one child under the age of six, who spend more than 50 minutes more per day (about 30% more) when their youngest is a boy. We also find suggestive evidence that the quality of childcare given to boys is higher. In addition our results show that boys are more likely to be vaccinated, to be breastfed longer and to be given vitamin supplementation. In general we find these inputs to be about 10% higher for boys. We do not however find evidence that boys fare better than girls in terms of anthropometric measures, outcomes rather than input measures. We discuss various explanations for this one anomalous result.

[3] Using time-use data from the US, Yeung et al (2001), Lundberg et al (2007), Mammen (2009) found suggestive evidence that boys receive more parental time than girls. Analyzing data from rural India, Rose (2000) reports that women work fewer days after the birth of a boy relative to a girl.

Our approach has several limitations. Like previous work we cannot address the issues of sex-selective abortion and differential mortality. These behaviors will likely bias our estimates of boy-girl differences towards 0, so our effects can be taken to be lower bounds. Our results suggest that the bias associated with sex-selective abortion is small. On the other hand, when we compute bounds to account for excess girl mortality, we find that our estimates are possibly underestimated by a large amount. Another limitation of our results is that we can only study children who are under the age of two. This is an important subset of the population because investments at this age have large returns in the short and long run.[4] However we cannot study discrimination at older ages.

It is difficult to estimate the impact of these inputs into outcomes. A back of the envelope calculation using estimates from the literature of the effects of breastfeeding and vaccinations on mortality suggests that differences in investments explain about 27% of excess girl mortality among children 12 to 36 months of age (this translates into 2 additional girl deaths per 1,000 children). We know of no good estimates of how parental time affects outcomes—but if we assume a modest effect of time inputs on mortality then we can explain an additional 3.4% of girl excess mortality.

Finally we cannot provide any evidence on why parents give girls fewer resources. There are several reasons why they might do so. They might prefer boys to girls; investments in boys might have larger returns (e.g., men have higher wage rates than women); boys might be seen as needing more resources (though given that girls die at larger rates this seems unlikely); and finally, families that have girls might anticipate that eventually their family size will be larger. Though we cannot distinguish between all of these explanation, we provide suggestive evidence that boys do not in fact "need" more than girls: if we look at South Africa, a developing country with data on investments and no evidence of son preference, we find that there are no systematic gender differences in the time allocation of parents or in breastfeeding and vaccination rates.

II. Identification issues in the presence of son-biased stopping rules

[4] A recent literature documents that conditions in the first years of life have large long term consequences on education, earnings and health, for example see Case, Fertig and Paxson (2005), van der Berg, Lindeboom and Portrait (2006) or Oreopolous et al (2008).

In this section, we present a simple model of a son-biased stopping rule and extend previous results by Yamaguchi (1989) and Jensen (2005) to show that—if families follow this rule—the estimates in the boy-girl discrimination literature, which assume that boys and girls live in families with similar characteristics, are biased. We then proceed to propose a method to overcome the problems that arise in this context.

We begin by presenting suggestive evidence that families in rural India follow son-biased stopping rules. Families follow son-biased stopping rules if *ceteris paribus* the probability of a family stopping having children after a boy is higher than after a girl. One implication of this behavior is that the probability of a family's youngest child being a boy is an increasing function of the age of the youngest child. At birth the sex ratio is determined by biological odds. However, as the youngest child ages, the sex ratio is increasingly determined by the stopping rule: since families are more likely to stop having children after a boy, the sex ratio will be skewed towards boys. This prediction is consistent with the data. In Figure 1, we use data from the Indian Demographic and Health Surveys (hereafter DHS) [5] to plot the fraction of boys by age for different groups of children. The figure shows that the fraction of boys among all living children is somewhat constant across ages. However, the fraction of boys is increasing in age if we restrict the sample to living children who are the youngest in their families: it increases from 51% for children 0-5 months old to 57% for children 54 to 59 months old. This is a large deviation from the natural sex ratio at birth. This evidence suggests that in India families do indeed follow son-biased stopping rules.[6] We now present a simple son-biased stopping rule model and use it to illustrate how these stopping rules bias estimates of discrimination.

We use the framework from Yamaguchi (1989). There is a continuum of families with heterogeneous preferences for sons and they all follow the same stopping rule. Family h wants to have a specific number of sons S_h, and it continues to have children until it reaches its desired number of sons. The total number of children of family h is given by N_h. The probability of a newborn being a boy p is assumed to be constant across families.

[5] We describe these data in more detail below.
[6] This pattern could also be driven by excess girl mortality. To gauge its importance, we compute the fraction of boys among all youngest children (including those that died according to the mother) and compare it to the fraction of boys among those that are alive. The graph suggests that there is excess girl mortality, since the fraction of boys is higher among the survivors. However, the extent of the bias is small relative to the effect of stopping rules. This is confirmed by the pattern that we observe among all children (rather than the youngest): the fraction of boys rises for this group but the increase is small, much smaller than what is observed among the youngest child.

In this stylized model, one can show that on average boys and girls live in families with different observed and unobserved characteristics. First, as shown by Yamaguchi (1989) (and by Jensen using a slightly different set-up[7]), girls will have on average more siblings than boys. A simple example provides intuition for the result. Take, for example, a family that wants one boy. If their first-born is a boy, then they stop having children immediately, whereas they will continue having children if their first-born is a girl. If all families behave this way then all girls have siblings but not all boys do. Simulations in Jensen (2005) suggest that the resulting differences in sib-ship size can be quite large.

Second, one can also show (Appendix 1) that if we compare children in families with the same size, on average girls are in families that desire fewer sons than the family of the average child. The intuition for this result is as follows. Suppose that we observe two families that stopped having children after their second child. Family A has a girl and a boy, and family B has two boys. Family A stopped having children in spite of the fact that they have only one boy, whereas family B stopped only because they had two boys, but would have continued otherwise. The example illustrates that for families of size 2 girls are in families that desire fewer sons than the average family. As a consequence, within families of the size, families with a larger number (share) of boys have a larger desired number of sons than families with many girls.

In the stylized model by Yamaguchi, families always obtain their desired number of sons, so we would observe their preferences by observing the number of sons they have. In reality however, since families have finite fertility rules and also imperfect fertility control, the desired number of sons is unobservable. Thus even if we control for the observed differences in family size and gender composition, there are unobserved differences in the families into which the average girl and average boy are born. If families that desire a larger number of sons invest less in girls (or more on boys) than other families, then these stopping rules imply that previous estimates of discrimination are biased, as we now discuss.

Suppose that we obtain estimates of discrimination by running a regression of some measure of child investment on a constant and a boy dummy (as in Sen and Sengupta 1983; Das Gupta 1987; Sommerfelt and Piani 1997):

[7] Jensen (2005) assumes that families will stop having children after they have a certain number of children, even if they haven't reached their desired number of boys.

$$Z_{ih} = \alpha_0 + \alpha_1 * B_{ih} + u_{ih},$$

where Z_{ih} is investment in child i in household h, B_{ih} is a dummy that is equal to 1 if child i in household h is a boy and u_{ih} is an error term. But son-biased stopping rules imply that B_{ih} is correlated with family size. Therefore, B_{ih} will be correlated with the error term and α_1 will be biased if child investment depends on the number of children in the family. The sign of the bias may be different for different measures of child investment. On the one hand, children in large families may have to share resources with more siblings (e.g., food)—this is the issue that Jensen (2005) investigates. On the other hand, children in large families may *ceteris paribus* receive more investments if there are large returns to scale to the child investment (e.g., vaccination in public campaigns, or supervision and teaching).

Given that girls tend to be in larger families than boys, it may seem reasonable then to control for family size. Suppose then that we estimate the following model instead (Oster 2009):

$$Z_{ih} = \alpha_0 + \alpha_1 * B_{ih} + X_{ih}\rho + u_{ih},$$

where now we are controlling for X_{ih}, a vector of controls that includes the number of siblings or dummies for the sex-composition of siblings. This strategy essentially compares outcomes of boys and girls in families of the same size. However, son-biased stopping rules imply that, conditional on family size, girls tend to be in families that want girls more than other families. In other words, the child's sex is not exogenous; it is correlated with parental preferences for the gender composition of children. In general the sign of the bias is unknown, and depends on the relationship between preferences for the gender composition of children and treatment of boys and girls (see Appendix 2). If for example all families invest the same in boys but families that want boys invest less in girls, then OLS estimate of α_1 is downward biased because the average girl is in a family that wants fewer boys and thus receives more child investments than she would have had she been "assigned" to a random family.

A similar argument applies to studies that use the adult goods method championed by Deaton (1989). Deaton's proposal was to estimate discrimination by observing whether families cut consumption of adults good by more when they have a baby boy compared to a baby girl. The empirical approach traditionally used in these studies is to estimate an equation like:

$$Z_h = \sum_{g=1}^{G} \beta_g \frac{M_h^g}{N_h} + \sum_{g=1}^{G-1} \varphi_g \frac{N_h^g}{N_h} + \ln N_h + X_h \delta + \varepsilon_h$$

where Z_h is the share of household expenditures devoted to a given adult good, M_h^g is the number of members of household *h* who are boys/men in age group *g*, N_h is the number of members of household *h*, N_h^g is the number of members of household *h* in age group *g*, X_h is a vector of controls that include total per capita expenditure and ε_h is an error term. There is evidence of discrimination if $\beta_g<0$—i.e., adults reduce their consumption of adult goods more when the child is a boy than when the child is a girl. However, son-biased stopping rules imply that the estimate of β_g is biased because the gender-composition of children (at any given age) is correlated with (unobserved) parental preferences, and again the sign of the bias is unclear.

III. Empirical Strategy

Our empirical strategy relies on the observation that—in the absence of sex-selective abortion—the child sex is random at birth. If the child sex is random, then families that just had a boy are identical to families that just had a girl in terms of predetermined characteristics. Therefore any differences we observe in terms of parental inputs can be attributed to the sex of the newborn. Over time, however, this is no longer true. Because families that follow a son-biased stopping rule are more likely to stop having children after a boy, in time a correlation will develop between the youngest child's sex and preferences: families with *N* children that stop after a girl tend to like girls more than families with *N* children that stop after a boy. To overcome this problem, we restrict our sample to families in which the youngest child is "young enough." We determine this using our data: we select the sample such that baby-boy and baby-girl families look identical. Formally, we estimate whether boys and girls are treated differently using the following equation:

$$Z_{ih} = \alpha_0 + \alpha_1 * B_{ih} + X_{ih}\rho + u_{ih}.$$

The OLS estimate of α_1 is an unbiased estimator of the parameter of interest if the child's sex is exogenous (conditional on X) —i.e., $Cov(b_{ih}, u_{ih}|X)=0$. Our identifying assumption is that the child's sex is exogenous at birth and for children that are young enough. In the next section, we provide evidence that predetermined characteristics (in particular number and gender of siblings) are not correlated with gender for very young children. We also show that, as the model above predicts, this no longer holds true as the family's youngest child gets older. Notice that, if gender is indeed random, then we do not need to condition on any variables.[8] Conditioning on predetermined variables should have no impact on our point estimates and should reduce the standard errors (if these variables predict parental investments).

Our assumption may fail if there is sex-selective abortion against girls or excess girl mortality. Like all the previous literature, we cannot directly address this issue. But sex-selective abortion and excess female mortality most likely bias our estimator against finding boy-girl differences. Since the surviving girls are expected to be in families that like girls more than the average family, they should receive more care than they would have in other families.[9] Thus, our estimates can be taken as lower bound estimates of gender differences in child investments. Although we cannot credibly estimate the extent of the bias generated by sex-selective abortion, we provide some estimates of the bias induced by differential mortality by making use of the complete fertility histories available in the DHS.

IV- Testing Random Assignment and Selecting the Estimation Sample

To test that the gender of the youngest child is uncorrelated with family characteristics we restrict the sample to children who are the youngest in their families and estimate the following linear equation:

$$I(boy_{ia} = 1) = X_i \beta_a + \varepsilon_{ia},$$

where the dependent variable is an indicator of whether child i in age category a is a boy, and X is a set of predetermined characteristics. Independence implies that $\beta_a = 0$, namely that X do not

[8] Some evidence suggests that the sex ratio at birth may be correlated with birth order, parental age, mother's education and marital status (see Almond and Edlund 2007 and Chahnazarian 1988 for a review). But these effects are small and can only be detected using very large samples of births (Yamaguchi 1989, Almond and Edlund 2007).

[9] One could also argue that our estimator is upwards biased for two reasons. First, the girls that survive might be healthier than boys and thus need less care than boys. However, since the mortality rates for girls remain higher than the mortality rates for boys for the entire postnatal period, it seems unlikely that surviving girls are healthier than boys. Second, due to nonrandom mating, girls might be consistently born into families with low SES relative to boys (Edlund 1999). In the next section we show evidence that this is not the case for the subsample of young children that we focus on.

jointly predict the gender of the child. The prediction is that we will not reject the null for very young children, but that we will always be able to reject it for children that are "old enough."

We perform this test using India's DHS,[10] a large representative survey that contains many variables that determined at birth. The DHS surveyed ever-married women of reproductive ages. Each woman was separately interviewed and asked questions on their characteristics and reproductive histories. The files contain full birth histories: there is a record for every child born, including date of birth and gender, whether the child has died, and whether s/he continues to live at home. Therefore, we know for every child born the characteristics of their mothers, and we can compute the number of their siblings by gender and age (including the number of those who have died) using the birth histories. To maximize sample size, we combine the 1992-1993 and 1998-1999 surveys.[11] We focus on rural households as the previous literature has done.[12] The final data contain one observation per family and include children born to women ages 15 to 49 living in rural areas in 25 states, excluding twins.[13]

To perform the test, we pool children into 12-month age-groups and run a joint test for each age-group.[14] We use the results of the test to determine the age at which the test starts to systematically reject the null. Table 1 shows all of the predetermined characteristics of the child and the mother that we can include—there are a total of 14.[15]

Figure 2 shows the results of our test graphically. It plots the p-value of the joint test that the *X*s do not predict the gender of the youngest child. The first point (at age 0) corresponds to children 0 to 11 months old. We cannot reject the null for the youngest age group. For living children, we reject the null at the 5% level for the first time for age group 6-17 months, and then again for age-group 8-19 months. For age-group 14-25 months and thereafter, we almost always

[10] This survey is also known as the National Family Health Survey.
[11] We do not use the 2005/06 survey because it does not include the survey design variables needed to correctly compute the standard errors. Although the documentation suggests that the standard errors can be computed by clustering at the state and rural level, using the 92 and 98 data we cannot reproduce the standard errors when following this procedure (see Appendix 7).
[12] There is a presumption in the literature that discrimination against girls is more prevalent in rural areas, but little evidence to support it. Subramaniam and Deaton (1991) find evidence of gender discrimination in rural but not in urban India. When explaining these results, Deaton (1997) justifies that "it can be argued that it is in the rural areas where discrimination is most likely to be found."
[13] These restrictions are made to make the 2 survey samples comparable: 1992-3 DHS does not include the state of Sikkim and includes women ages 13 and 14. We exclude twins so we can define the family's youngest child's sex.
[14] We pool children into age groups to minimize the likelihood that we reject the null of due to small sample sizes.
[15] We do not include prenatal care because this is only available for young children. To the extent that gender is known in utero, these variables could also be considered endogenous and thus might be better treated as outcomes.

reject the null. Based on these results, we construct 2 samples for estimation. Our first sample is a conservative sample ages 0-12 months. The second sample is ages 0-23 months (just under 2 years of age). This second sample is more marginal: clearly it contains some older children for whom the covariates are not balanced across genders. But we are interested in keeping children that are as old as possible in order to study as many parental investments as we can. Also this is the sample we will concentrate on for the time use survey.

Table 2 shows the results of our test for these two estimation samples. For each sample, we test whether the means for different characteristics are the same for both genders. At the bottom we report the p-value from the joint test. For the conservative sample, no differences are significant at the 5% or 10% level, and the joint test cannot reject the null. For the somewhat older sample, the number of brothers is statistically significantly different at the 5% level, and three other differences are significant at the 10% level. As a result, the joint test does not reject the null at the 5% level but does reject it at the 10% level. Clearly this sample is not as good as the younger sample. These results are consistent with the distribution of birth intervals in the data: the median birth interval (based on mother reports) is about 31 months, and the 2nd percentile is 12 months. For comparison, in the last two columns we report the result of the test for the youngest children ages 24-59 months. For this sample, despite its substantially smaller size, three out of the 14 variables we examine are statistically different at the 5% level and two others are significant at the 10% level. Interestingly we now observe that if the youngest is male he is more likely to have more sisters and fewer brothers, consistent with son-biased stopping rules. Also some mother characteristics (religion, language) are significant—these have been documented to predict son preferences (Pande and Astone 2003). The null of the joint test for this group can be rejected at the 1% level.

The mean differences we observe for older children could result from sex-selective abortion, excess girl mortality or stopping rules. We cannot assess the bias that results from sex-selective abortion, but if it were large enough even our youngest sample would not appear to be balanced between boys and girls. To assess the effect of excess girl mortality, Figure 2 also plots the p-value of the test for the sample of ever born children (including children that mothers reported to have died by the time of the survey). For this sample we do not reject the null for the first time until children reach age 14-25 months—thus in the absence of excess-girl mortality, gender would appear to be balanced among the youngest child for a much larger (older) group of children. However, even in this group, we reject the null consistently for older children after a certain age, as stopping rules predict.

There are a couple of caveats to the results in this section. As in other tests of random assignment, our test is imperfect in that we can only observe that the samples are identical based on observables—it is possible they are different based on unobservables. Also, although our samples are large, they are not large enough to precisely identify the age at which the covariates become unbalanced.

In summary, the data are supportive of the assumption that gender is as good as "randomly assigned" among the youngest children 12 months and younger, and perhaps for those under 2 years of age but not for those that are older. We use these two samples to estimate whether girls are given fewer resources than boys. We start by looking at time, and then look at other investments that have been studied in the literature.

V-Results from the Time Use Survey

We begin investigating whether boys and girls are treated differently by examining whether families spend more time taking care of children if the youngest child is a boy. We use data from the Indian Time Use Survey (hereafter TUS) conducted from July 1998 to June 1999 by the Social Statistics Division of the Central Statistical Organization of India. The TUS asked about the time use of all household members over 5 years of age during the previous 24 hours. The diary section was open-ended in terms of both describing the activities and giving beginning and ending times, with each activity identified as multiple (simultaneous) or not.[16] The survey collected data in six states selected to be representative of the different regions of the country (Gujarat, Haryana, Madhya Pradesh, Meghalaya, Orissa and Tamil Nadu) and interviewed 12,750 rural and 5,841 urban households, for a total of 75,000 respondents.

We analyze time use data corresponding to "normal" days only (excluding, for example, holidays).[17] The main variable of interest is the amount of time spent on childcare by the

[16] The activities were detailed coded into 176 different types. For simultaneous activities, field workers determined the main activity and distributed the total time spent according to the relative importance of activities.
[17] This excludes "abnormal" days when there are guests, someone is sick or there is a festival; and "weekly variants." Most days are included, because as Hirway (2000) notes "in rural areas people continue their normal

household. We follow Guryan et al. (2008) as closely as possible, and classify the following activities as childcare: physical care of children (washing, dressing, feeding); teaching, training and instruction of own children; accompanying children to places; travel related to care of children; and supervising children (with or without other activities).

These data have a couple of limitations for our analyses. Aside from containing information only on 6 states, there is very little information about the participants. Also, families cannot be identified, only households. We can only identify the youngest child in the household—for this reason, we restrict the sample to children who are the children or grandchildren of the household head.[18] The most important limitation of the TUS is that we do not know the identity of the child who was being cared for, we just know whether individuals reported being occupied with childcare.[19] This feature however does have an advantage: since the questions on childcare do not refer to a particular child, it is less likely that respondents are systematically biasing their responses based on the gender of the youngest child.

Based on the results from the previous section, we divide households into 2 samples: those whose youngest is under 2 years of age, and those whose youngest is older.[20] Because the TUS is small and because there is substantial age-heaping at age 1 (which appears to be differential by gender, Coale and Demeny 1967, Bhat 1990), we do not report results for children under age 1, though they are qualitatively similar.[21] Table 3 presents summary statistics for these two samples. Households with infants spent on average more than 3 hours on childcare per day, while households with older children spent a little less than 2 hours.[22] Women provided more than 80% of the total time spent on child care by the household. About 70% of childcare

activities on holidays also." No households are dropped by making this restriction since they were all interviewed for at least one normal day.

[18] Children that do not live with their biological parents receive less care on average, and it is possible this differs by gender—for example families are much more likely to adopt girls than boys. We restricted the sample to avoid these complications. We also exclude households that had more than one child with the youngest age so we can define the sex of the youngest (if there is a boy and a girl both aged 3, we cannot tell who is the youngest).

[19] The survey did not ask the respondent who was present when an activity was performed.

[20] We tested our identifying assumption for these TUS samples. The results are in Appendix Table 1.

[21] See Appendix Table 2.

[22] Although these numbers might seem small, they are roughly comparable to those from time use surveys in other countries. For example Guryan et al. (2008) in table 4 report that the average time per week spent on childcare for adults with children ranges from 4 hours (South Africa) to about 9 hours (US). Assuming that there are 3 adults per household on average this translates into 1 (South Africa) to 4 (US) hours per day at the household level. The most likely reason why the numbers are so low is that individuals are only reporting childcare when it is being performed as a primary activity (exclusively)—previous research (Fedick et al. 2005) suggests that estimates of total childcare time are about 3 to 4 times larger when time spent with children (though not reported as childcare) is included.

consisted of physical care of children. Roughly half of the time devoted to childcare consisted of exclusive childcare, in which the caregiver did not report any simultaneous activity—we use this as a measure of quality of childcare.

Preliminary evidence of differential treatment by gender is presented in Figure 3, which shows the cumulative distribution of childcare separately by gender of the youngest child under age 2. About 10% of all households report spending no time (collectively) on childcare, even though the youngest child is under age 2, but this does not appear to differ by gender. It is quite clear though that households with baby boys spend more time on childcare than those with baby girls, as the baby boy distribution appears to first-order stochastically dominate that of baby girls.

To obtain estimates of the effect of gender on childcare time we estimate

$$Z_h = \alpha_0 + \alpha_1 * B_h + X_h \rho + u_h,$$

where Z_h is the total amount of time that adult members in the household spent on childcare, and B_h is a dummy for whether the household's youngest child is a boy. We present results with and without controlling for predetermined household-level covariates X_h. The standard errors are estimated using White's correction for heteroskedascity and we use the survey weights.[23]

The main results are in Table 4. Since our dependent variable is the number of minutes spent on childcare, we estimate various models. The first column estimates a simple OLS model where the dependent variable is the total number of minutes spent on childcare including 0s. It shows that households whose youngest child is a boy spend roughly 19 minutes more per day taking care of children than households whose youngest child is a girl, about 9% more relative to the mean. Column 2 shows that this estimate is robust to controlling for religion, ethnicity and the area of the land that the household owns. In column 3, we estimate a logit of whether the household spends any time on childcare. As suggested by Figure 3, there is no statistically significant effect of gender, its coefficient is insignificant and the implied marginal effect is very small (less that 1%). If we estimate instead an OLS model only for those that report some care, we find that households whose youngest child is a boy spend roughly 22 minutes more (about

[23] We also estimated standard errors taking the survey design into account (Appendix Table 3). The TUS had a sophisticated sampling scheme with three levels of stratification and clustering, and consequently there were many strata with one sampling unit. Given that some assumptions have to be made in order to calculate the standard errors (see Appendix 3), we opted for showing the OLS standards errors in the main tables.

10% more) per day on childcare than households whose youngest child is a girl. Column 5 estimates a Tobit model, which accounts for censoring at 0. Again we find a statistically significant increase in childcare of about 10%.

The second part of the table presents results for households whose youngest is between 2 and 5 years of age. Regardless of the specification, we do not find any statistically significant effect of gender, and in fact several coefficients have the “wrong” sign. Thus, we fail to find evidence of differential treatment among the older group.[24]

Table 5 looks at whether the effects of gender differ depending on the number of children in the household. The first column of Table 5 reproduces our main estimates from Table 4 column 1. In column 2, we interact gender of the youngest with the number of other children in the household under the age of 6, which is also added as a control. Not surprisingly families with many young children spend more time on childcare, about 12 minutes per child under 6. The coefficient on gender is now larger, about 50 minutes, but the interaction with number of children is negative. If the youngest is the only child under 6, and he is a boy, the household spends 50 minutes more on childcare but not if there are two or more other children under 6—so it would appear that when there are many small children, there is simply “no room” to provide differential treatment. Our main results so far suffer from one important caveat: a significant coefficient on the youngest’s gender could be interpreted as parents reallocating time across children depending on the sex of their youngest child (rather than giving more time to the youngest when he is a boy). Column 3 restricts attention to families with no other children under 6. It confirms the estimates from our interacted specification (column 2). More importantly, it suggests that boys do indeed receive more care than girls.

In columns 4-7 we look at whether the estimates differ depending on whether childcare is private (such as physical care) or public (supervising) in nature. Households spend more time in both private and public childcare if the baby is a boy than if it is a girl (columns 4 and 6). The amount of private and public childcare time per child is increasing with the number of children, but it increases faster for public care (columns 5 and 7). Most interestingly, time allocated into physical care is larger in baby-boy households, and the effect does not decline with the number of other children under 6 (or does so at a small rate); whereas for public care the effect of gender

[24] Appendix Table 2 shows results by single year of age: around age 2 the estimates of the effect of gender of the youngest child start to be substantially biased with girls getting more time than boys, but this is not true for age 5.

disappears if there is one more child under 6. These patterns can be explained by the private versus public nature of childcare activities. Since supervising is a public type of care, it makes sense that, as the number of young children in the household increases, members will spend disproportionally more time in this type of care and this time will not be closely related to the sex of the youngest child. In contrast, physical care is private so there is room for differential treatment of boys and girls even when other young children are present.

The last column shows estimates of the effect of gender on "exclusive child care time", our proxy for quality care, defined as the number of minutes that adults spent caring for children and not doing anything else at the same time. The results show that households whose youngest child is a boy provide more exclusive childcare than households whose youngest child is a girl, roughly 37 minutes more per day if there are no other children under 6.

In table 6 we document how much time different household members spend on childcare. We compute total minutes spent on childcare by adult females (age 15 and older), adult males (age 15 and older) or juvenile females (age 6-14). Adult women are more likely to spend time on childcare, and to spend more time (about 10% more), if the youngest is a boy—and the effect is twice as large if there are no other children under 6. The patterns are the same for time spent by adult males—they spend more time (about 17% more) when the youngest is a boy, though this is not statistically significant. But if no other child under 6 is present, they spend about 50% more time when the youngest is a boy, and this is statistically significant. Time allocated to childcare by juvenile females is also larger when the youngest is a boy, but the effects are only significant when no other child under 6 is present—in which case juvenile females spend triple the time with boys than they do with girls. Although these results suggest that all members of the house spend more time in childcare when the youngest is a boy, it could also be that when the youngest is a boy the composition of the household changes (if, for example, some relative moves in to help take care of the baby boy).[25] We examine this directly in the final section of the paper. In column 7 we repeat the estimation for urban households who are deemed to discriminate less. Indeed we do not find that the youngest's gender is statistically significant—in fact the effect is large and negative, but the sample is small and the standard errors are large.

[25] We also estimated results for boys 6-14 (available upon request). The coefficients on gender have the "wrong" sign, and are never statistically significant. This is most likely due to the fact fewer older boys remain in the household when a baby boy is born—see results in Table 10.

Overall we find that more time is spent in childcare in households whose youngest is a boy, by both adult males and females, as well as by juvenile females. And the quality of this time is higher. Because the result holds in households with a single child, we interpret them as meaning that boys are allocated more time than girls. These findings raise the question of whether in part parents and other members of the households are allocating more time to young boys because they need it more—this could be the case if boys are more active, or they get sicker more frequently. To assess this possibility we look at South Africa, a developing country with a time use survey[26] for which there is no evidence of son preference (Gangadharan and Maitra 2003). We find no evidence that South African households whose youngest is a boy spend more time on childcare than households whose youngest is a girl: the cumulative distribution of household childcare time is very similar for both sexes (see Figure 4) with baby-boy households spending on average 2 minutes more per day in childcare than baby-girl households (mean differences are not statistically significant, results available upon request). Although this evidence is only suggestive—since it is not necessarily clear that South African provides a good counterfactual for India, it does not support the hypothesis of differential time needs by gender. We now look at whether gender affects other parental investments in India.

VI-Gender differences in other inputs: additional results from the DHS.

We now proceed to investigate whether there are boy-girl differences in other child investments using DHS data on breastfeeding, vitamin A supplementation and vaccination.[27] We estimate whether the gender of the youngest child of a given mother predicts parental inputs. All estimations use survey weights and correct the standard errors for survey design. The results are reported in Table 7. We first look at breastfeeding, which is deemed to be the ideal source of nutrition for infants, particularly in developing countries where food and water quality are low, and sanitation is poor.[28] We do not find that boys are more likely to have ever been breastfed

[26] We use data from the South African Time Use Survey, conducted in 2000 by Statistics South Africa. Information on time use was collected for persons aged 10 years or more, with two respondents randomly chosen per household (or only one if there was only one household member aged 10 years or more). Data were collected for 8,564 households (14,553 respondents). We use data from 1,025 households whose youngest member is under 2 years old.
[27] Mothers reported the information for all children under three in 1998 and under four in 1992.
[28] See Jayachandran and Kuziemko (2009) for a more detailed discussion on the benefits of breastfeeding in the context of developing countries.

(defined as: ever breastfed, or breastfed less than a month), and this is true in both the linear and non-linear specifications; and for both samples. The effect sizes are precisely estimated zeroes. Most likely this is because more than 95% of children are ever breastfed.

In the next set of columns we look at the duration of breastfeeding. We estimate several specifications for this outcome. We estimate censored linear regressions, since many children are still being breastfed at the time of the interview. Alternatively we estimate a censored log-linear model or a standard accelerated-failure time model. We find a positive but small (about 2-4%) and statistically insignificant effect of gender for children 0 to 12 months old. If we estimate a proportional hazard model, we find that the odds of stopping breastfeeding are lower for males but again the coefficient is not statistically significant. These results are perhaps not surprising given that 98% of children 0-12 months are still being breastfed (conditional on breastfeeding) so there is very little variation in the outcome for this group. By contrast, for children 0-23 months (Panel B) all the coefficients are statistically significant. The magnitudes suggest that breastfeeding duration increases by about 7-14% when the child is a boy (depending on the specification), consistent with the findings in Jayachandran and Kuziemko (2009). The results for breastfeeding duration underscore the tradeoffs that we face using our identification strategy: the omitted variable bias problem should not exist for the very young, but the number of outcomes and the variation in outcomes may not be sufficient to detect differences by gender.

Next we look at whether children are given Vitamin A, which protects against night blindness, measles and diarrhea.[29] For both samples, and regardless of whether we estimate a linear or non-linear model, we find that boys are more likely to receive vitamin A: about 11% more for ages 0-12 months, and 7% more for ages 0-23. Finally, we look at whether mothers had a vaccination card at the time of the interview. Only about 30% of mothers have a vaccination card, but they are 11% more likely to have the vaccination cards of boys. For both samples and all outcomes, the results are not sensitive to the inclusion of covariates.

We also investigate whether boys are more likely to be vaccinated.[30] At the interview, mothers were first asked for the vaccination cards. If the mother had it, then all the vaccination

[29] Children between 6 months of age and 5 years are supposed to take a Vitamin A supplements every 6 months. The first two doses can be given at the same time required vaccinations are given.
[30] The recommended vaccination schedule in India for children is as follows: BCG at birth, polio at birth, 6 weeks, 10 weeks and 14 weeks; DPT at 6 weeks, 10 weeks and 14 weeks and measles at 9 months. BCG protects against tuberculosis and DPT protects against diphteria, pertussis and tetanus.

information was recorded directly from the card. If the card was not available, mothers were asked to provide information on each type of vaccination. For each vaccine (except for polio at birth which was not collected in one of the waves), we construct two measures of whether the child had received the dose: (1) the mother reported that the child had received the dose, or the dose had been recorded in the vaccination card; or (2) the dose had been recorded in the vaccination card. The means in Table 8 show that vaccination rates are much lower for the full sample (which includes mother reports) than for the sample with a vaccination card. For example, 47% of children 0-12 months old are vaccinated against BCG, but among those with a vaccination card the rates are 89%. When using the full sample, we find that boys are more likely to be vaccinated. These results are shown in Panels A, B and C of Table 8.The magnitudes vary depending on the vaccination but they range from 6 to 17% for all three panels.[31] Oster (2009) and Jayachandran and Kuziemko (2009) find similar results.

One issue with these results is that they partially rely on mothers' reports, which could be biased. The bottom part of the Table 8 presents the results when we restrict the sample to those whose vaccination information was recorded from the vaccination cards. The results are dramatically different. Not only are most of the coefficients on gender statistically insignificant—which we might expect given that the sample size falls to about 1/3 of its original size—they are also small and sometimes negative (with a few exceptions for the youngest sample). The results using the vaccination card sample are however subject to non-trivial sample selection bias given that Table 7 showed that mothers were much more likely to have cards for boys and that less than 1/3 of the sample has cards.[32] The vaccination results suggest reasons why some previous research has not found large gender differences in vaccinations: some was based on vaccination cards (Borooah 2004) and others are from surveys based entirely on mothers reports, which could be more (or less) reliable than in the DHS—this might be the case in the NSS, which Deaton (2003) uses to draw his conclusions.

[31] The results are identical if we restrict the sample only to those "at risk," namely those that are older than the recommended age for the vaccination (Panel C). Appendix Table 4 shows that the results are the same using non-linear models and again they are not affected by the inclusion of controls.

[32] Differential misreporting of mothers by gender would have to be large to account for our findings. For example using children 0-12, we calculate that in order for the BCG vaccination rates in the full sample to be similar across boys and girls, mothers of boys without a vaccination card would have to be at least 13% more likely to misreport that her child had received the BCG than mothers of girls without a vaccination card. These estimates are different for each dose and sample. They range from 9% for the first dose of polio (children 0 to 24 months old) to 28% for the third dose of DPT (sample of children 0 to 12 months old).

In Appendix Table 5 we estimate the bias generated by mortality. Since we have reports on all children born, we can impute outcomes for them under best- and worst-case scenarios (for boy-girl differences) and obtain bounds on the coefficients for gender. These results suggest that our estimates are likely to be downward biased (given that the lower bounds seem to be based on unreasonable assumptions) perhaps by as much as 50%. However the results are just suggestive since mothers are likely to misreport deaths differentially by gender. This table also shows results for urban children. The coefficients for boys are somewhat smaller in magnitude relative to the means (with the possible exception of breastfeeding) but not statistically significant.

Appendix Table 6 shows results by single year of age. We use the 1992 data which includes children up to 47 months old and include those who are not the youngest. The patterns of the coefficients by age differ depending on the outcome: the effect of gender on breastfeeding falls with age, whereas the effects of gender for vitamin A, vaccination cards and vaccination increase (in percentage terms) from age 0 to 1 and then decrease from 1 to 2, and then increase again from age 2 to 3.[33] It is not entirely clear why the coefficients for ages 2 and 3 differ so much—but it is clear that the estimated size of the effects would change a lot when looking at older samples.

Finally we investigate how sex-selective abortion potentially affects our findings. We first investigate whether our "test"that gender is as good as random would fail for states where SSA is suspected to be high. We then re-estimate our main equations separately for states with high and low SSA. Appendix Table 8 shows the results using 3 different criteria as to which states have high SSA: we classify states as being high SSA if a-sex-ratios at birth were larger than 110 in either 1992 and 1998; b-sex ratio at birth was larger than 110 in both 1992 and 1998, and c-based on Retherford and Roy (2003)'s characterization.[34] In all three cases we find that at the 10% level we would reject the null than gender is random at birth for the high SSA states but not for the low SSA states. For all groups however we continue to find that boys receive more inputs than girls, in spite of the fact that SSA biases our estimates downwards.

In summary, for all the measures we looked at, we find that boys are given more inputs than girls; in general, girls receive about 10% less than boys. Again these differences could be

[33] We also estimated the coefficients by age using the 1998 data, although we can only go up to 35 months of age. The coefficients fall with age for most outcomes. It is not clear why the age pattern differs in 1992 and 1998.
[34] These states have high sex ratios that increase with birth order, and high rates of prenatal care and ultrasound use.

driven by differential needs of boys. However in South Africa, there are no systematic differences between boys and girls in terms of breastfeeding or vaccination rates.[35]

To assess the magnitude of these differences, we estimate how much gender differences in investments can explain the higher mortality rates among girls using estimates from the literature of the effects of breastfeeding, vitamin A supplementation and vaccinations on mortality.[36] Mortality rates among children 12 to 36 months of age are 16.7 per 1,000 for boys and 24.2 per 1,000 for girls. A back of the envelope calculation suggests that the observed differences in investments explain about 27% of excess girl mortality among children in this age group (this translates into 2 additional girl deaths per 1,000 children). We can further explain 3.4% of girl excess mortality if we assume a modest effect of time inputs on mortality.[37]

VII-Other results from the DHS: anthropometric measures and living arrangements

We now look at the effect of gender on height-for-age, weight-for-age and weight-for-height Z-scores, which are computed by subtracting the median of the reference population and dividing by the standard deviation of the standard population.[38] It is important to normalize outcomes since boys are otherwise known to be taller and heavier than girls. We further examine whether gender determines the likelihood of a child being stunted, underweight or wasted.[39]

---

[35] The 1998 DHS for South Africa can be used to assess this. Among children ages 12-23 months of age, girls are slightly more likely to have been vaccinated for DPT, and Polio, and slightly less likely to have been vaccinated from BCG or Measles (Table 7.9, DOH 2002). Among children under the age of 3, girls are more likely to be breastfed and the median duration of breastfeeding is longer for girls (Tables 8.1 and 8.3, DOH 2002).

[36] For each investment we first calculate the (gender neutral) probability of death conditional on not receiving the investment (p_0) and the (gender neutral) probability of death conditional on receiving it (p_1) using the relative risks estimated in the literature [vitamin A (Rahmathullah et al 2003); breastfeeding (Briend, Wojtyniak, and Rowland, 1988 and WHO 2000); measles (Koenig et al 1990); Polio, BCG and DPT (Moulton et al 2005)], the mortality rate for children 12 to 36 months old (20.3 per 1,000 children) and the fraction of children in this age group receiving the investment (see Jayachandran and Kuziemko 2009 for a more detailed discussion). Let θ_b be the fraction of boys and θ_b be the fraction of girls (in the age group) who receive the investment. The difference in the mortality rates of girls versus boys associated to gender differences in the investment is equal to $(\theta_b - \theta_g)*(p_0 - p_1)$. We sum these differences over all investments and divide the total by the difference in mortality rates of girls and boys.

[37] We know of no good estimates of how parental time affects mortality. Assuming that one additional hour of childcare reduces the probability of death (in absolute terms) by 0.0003, the boy-girl difference in time use of roughly 51 minutes translates into a boy-girl difference in mortality of 0.000255.

[38] About 15% of children were not measured but this was not different by gender.

[39] A child is stunted if the height-for-age is 2 s.d. below the median of reference population (measures chronic under-nutrition); a child is underweight if the weight-for-age is 2 s.d. below the median (measures both chronic and acute under-nutrition); a child is wasted if the weight-for-height is 2 s.d. below (measures acute under-nutrition).

Before we discuss the results, it is important to notice these are not ideal measures to investigate whether boys and girls are treated differently. Anthropometric measures are outcomes, not inputs. Height and weight are the result of caloric inputs, but also of other factors such as the incidence of disease and caloric expenditure, which may differ by gender for biological reasons. Indeed, Sommerfelt and Arnold (1998), using data for almost all developing countries for which DHS data is available (41 surveys were used), find that girls under 2 have better anthropometric measures that boys. Also, it is possible that one may find differences because of the standardization methods used.

Table 9 shows that for both of our experimental samples boys fare *worse* than girls if we use the Z-scores provided by the DHS, which are based on the WHO standards. They are more likely to be short or extremely short for their age, and more likely to be wasted. The other measures are not statistically significant, but they also point in the same direction. We also present results using the 1990 British Standards for height-for-age and weight-for-age and the 2000 CDC standards for weight-for-height.[40] The table shows that we obtain very different results with these alternative standards, where now boys appear to do better than girls.

These results are hard to reconcile with the previous evidence presented here that inputs are higher for boys. We cannot conclusively say why, in spite of differential inputs, boys do not always appear to do better on anthropometric measures. The answer we obtain depends on the standard, so perhaps the standards are not correct.[41] It has been hypothesized that baseline differences in height and weight between boys and girls are larger in well-fed populations (boys' growth is more responsive to environmental conditions, Stinson 1985), thus to estimate differential treatment we would need to compare Indian children to a equally well fed population that does not treat boys better. We cannot resolve this puzzle here, but we simply note that anthropometric measures paint a very different picture than other measures in terms of differential treatment of boys and girls.

Finally we investigate whether living arrangements are affected by the gender of the youngest boy. Studies in the US report that having a son reduces the probability of parents getting divorced (Katzev, Warner and Acock 1994; Morgan, Lye and Condran 1988; Mott 1994;

[40] There are no British standards available for weight-for-height. The 2000 CDC standards for height are not available for children under 2.
[41] Moetsue (2009) also finds that for Blangladesh boy-girl differences depend on the standard used.

Spaneir and Glick 1981) and that daughters are less likely to live with their fathers (Dahl and Moretti 2004). There is, however, little research in the context of developing countries, where boy-girl discrimination is thought to be a greater concern. Table 10 reports the results, for children 23 months of age and younger. Panel A uses the DHS and examines whether gender of the youngest affects the likelihood that different family members live together. We do not look at marital status as an outcome because in our samples all mothers are married. But we look at whether women report that their husband's live at home. We do find that if the youngest is a boy then the husband is more likely to live at home, but the effect insignificant and very small, less than 1%. There is no evidence that the youngest's gender affects the likelihood that the mother is the household head's wife—which might occur if for example families move in with their parents after a boy is born. There is also no discernable effect of the youngest gender on the number of siblings living at home. But we do find that the gender composition of the siblings living at home changes: sisters are more likely to live at home whereas brothers are less likely to live at home if the youngest is a boy. Again although these effects are statistically significant, they are small in magnitude about 3 to 4% relative to the sample mean. In Panel B we look at household composition in the TUS. Although the unit of observation is now the household, we find somewhat similar results. If the youngest is a boy, then there are more men in the household over age 15, but the effect is small and statistically insignificant. There are more women over 15 (about 7%), and this effect is statistically significant. There is again no apparent effect on the number of total children under age 14, but these are more likely to be girls and less likely to be boys. Together these results suggest that when the youngest is a boy, the family is more likely to retain their daughters for caregiving, or to have another female adult move in to provide help.

VII-Conclusion

Despite the evidence showing that girls have higher mortality rates than boys in a number of South and East Asian countries, there is considerable debate in the literature as to whether boys and girls are treated differently. In this article, we developed a novel empirical strategy to address possible biases in previous estimates of differential treatment and we used it to investigate whether families treat boys better than girls in India. In particular we examined whether families spend more time with childcare when the baby is a boy than when the baby is a girl. Time investments have not been previously studied in the context of developing countries. We also used our identification strategy to look at differential treatment along measures previously used in the literature.

We find evidence that boys receive more child investments than girls. Households with a boy under two spend roughly 20 minutes more per day (about 10% more time) on childcare than households with a girl under two. This difference is even larger for one-child households: households with one boy under six spend roughly 50 minutes more per day on childcare than households with one girl under six. We also find suggestive evidence that the quality of childcare given to boys is higher. Moreover, we find that boys are more likely to be vaccinated, to be breastfed longer and to be given vitamin supplementation. In general we find these inputs to be about 10% lower for girls. However we find that the effects for anthropometric measures (which are outcomes, not investments) are equivocal, and sensitive in particular to specific standard chosen.

We interpret our results as evidence that families discriminate against girls because they have a preference for sons, because the returns to these investments are higher for boys (e.g., men have higher wage rates than women) or because families that have girls anticipate having larger families. An alternative explanation would be that boys receive more investments because they "need more," as would be the case if boys were frailer, for example. However, this hypothesis seems unlikely, as it would imply that—in investing their time and other inputs—families end up "overcompensating," given that girls in India have higher mortality rates than boys. It also seems unlikely that families see boys as needing more of all inputs, including inputs such as vaccinations. In other developing countries with no son-preference there does not appear to be any evidence that boys need more. In South Africa for example, we find that there are no systematic gender differences in the time allocation of parents or in the breastfeeding and vaccination rates. Thus we view our results as providing evidence that on average parents do indeed discriminate against girls. But we cannot provide any evidence about why do they so.

References

Almond, D. and, L. Edlund. 2007. "Trivers-Willard at birth and one year: evidence from US natality data 1983-2001." *Proceedings of the Royal Society,* 274: 2491-2496.

Arnold, F., M. K. Choe, and T. K. Roy. 1998. "Son preference, the family-building process and child mortality in India." *Population Studies,* 52: 301-315.

Basu, A. M. 1989. "Is Discrimination in Food Really Necessary for Explaining Sex Differentials in Childhood Mortality?" *Population Studies* 43 (2), 193–210.

Bhat, P. N. M. 1990. "Estimating Transitional Probabilities of Age Misstatement." *Demography*, 27(1): 149-63.

Borooah, Vani K. 2004. "Gender bias among children in India in their diet and immunisation against disease." *Social Science and Medicine* 58(9): 1719 – 1731.

Becker, G. 1991. A Treatise on the Family. Cambridge: Harvard University Press.

Briend, A., B. Wojtyniak, and M. G. Rowland. 1988. "Breastfeeding, Nutritional State, and Child Survival in Rural Bangladesh." *British Medical Journal*, 296(6626): 879-882.

Case, A., A. Fertig and C. Paxson. 2005. "The lasting Impact of Childhood health and Circumstances." *Journal of Health Economics*, 24: 365-389.

Chahnazarian, A. 1988. "Determinants of the sex ratio at birth: Review of recent literature." *Social Biology,* 35(3-4): 214-235.

Chen, L. C., E. Huq, and S. D'Souza. 1981. "Sex bias in the family allocation of food and health care in rural Bangladesh," *Population and Development Review,* 7: 55-70.

Clark, S. 2000. "Son Preference and Sex Composition of Children: Evidence from India." *Demography,* 37: 95-108.

Coale, A. J. and P. Demeny. 1967. "Methods of Estimating Basic Demographic Measures from Incomplete Data." Manual IV, Population Studies No 42, United Nations, New York.

Dahl, G.B. and E. Moretti. 2004. "The Demand for Sons." *Review of Economic Studies,* 75(4): 1085-1120.

Das Gupta, M. 1987. "Selective discrimination against female children in rural Punjab, India." *Population and Development Review*, 13(1): 77-100.

Deaton, A. 1989. "Looking for Boy-Girl Discrimination in Household Expenditure Data." *World Bank Economic Review,* 3: 1-15.

Deaton, A. 1997. The analysis of household surveys: A microeconometric approach to development policy. Baltimore and London: John Hopkins University Press.

Deaton, A. 2003. " Health, Inequality, and Economic Development," *Journal of Economic Literature*, 41(1): 113-158.

Department of Health, Medical Research Council & Measure DHS+. South Africa Demographic and Health Survey 1998, Full Report. Pretoria: Department of Health; 2002.

Dibbley, M. J., N. Staehling, P. Nieburg and F. L. Trowbridge. 1987. "Interpretation of Z-score anthropometric indicators derived from the international growth reference," *American Journal of Clinical Nutrition*, 46: 749-62.

Duflo, E. 2005. "Gender Equality in Development," MIT Working Paper.

Edlund, L. 1999. "Son Preference, Sex Ratios, and Marriage Patterns." *Journal of Political Economy,* 107(6): 1275-1304.

Fedick, C. B., S. Pacholok and A. H. Gauthier. 2005. "Methodological Issues In The Estimation Of Parental Time – Analysis Of Measures In A Canadian Time-Use Survey," *International Journal of Time Use Research,* 2(1): 67-87

Ganatra, B. and S. Hirve. 1994. "Male bias in health care utilization for under-fives in a rural community in western India." *Bulletin of the World Health Organization,* 72(1): 101-104.

Guryan, J., E. Hurst and M.Kearney, 2008. "Parental Education and Parental Time with Children." *Journal of Economic Perspectives,* 22(3): 23-46.

Hirway, I. 2000. "Tabulations and Analysis of the Indian time use survey data for improving measurement of paid and unpaid work" Expert group meeting on methods for conducting time use surveys 23-27 October 2000, United Nations Secretariat, Statistics Division, New York.

Jayachandran, S. and Kuziemko, I. 2009. "Why Do Mothers Breastfeed Girls Less than Boys: Evidence and Implications for Child Health in India." NBER Working Paper 15041.

Jensen, R. 2005. "Equal Treatment, Unequal Outcomes? Generating Sex Inequality through Fertility Behavior." Unpublished Manuscript, Harvard University.

Khanna R., A. Kumar, J. F. Vaghela, V. Sreenivas, J. M. Puliyel. 2003. “Community Based Retrospective Study of Sex in Infant Mortality in India,” *British Medical Journal*, 327:126-8.

Katzev, A. R., R. L. Warner, and A. C. Acock. 1994. “Girls or Boys? Relationship of Child Gender to Marital Instability.” *Journal of Marriage and the Family,* 56: 89-100.

Koenig, M. A., M. A. Khan, B. Wojtyniak, J. D. Clemens, J. Chakraborty, V. Fauveau, J. F. Phillips, J. Akbar, and U. S. Barua. 1990. “Impact of Measles Vaccination on Childhood Mortality in Rural Bangladesh,” *Bulletin of the World Health Organization* 4: 441–447.

Lundberg, S., S.W. Pabilonia and J. Ward-Batts. 2007. “Time Allocation of Parents and Investments in Sons and Daughters.” Working paper. Department of Economics, University of Washington.

Mammen, K. 2008. “The Effect of Children's Gender on Living Arrangements and Child Support.” *American Economic Review: Papers and Proceedings,* 98(2): 408-412.

Mammen, K. 2009. “Fathers' Time investments in children: do sons get more?” *Journal of Population Economics.*

Marcoux, A. 2002. “Sex differentials in undernutrition: A look at survey evidence.” *Population and Development Review*, 28(2): 275-284.

Miguel, E. 2005. “Poverty and witch killing." *Review of Economics Studies* 72(4): 1153-1172.

Morgan, S. P., D.N. Lye, and G. A. Condran. 1988. “Sons, Daughters, and the Risk of Marital Disruption.” *American Journal of Sociology,* 94: 110-29.

Mott, F.L. 1994. “Sons, Daughters and Fathers' Absence: Differentials in Father-Leaving Probabilities and in Home Environments.” *Journal of Family Issues,* 15: 97-128.

Moestue H, de Pee S, Hall A, Hye A, Sultana N, Ishtiaque MZ, Huq N & Bloem MW. 2004. “Conclusions about differences in linear growth between Bangladeshi boys and girls depend on the growth reference used,” *European Journal of Clinical Nutrition* 58: 725–731.

Moestue, H. 2009. “Can anthropometry measure gender discrimination? An analysis using WHO standards to assess the growth of Bangladeshi children." *Public Health Nutrition* 12(8): 1085-1091.

Moulton, L. H., L. Rahmathullah, N. A. Halsey, R. D. Thulasiraj, J. Katz, J. M. Tielsch. 2005. "Evaluation of Non-Specific Effects of Infant Immunizations on Early Infant Mortality in a Southern Indian Population." *Tropical Medicine & International Health* 10(10): 947-955.

Oreopolous, P., M. Stabile, R. Walld and L. Roos. 2008. “Short, Medium and Long Run Consequences of Poor Infant Health: An Analysis using Siblings and Twins.” *Journal of Human Resources*, 43: 88-138.

Oster, E. 2005 “Hepatitis B and the Case of Missing Women,” *Journal of Political Economy*, 113(6): 1163-1216

Oster, E. 2009. “Does increased access increase equality? Gender and child health investments in India.” *Journal of Development Economics*, 89(1): 62-76.

Pande, R. and N. Astone. 2007. “Explaining son preference in rural India: the independent role of structural versus individual factors.” *Population Research and Policy Review* 26(1): 1-29.

Rahmathullah, L., J. M. Tielsch, R. D. ThulasirajD, J. Katz, C. Coles, S. Devi, R. John, K. Prakash, A. V. Sadanand, N. Edwin N, and C. Kamaraj. 2003. "Impact of Supplementing Newborn Infants with Vitamin A on Early Infant Mortality: Community Based Randomised Trial in Southern India." *British Medical Journal*, 327(7409): 254-256.

Retherford R. D. and T. K. Roy. 2003. *Factors affecting sex-selective abortion in India and 17 major states.* NFHS Subject Reports, No. 21. Mumbai: IIPS; and Honolulu: East-West Center.

Rose, E. 2000. “Gender Bias, Credit Constraints and Time Allocation in Rural India.” The *Economic Journal* 110: 738-758.

Sen, A. and S. Sengupta. 1983. “Malnutrition of rural children and the sex bias.” *Economic and Political Weekly,* 18: 855-864.

Sen, A. 1990. “More Than 100 Million Women Are Missing.” The New York Review of Books (December 20, 1990).

Sommerfelt, A. E. and A. L. Piani. 1997. Childhood Immunization: 1990-1994. DHS Comparative Studies No. 22. Calverton, MD: Macro International Inc.

Sommerfelt, A. E. and F. Arnold. 1998. “Sex differentials in nutritional status of young children,” in Too Young to Die: Genes or Gender?” pp. 133-153. New York: Population Division, United Nations.

Stinson, S. 1985 “Sex Differences in Environmental Sensitivity during Growth and Development,” *American Journal of Physical Anthropology*, 28(S6): 123-147.

Subramanian S. and Deaton, A. S. 1991. “Gender Effects in Indian Consumption Patterns.” *Sarvekshana*, 14: 1-12.

United Nations Secretariat. 1988. “Sex differentials in life expectancy and mortality in developed countries: An analysis by age groups and causes of death from recent and historical data.” *Population Bulletin of the United Nations,* 25: 65-107.

Van der Berg, G., M. Lindeboom and F. Portrait. 2006. “Economic Conditions early in Life and Individual Mortality.” *American Economic Review*, 96(1): 290-302.

World Health Organization. 2000. "Effect of Breastfeeding on Infant and Child Mortality Due to Infectious Diseases in Less Developed Countries: A Pooled Analysis." *Lancet*, 55(9202): 451-455.

Yamaguchi, K. 1989. “A Formal Theory for Male-Preferring Stopping Rules of Chilbearing: Sex Differences in Birth Order and the Number of Siblings.” *Demography,* 26(3): 451-65.

Yeung, W.J., J.F. Sandberg, P.E. David-Kean and S.L. Hofferth. 2001. “Children’s Time with Fathers in Intact Families.” *Journal of Marriage and the Family,* 63: 136-5.

FIGURE 1

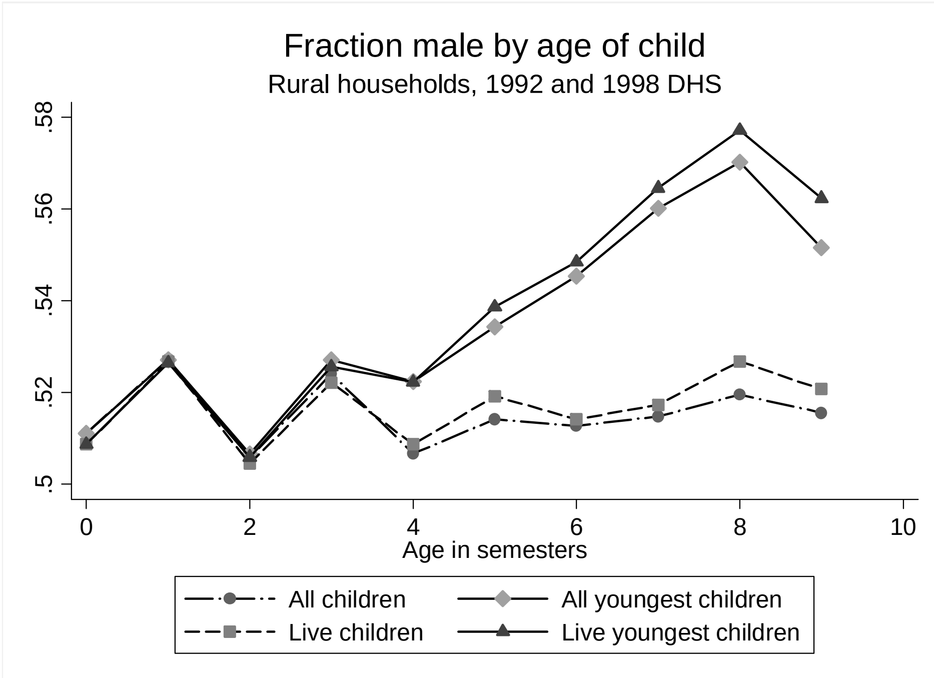

FIGURE 2

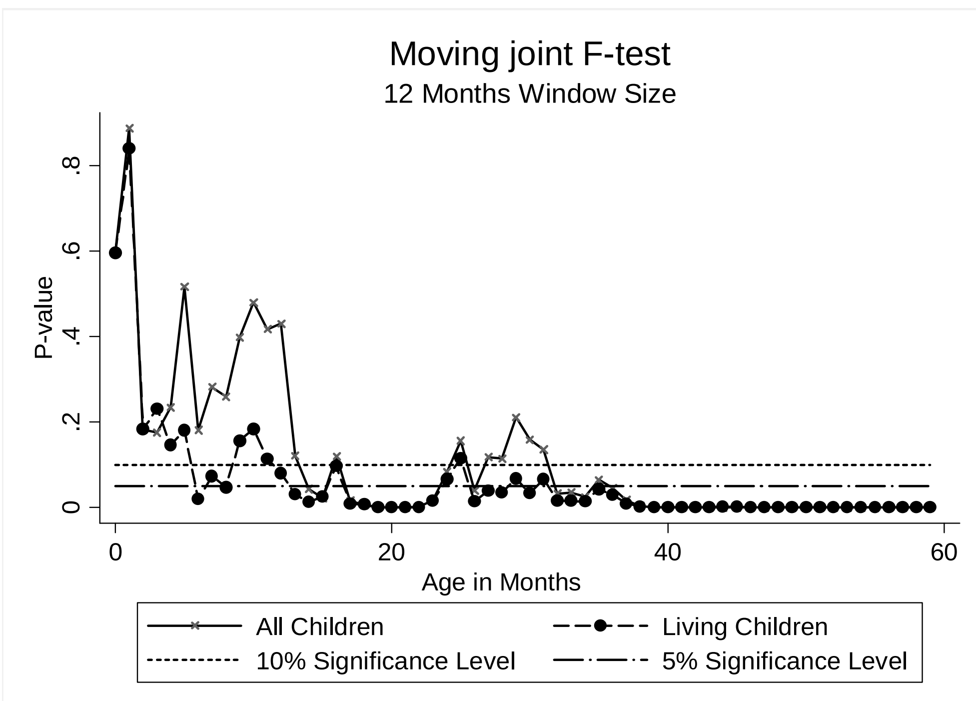

FIGURE 3: Childcare Time by gender, Indian Time Use Survey 1998 -1999

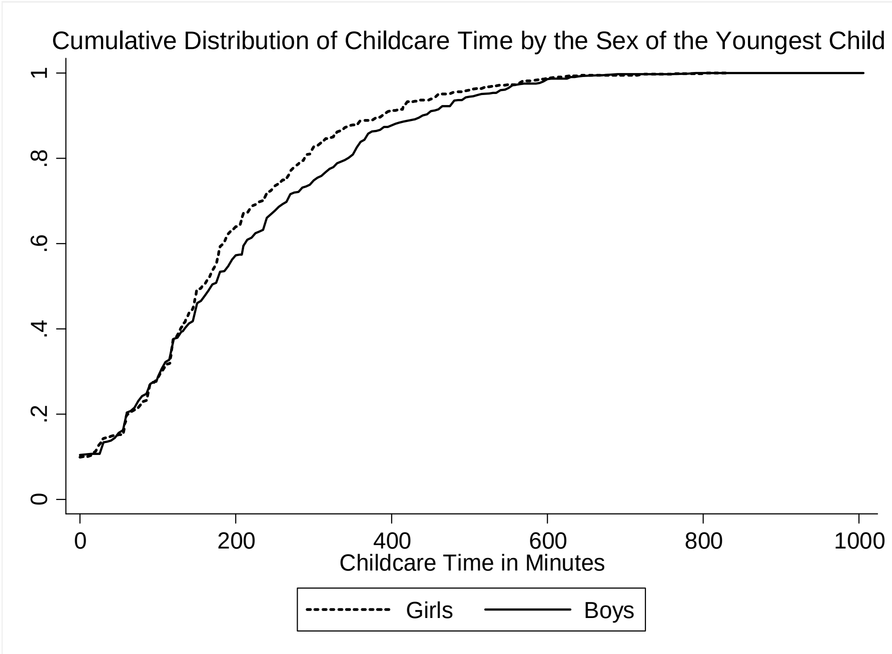

FIGURE 4: Childcare Time by gender, South Africa Time Use Survey 2000

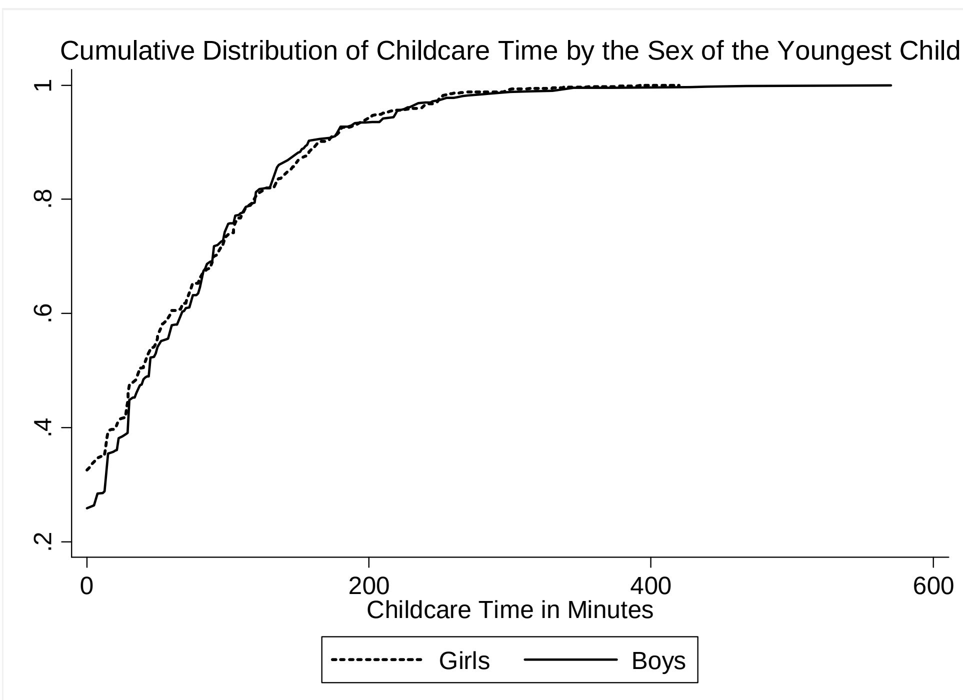

TABLE 1: SUMMARY STATISTICS DHS 1992 AND 1998. RURAL AREAS.

Sample:	Youngest child age 0-12 months N=17385		Youngest child age 0-23 months N=30,985	
	Mean	s.e.	Mean	s.e.
Child characteristics				
Male=1	0.517	0.005	0.516	0.003
Age in months	5.770	0.032	10.912	0.047
Birth month	6.866	0.032	6.908	0.024
# of siblings ever born	2.007	0.019	1.997	0.015
# of brothers ever born	0.942	0.012	0.944	0.009
# of sisters ever born	1.065	0.012	1.053	0.009
Mother's characteristics				
Mother's age	24.395	0.050	24.742	0.040
Mother's ethnicity (scheduled caste omitted)				
Scheduled tribe	0.112	0.004	0.112	0.004
Other	0.715	0.005	0.714	0.005
Mother's religion (other omitted)				
Hindu	0.820	0.006	0.822	0.005
Muslim	0.133	0.006	0.132	0.005
Christian	0.021	0.001	0.021	0.001
Mother's years of education	2.357	0.039	2.331	0.032
Mother born in urban area	0.067	0.002	0.069	0.002
Mother's age first married	16.481	0.028	16.432	0.022
Mother's age at first birth	18.577	0.030	18.540	0.024
Mother speaks Hindi	0.483	0.005	0.484	0.004

Means computed taking survey design into account.

TABLE 2: TESTING RANDOM ASSIGMENT (DHS 1992 AND 1998). MEAN DIFFERENCES BY GENDER.

Sample:	**Youngest live child**					
	Age 0-12 months (N=17,385)		**Age 0-23 months (N=30,985)**		**Age 24-59 months (N=23,058)**	
<u>Child characteristics</u>						
# of siblings ever born	-0.002	[0.034]	0.000	[0.026]	0.032	[0.032]
# of brothers ever born	-0.021	[0.021]	-0.032	[0.015]**	-0.04	[0.019]**
# of sisters ever born	0.02	[0.023]	0.032	[0.017]*	0.071	[0.021]***
Birth month	-0.036	[0.061]	-0.042	[0.046]	0.003	[0.053]
<u>Mother's characteristics</u>						
Mother's age	0.052	[0.092]	0.063	[0.072]	-0.02	[0.092]
Mother's ethnicity (scheduled caste omitted)						
Scheduled tribe	-0.004	[0.006]	-0.007	[0.004]*	0.007	[0.005]
Other	-0.003	[0.008]	0.003	[0.006]	-0.01	[0.007]
Mother's religion (other omitted)						
Hindu	0.002	[0.007]	-0.001	[0.005]	0.006	[0.006]
Muslim	0.001	[0.006]	0.001	[0.005]	-0.01	[0.005]*
Christian	0.001	[0.002]	0.000	[0.002]	-0	[0.002]
Mother's years of education	-0.014	[0.068]	0.066	[0.049]	0.03	[0.056]
Mother born in urban area	-0.001	[0.004]	0.002	[0.003]	0.001	[0.004]
Mother's age first married	0.044	[0.048]	0.063	[0.035]*	-0.04	[0.045]
Mother's age at first birth	0.045	[0.053]	0.042	[0.040]	-0.09	[0.049]*
Mother speaks Hindi	-0.008	[0.009]	-0.008	[0.007]	0.023	[0.008]***
Pvalue (Joint Test)		0.8492		0.095		0.000

Standard errors (in brackets) are computed taking survey design into account. Coefficients reported from separate linear regressions, where each characteristic is regressed on a dummy for male and a constant. The coefficient on male is reported here. The p-value for the joint test comes from regressing the youngest child's gender on all the charateristics and testing whether they are jointly significant. ** p<0.05, * p<0.1

TABLE 3. DESCRIPTIVE STATISTICS, TIME USE SURVEY (1998-1999). RURAL AREAS.

	HHs with youngest ages 0-1		HHs with youngest ages 2-5	
	Mean	S.D.	Mean	S.D.
Percentage of all households	0.15		0.30	
Time Use:				
Time spent on child care (minutes per day)	196.90	152.19	107.19	129.05
Time spent on child care by female members	166.23	132.76	88.87	108.79
Time spent on child care by male members	30.67	64.70	18.31	51.21
Time spent on physical care	137.89	121.83	73.06	94.29
Time spent supervising children	48.96	105.50	24.26	76.10
Time spent instructing children	4.10	23.10	5.24	26.62
Time spent taking children to places	5.94	42.81	4.62	39.92
Time spent on exclusive child care	95.55	137.45	57.43	105.52
Household characteristics:				
Household size	4.54	1.83	3.95	1.54
Male youngest	0.51	0.50	0.55	0.50
Scheduled tribe	0.23	0.42	0.21	0.41
Scheduled caste	0.18	0.38	0.20	0.40
Hindu	0.91	0.29	0.92	0.28
Per capita expenditure	393.83	188.92	408.61	196.36
Land owned and possessed	4.51	8.25	3.89	9.13
Observations	1,947		3,815	

Weighted statistics for households in each sample. The statistics below the first row are for the samples of interest, that is, households where the youngest child is under 2 (first two columns) and households where the youngest is between 2 and 5 years of age (last two columns).

TABLE 4. EFFECT OF CHILD GENDER ON HOUSEHOLD CHILD CARE TIME, TIME USE SURVEY (1998-1999)

Sample:	Youngest children under 2					Youngest children 2-5			
Model:	OLS	OLS	Logit	OLS	Tobit	OLS	Logit	OLS	Tobit
Dependent variable:	Number of minutes per day, including 0s	Number of minutes per day, including 0s	Any care? (Beta reported)	Number of minutes per day>0	Number of minutes per day	Number of minutes per day, including 0s	Any care? (Beta reported)	Number of minutes per day>0	Number of minutes per day
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Male=1	18.689	16.602	-0.052	21.951	18.647	-1.745	-0.013	-1.888	-2.205
	[8.643]**	[8.593]*	[0.182]	[8.629]**	[9.509]**	[5.086]	[0.089]	[5.991]	[6.905]
Controls?	no	yes	no	no	no	no	no	no	no
Obs	1947	1947	1947	1747	1947	3815	3815	2765	3815
Mean Y	196.90	196.90	0.90	219.17	196.90	107.19	0.72	149.79	107.19

Robust standard errors in brackets. The dependent variable in all columns except (3) and (7) is the number of minutes per day spent with child care by all household members. The dependent variable in columns (3) and (7) is an indicator variable for positive childcare time. Columns (1)-(5) report results for households whose youngest child was under 2 years old. Columns (6)-(9) show results for households whose youngest child is between 2 and 5 years old. The controls include dummies for household caste (2 dummies), a dummy for whether the household was Hindu and the area of the land owned and possessed by the household. Survey weights are used in estimation. *** p<0.01, ** p<0.05, * p<0.1

TABLE 5. HETEROGENEITY IN CHILDCARE TIME, TIME USE SURVEY (1998-1999).

	Childcare		Households w/ ONLY 1 child under 6	Physical Care		Supervising		Exclusive Care
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Male = 1	18.689	49.654	50.777	14.673	20.442	10.643	35.414	37.278
	[8.643]**	[14.179]***	[15.969]***	[6.597]**	[11.150]*	[6.077]*	[9.874]***	[12.577]***
Male* (# other children under 6)		-25.206			-4.725		-20.231	-23.943
		[9.042]***			[7.181]		[6.003]***	[7.496]***
# Other children under 6		12.439			1.606		8.315	-3.696
		[5.904]**			[4.471]		[4.061]**	[5.001]
Constant	187.332	171.739	168.946	130.381	128.368	43.516	33.092	95.753
	[6.143]***	[9.715]***	[10.742]***	[4.346]***	[7.393]***	[3.723]***	[5.369]***	[8.097]***
Observations	1947	1947	481	1947	1947	1947	1947	1947
R-squared	0.004	0.01	0.028	0.004	0.004	0.003	0.011	0.019

Robust standard errors in brackets. The sample is restricted to households whose youngest child was under 2 years old. The dependent variable in columns (1) and (2) is the number of minutes per day spent with child care. In column (3), the dependent variable is childcare time and the sample is further restricted to households with only one child under 6. The dependent variable in columns (4) and (5) is the amount of time spent taking physical care of children (e.g., washing, dressing and feeding). The dependent variable in columns (6) and (7) is the amount of time spent supervising children. The dependent variable in column (8) is the amount of time spent exclusively on childcare -- i.e., the caretaker was not multi-tasking. The variable "# Other children under 6" excludes the youngest child. Its mean is equal to 1.23 children. All columns are estimated using OLS. Survey weights are used for estimation. *** p<0.01, ** p<0.05, * p<0.1

TABLE 6: EFFECT OF CHILD GENDER ON HOUSEHOLD CHILDCARE TIME BY DEMOGRAPHIC GROUP, TIME USE SURVEY (1998-1999)

Dependent variable:	Childcare by females 15 and older				Childcare by males 15 and older				Childcare by females 14 and younger				Minutes of care, urban households	
	Any care?	Any care?	minutes of care≥0	minutes of care≥0	Any care?	Any care?	minutes of care≥0	minutes of care≥0	Any care?	Any care?	minutes of care≥0	minutes of care≥0		
Model:	Probit	Probit	OLS	OLS	Probit	Probit	OLS	OLS	Probit	Probit	OLS	OLS	OLS	OLS
Male = 1	-0.002	0.062	14.507	33.564	0.02	0.075	4.208	10.104	0.086	0.2	15.334	67.198	-15.867	-17.886
	[0.017]	[0.032]**	[6.696]**	[11.627]***	[0.026]	[0.042]*	[2.887]	[4.486]**	[0.066]	[0.113]*	[14.516]	[27.503]**	[16.173]	[25.307]
Male * (# other children under 6)		-0.051		-15.668		-0.045		-4.752		-0.082		-36.441		2.135
		[0.020]**		[7.515]**		[0.028]		[2.876]*		[0.062]		[14.866]**		[18.850]
# other children under 6		0.022		3.253		0.022		3.327		0.027		18.503		-0.327
		[0.014]		[4.999]		[0.020]		[2.035]		[0.044]		[9.141]**		[13.904]
Constant			148.327	144.247			24.458	20.281			48.055	22.041	235.373	235.685
			[4.600]***	[7.984]***			[1.871]***	[2.840]***			[8.148]***	[10.417]**	[11.733]***	[20.446]***
Observations	1936	1936	1936	1936	1907	1907	1907	1907	408	408	408	408	677	677
R-squared			0.003	0.008			0.002	0.004			0.006	0.032	0.003	0.003

Robust standard errors in brackets. The dependent variable in columns (1), (3) and (5) is an indicator for whether household members of a given demographic group reported spending time taking care of children. The dependent variable in columns (2), (4) and (6) is the number of minutes per day spent with child care by all household members of a given demographic group. The sample is restricted to households whose youngest child is under 2. Columns (1) to (6) show results for rural households. Column (7) shows results for urban households. The dependent variable in column (7) is the amount of time spent with childcare by all household members. The variable "# Other children under 6" excludes the youngest child. Its mean is equal to 1.23 children. Survey weights are used for estimation. *** p<0.01, ** p<0.05, * p<0.1

TABLE 7: EFFECT OF CHILD GENDER ON PARENTAL INPUTS, DHS (1992 AND 1998)

Dependent variable:		Was child ever breastfed?		# months breastfed	log(# months breastfed)	# months breastfed		Vitamin A supplement?		Did mother have vaccination card at interview?	
Model:	controls?	OLS	Logit (beta reported)	censored regression	censored regression	Accelerated Failure Time model	Proportional Hazard Model	OLS	Logit (beta reported)	OLS	Logit (beta reported)
Panel A: Youngest kids 0-12 months old											
				(# censored obs: 16,140)							
Male = 1	no	0.004	0.072	0.12	0.029	0.037	-0.038	0.009	0.109	0.033	0.157
		[0.004]	[0.079]	[0.510]	[0.097]	[0.139]	[0.139]	[0.005]*	[0.065]*	[0.008]***	[0.038]***
Male = 1	yes	0.004	0.078	0.317	0.064	0.065	-0.074	0.008	0.109	0.031	0.169
		[0.004]	[0.079]	[0.501]	[0.097]	[0.138]	[0.138]	[0.005]*	[0.066]*	[0.007]***	[0.041]***
Obs		17,369		16,444	16,444	16,444	16,444	16,929		17,372	
Mean of Y		0.946		5.981	1.574			0.0869		0.293	
Panel B: Youngest kids 0-23 months old											
				(# censored obs: 27,733)							
Male = 1	no	0.003	0.085	0.705	0.068	0.128	-0.13	0.011	0.089	0.031	0.152
		[0.002]	[0.072]	[0.336]**	[0.035]*	[0.055]**	[0.056]**	[0.005]**	[0.038]**	[0.006]***	[0.029]***
Male = 1	yes	0.003	0.088	0.872	0.087	0.16	-0.166	0.009	0.074	0.028	0.148
		[0.002]	[0.072]	[0.326]***	[0.034]**	[0.054]***	[0.056]***	[0.005]*	[0.039]*	[0.006]***	[0.031]***
Obs		30,940		29,866	29,866	29,866	29,866	29,861		30,962	
Mean of Y		0.966		10.77	2.118			0.152		0.293	

Standard errors [in brackets] are computed taking survey design into account. Child ever breastfed is equal to zero if mother reports that child was not breastfed or if breastfeeding duration was less than a month. Each coefficient corresponds to a separate estimation, and survey weights are used. The number of observations for each age group varies from outcome to outcome because there are a few missing values. Controls include all variables in Table 3: # of brothers, # of sisters, birth month, mother's age, mother's caste (2 dummies), mother's religion (3 dummies), mother's years of education, whether mother was born in rural area, mother's age at first marriage, mother's age at first birth, and whether mother speaks Hindi. *** p<0.01, ** p<0.05, * p<0.1

TABLE 8: EFFECT OF CHILD GENDER ON VACCINATIONS, OLS, DHS (1992 AND 1998)

	controls?	BCG	DPT 1st dose	DPT 2nd dose	DPT 3rd dose	Polio 1st dose	Polio 2nd dose	Polio 3rd dose	Measles
		From vaccination cards or mothers							
Panel A: Youngest kids 0-12 months old									
Male = 1	no	0.039	0.046	0.034	0.031	0.038	0.031	0.03	0.017
		[0.009]***	[0.009]***	[0.008]***	[0.008]***	[0.009]***	[0.009]***	[0.008]***	[0.005]***
Male = 1	yes	0.037	0.044	0.032	0.03	0.037	0.029	0.028	0.017
		[0.008]***	[0.008]***	[0.008]***	[0.007]***	[0.008]***	[0.008]***	[0.007]***	[0.005]***
Obs		17325	17304	17299	17299	17355	17354	17354	17151
Mean of Y		0.465	0.461	0.337	0.233	0.524	0.381	0.24	0.1
Panel B: Youngest kids 0-23 months old									
Male = 1	no	0.037	0.046	0.039	0.034	0.039	0.035	0.028	0.022
		[0.007]***	[0.007]***	[0.007]***	[0.007]***	[0.006]***	[0.007]***	[0.006]***	[0.006]***
Male = 1	yes	0.032	0.041	0.034	0.029	0.035	0.031	0.024	0.018
		[0.006]***	[0.006]***	[0.006]***	[0.006]***	[0.006]***	[0.006]***	[0.006]***	[0.005]***
Obs		30860	30771	30759	30759	30922	30912	30912	30384
Mean of Y		0.535	0.544	0.443	0.346	0.608	0.501	0.367	0.242
Panel C: Youngest kids--4-23 months old (DPT and Polio), 9-23 months old (measles)									
Male = 1	no		0.05	0.043	0.04	0.043	0.038	0.033	0.036
			[0.007]***	[0.007]***	[0.008]***	[0.007]***	[0.007]***	[0.007]***	[0.009]***
Obs			25312	25300	25300	25460	25450	25450	17715
Mean of Y			0.62	0.527	0.417	0.685	0.593	0.443	0.39
		From vaccination card only							
Panel D: Youngest kids 0-12 months old									
Male = 1	no	-0.002	0.004	0.005	0.023	0.003	0.008	0.028	0.02
		[0.011]	[0.009]	[0.016]	[0.017]	[0.009]	[0.016]	[0.017]*	[0.013]
Male = 1	yes	0.002	0.004	0.009	0.029	0.004	0.012	0.035	0.022
		[0.011]	[0.009]	[0.016]	[0.017]*	[0.009]	[0.016]	[0.017]**	[0.013]*
Obs		5201	5201	5201	5201	5201	5201	5201	5201
Mean of Y		0.885	0.925	0.702	0.509	0.916	0.705	0.51	0.179
Panel E: Youngest kids 0-23 months old									
Male = 1	no	-0.006	-0.001	-0.001	0.008	0	0.001	0.011	0.005
		[0.008]	[0.006]	[0.010]	[0.012]	[0.006]	[0.010]	[0.012]	[0.012]
Male = 1	yes	-0.005	-0.001	0.001	0.011	0	0.003	0.015	0.01
		[0.008]	[0.006]	[0.010]	[0.012]	[0.006]	[0.010]	[0.012]	[0.012]
Obs		9250	9250	9250	9250	9250	9250	9250	9250
Mean of Y		0.899	0.95	0.795	0.653	0.942	0.796	0.654	0.396

Standard errors [in brackets] are computed taking survey design into account. Each coefficient corresponds to a separate linear regression of the dependent variable on a dummy variable equal to one if the child is a boy. Controls include all variables in Table 3: # of brothers, # of sisters, birth month, mother's age, mother's caste (2 dummies), mother's religion (3 dummies), mother's years of education, whether mother was born in rural area, mother's age at first marriage, mother's age at first birth, and whether mother speaks Hindi. Survey weights are used for estimation. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

TABLE 9: EFFECT OF CHILD GENDER ON ANTHROPROMETRIC MEASURES, DHS (1992 AND 1998)

Dependent variable:		Height-for-age Z score		Stunted = 1 (height-for-age Z score< 2 s.d. below reference median)		Weight-for-age Z score		Underweight=1 (weight-for-age Z score< 2 s.d. below reference median)		Weight-for-height Z score		Wasted=1 (Weight-for-height Z score< 2 s.d. below reference median)	
	controls?	DHS	UK	DHS	UK	DHS	UK	DHS	UK	DHS	CDC	DHS	CDC
Panel A: Youngest kids 0-12 months old													
Male = 1	no	-0.108	0.011	0.023	-0.01	-0.07	0.086	0.007	-0.027	-0.02	0.086	0.019	-0.007
		[0.032]***	[0.038]	[0.010]**	[0.011]	[0.026]***	[0.031]***	[0.009]	[0.011]**	[0.028]	[0.037]**	[0.007]***	[0.010]
Male = 1	yes	-0.113	0.003	0.024	-0.008	-0.071	0.084	0.007	-0.026	-0.017	0.087	0.018	-0.007
		[0.031]***	[0.037]	[0.010]**	[0.011]	[0.025]***	[0.030]***	[0.009]	[0.010]**	[0.028]	[0.037]**	[0.007]***	[0.010]
Obs		11235	11235	11235	11235	12959	12959	12959	12959	11129	11129	11129	11129
Mean of Y		-1.064	-1.113	0.254	0.302	-1.247	-1.784	0.286	0.451	-0.518	-1.064	0.0943	0.267
Panel B: Youngest kids 0-23 months old													
Male = 1	no	-0.075	0.057	0.013	-0.02	-0.1	0.077	0.017	-0.021	-0.05	0.069	0.031	-0.011
		[0.025]***	[0.028]**	[0.008]	[0.009]**	[0.019]***	[0.023]***	[0.008]**	[0.008]***	[0.021]**	[0.027]**	[0.007]***	[0.008]
Male = 1	yes	-0.081	0.049	0.015	-0.018	-0.1	0.074	0.017	-0.021	-0.047	0.068	0.03	-0.011
		[0.024]***	[0.027]*	[0.008]*	[0.008]**	[0.019]***	[0.022]***	[0.008]**	[0.008]***	[0.021]**	[0.027]**	[0.007]***	[0.008]
Obs		20219	20219	20219	20219	23162	23162	23162	23162	20508	20508	20508	20508
Mean of Y		-1.563	-1.59	0.402	0.427	-1.655	-2.121	0.437	0.555	-0.831	-1.244	0.164	0.307

Standard errors [in brackets] are computed taking survey design into account. Each coefficient corresponds to a separate estimation. Controls include all variables in Table 3: # of brothers, # of sisters, birth month, mother's age, mother's caste (2 dummies), mother's religion (3 dummies), mother's years of education, whether mother was born in rural area, mother's age at first marriage, mother's age at first birth, and whether mother speaks Hindi. The other measures are standardized using the UK (1990) standards or the 2000 CDC standards. The UK standards are not available for height for age. CDC standards for height are not available for children under 2. Survey weights are used in estimation. For children under 2 the standards use length rather than height, which is measured while lying instead of standing. *** p<0.01, ** p<0.05, * p<0.1

TABLE 10: EFFECT OF CHILD'S GENDER ON LIVING ARRANGEMENTS, YOUNGEST CHILDREN 23 MONTHS AND YOUNGER. RURAL HOUSEHOLDS.

Panel A: Effect of gender on family living arrangements in the DHS (1992 & 1998)

Dependent variable:	Husband lives home?		Is mother the wife of the household head?		# of other sibs living at home		# of sisters living at home		# of brothers living at home	
	OLS	Logit (beta reported)	OLS	Logit (beta reported)	OLS	Negative binomial (IRR reported)	OLS	Negative binomial (IRR reported)	OLS	Negative binomial (IRR reported)
Male=1	0.004	0.05	0.001	0.003	0.004	1.0026	0.035	1.04	-0.031	0.96
	[0.004]	[0.048]	[0.007]	[0.026]	[0.021]	[0.013]	[0.014]**	[0.017]**	[0.013]**	[0.017]**
Obs	30648	30648	30982	30982	30985	30985	30985	30985	30985	30985
Mean of Y	0.912	0.912	0.48	0.48	1.579	1.579	0.83	0.83	0.749	0.749

Panel B: Effect of gender on household composition TUS (1998-1999)

Dependent variable:	# Men 15 and older		# Women 15 and older		# Children 14 and younger		# Girls 14 and younger		# Boys 14 and younger	
	OLS	Poisson (IRR reported)	OLS	Poisson (IRR reported)	OLS	Poisson (IRR reported)	OLS	Poisson (IRR reported)	OLS	Poisson (IRR reported)
Male=1	0.059	1.043	0.098	1.071	-0.085	0.951	0.011	1.012	-0.096	0.891
	[0.04]	[0.03]	[0.036]***	[0.027]***	[0.07]	[0.039]	[0.052]	[0.061]	[0.051]*	[0.055]*
Obs	1947	1947	1947	1947	1947	1947	1947	1947	1947	1947
Mean of Y	1.41	1.41	1.43	1.43	1.70	1.70	0.83	0.83	0.87	0.87

The standard errors [in brackets] are computed taking survey design into account in the DHS and in the TUS they allow for heteroskedasticity. Each coefficient corresponds to a separate estimation, where the dummy for the youngest child's gender is the only covariate. In the TUS (Panel B) we estimated Possion rather than negative binomial models because some of the negative binomial models in the TUS did not converge. Survey weights are used in estimation. *** p<0.01, ** p<0.05, * p<0.1

Appendix 1

Claim: Let $E[S|N]$ be the desired number of sons of the family of the average child in a family with N children and $E[S|N,G]$ the desired number of sons of the family of the average girl in a family with N children (G is an indicator variable for whether the child is a girl). If families continue to have children until they achieve their desired number of sons S, then $E[S|N] \geq E[S|N,G]$.

Proof: It is sufficient to show that $p(S \leq s|N)$ first-order stochastically dominates $p(S \leq s|N,G)$. By definition, we have:

$$p(S=s|N)=p(S=s|N,G)\ p(G|N)+p(B|S,N)\ p(S=s|N),$$

which can be rewriten as:

$$p(S=s|N)=p(S=s|N,G)*[p(G|N)/p(G|S,N)].$$

Finally, notice that the term between brackets is increasing in S since $p(G|S,N)$ is decreasing in S. QED.

Appendix 2

We consider a model in which there is heterogeneity in the way that households treat boys and girls. Formally, the model can be written as:

$$Z_{ih} = \delta_{0,h}\left(1 - B_{ih}\right) + \delta_{1,h} B_{ih} + \xi_{ih}, \tag{1}$$

where Z_{ih} is the investment in child i in household h, $\delta_{0,h}$ ($\delta_{1,h}$) is the average investment made by household h in girls (boys), B_{ih} is a dummy that is equal to 1 if child i in household h is a boy and ν_{ih} is an error term.

Unconditional

Suppose we run a regression of Z_{ih} on a constant and B_{ih}. In this case, the OLS bias is given by:

$$\frac{Cov\left(\xi_{ih}, B_{ih}\right)}{Var\left(B_{ih}\right)}$$

because the average boy and the average girl are in families with identical preferences for children – i.e., $E\left[\delta_{0,h}|B_{ih}\right] = E\left[\delta_{0,h}\right]$ and $E\left[\delta_{1,h}|B_{ih}\right] = E\left[\delta_{1,h}\right]$. However, because girls are on average in larger families than boys, the OLS estimate is biased if N_h is part of the error term ξ_{ih}.

Conditional on Family Size

We can rewrite (1) as:

$$Z_{ih} = E\left[\delta_{0,h}|N_h\right] + \left(E\left[\delta_{1,h}|N_h\right] - E\left[\delta_{0,h}|N_h\right]\right) B_{ih} + \nu_{ih} + \xi_{ih}, \tag{2}$$

where

$$\nu_{ih} = \delta_{0,h} - E\left[\delta_{0,h}|N_h\right] + \left[\left(\delta_{1,h} - E\left[\delta_{1,h}|N_h\right]\right) - \left(\delta_{0,h} - E\left[\delta_{0,h}|N_h\right]\right)\right] B_{ih}. \tag{3}$$

Suppose we run a regression of Z_{ih} on a constant and B_{ih}, controlling for family size N_h. In this case, the OLS bias is given by:

$$\frac{Cov\left(\nu_{ih}, B_{ih}|N_h\right)}{Var\left(B_{ih}|N_h\right)} + \frac{Cov\left(\xi_{ih}, B_{ih}|N_h\right)}{Var\left(B_{ih}|N_h\right)}.$$

Here we are interested in the bias that may arise from how families with different preferences treat boys and girls so we will concentrate in the first term. Let us write:

$$\frac{Cov\left(\nu_{ih}, B_{ih}|N_h\right)}{Var\left(B_{ih}|N_h\right)} = \frac{E\left[\left(\nu_{ih} - E\left[\nu_{ih}\right]\right) B_{ih}|N_h\right]}{Var\left(B_{ih}|N_h\right)}$$

and the law of iterated expectations implies:

$$\begin{aligned}\frac{Cov\left(\nu_{ih}, B_{ih}|N_h\right)}{Var\left(B_{ih}|N_h\right)} &= \frac{E\left[E\left[\left(\nu_{ih} - E\left[\nu_{ih}\right]\right) B_{ih}|B_{ih}, N_h\right]\right]}{Var\left(B_{ih}|N_h\right)} = \\ &= \frac{pr\left(B_{ih} = 1|N_h\right)}{Var\left(B_{ih}|N_h\right)} E\left[\left(\nu_{ih} - E\left[\nu_{ih}\right]\right)|B_{ih} = 1, N_h\right]. \end{aligned} \tag{4}$$

Furthermore, notice that we can rewrite $E\left[\nu_{ih}|N_h\right]$ as:

$$\begin{aligned} E\left[\nu_{ih}|N_h\right] &= pr\left(B_{ih} = 1|N_h\right) E\left[\nu_{ih}|B_{ih} = 1, N_h\right] + \\ &\quad + pr\left(B_{ih} = 0|N_h\right) E\left[\nu_{ih}|B_{ih} = 0, N_h\right]. \end{aligned} \tag{5}$$

Substituting (5) into (4) yields:

$$\begin{aligned}\frac{Cov\left(\nu_{ih},B_{ih}|N_h\right)}{Var\left(B_{ih}|N_h\right)} &= \kappa\left\{E\left[\nu_{ih}|B_{ih}=1,N_h\right]-E\left[\delta_{0,h}|B_{ih}=0,N_h\right]+E\left[\delta_{0,h}|,N_h\right]\right\}= \\ &= \kappa\left\{\left(E\left[\delta_{1,h}|B_{ih}=1,N_h\right]-E\left[\delta_{1,h}|N_h\right]\right)-\left(E\left[\delta_{0,h}|B_{ih}=0,N_h\right]-E\left[\delta_{0,h}|N_h\right]\right)\right\},\end{aligned} \tag{6}$$

where

$$\kappa=\frac{pr\left(B_{ih}=1|N_h\right)\left[1-pr\left(B_{ih}=1|N_h\right)\right]}{Var\left(B_{ih}|N_h\right)}. \tag{7}$$

Therefore, the sign of the bias is determined by the expression between curly brackets in (6):

$$\left(E\left[\delta_{1,h}|B_{ih}=1,N_h\right]-E\left[\delta_{1,h}|N_h\right]\right)-\left(E\left[\delta_{0,h}|B_{ih}=0,N_h\right]-E\left[\delta_{0,h}|N_h\right]\right).$$

Notice that the same formula applies if one conditions on a vector X_{ih} that contains family size N_h.

Appendix 3: Sampling and Clustering

Indian Time Use Survey

The Indian Time Use Survey (ITUS) adopted a sophisticated sampling scheme with three levels of stratification and clustering: districts, villages and households.[1] Pandey (2000) provides a detailed description of the ITUS sampling procedure. First, all districts in a given state were grouped into four district-level strata and districts were randomly selected within each district-level stratum. Second, all the villages in a given selected district were grouped into three village-level strata and villages were selected within each village-level stratum. Finally, all households in a selected village were grouped into six household-level strata and households were randomly selected within each household-level stratum.

For our purposes, it is important to notice that the survey intended to interview twelve households per village and that these households were distributed into six different strata according to the type of job (rural labor, self-employed and other) and land possession (less than two acres, between two and seven acres and more than seven acres). As a consequence, there were multiple household-level strata with a single sampling unit. To calculate standard errors, it is necessary to "remove" these sampling units to different strata. Something similar applies to village-level strata, in which case a village may be the only village in a village-level stratum. Here we explain the procedure we used to remove these special sampling units.

The algorithm is run at the household level, then at the village level and finally at the district level.[2]

Household level

A1) Identify households that are the single sampling units in their household-level strata
A2) Find the mode household-level stratum for each village
A3) Remove households identified in (A1) to (A2)

B1) Identify households that are the single sampling units in their villages
B2) Find the mode village for each village-level stratum
B3) Find the mode household-level stratum for the village in (B2)
B4) Remove households identified in (B1) to the village in (B2) and to the household-level stratum in (B3)

Village level

[1] In the urban areas, there is an additional level of clustering and stratification (Pandey 2000).
[2] Finite population corrections were calculated using: (1) the number of villages in each state (according to the 1991 Census: 6,988 villages in Haryana; 55,842 in Madhya Pradesh; 18,509 in Gujarat; 51,057 in Orissa; 16,780 in Tamil Nadu and 5,629 in Meghalaya); (2) estimates of the number of households in each district – which were calculated using survey weights – and (3) the number of districts for each state is (available in Pandey 2000).

A1) Identify villages that are the single sampling units in their village-level strata
A2) Find the mode village-level stratum for each district
A3) Remove villages identified in (A1) to (A2)

B1) Identify villages that are the single sampling units in their districts
B2) Find the mode district for each district-level stratum
B3) Find the mode village-level stratum for the district in (B2)
B4) Remove villages identified in (B1) to the district in (B2) and to the village-level stratum in (B3)

District level
A1) Identify districts that are the single sampling units in their district-level strata
A2) Find the mode district-level stratum for each state
A3) Remove districts identified in (A1) to (A2)

Indian Demographic and Health Surveys (DHS) 1992 and 1998

The Indian DHS (ITUS) adopted a two stages sampling scheme for rural households. In each state, villages were the primary sampling units and households were selected in the second state within each village. If the household is the single household in its village, then its removed to the mode village in the state.

APPENDIX TABLE 1: TESTING IDENTIFYING ASSUMPTION, TIME USE SURVEY (1998-1999)

Sample:	**Youngest child**					
	Age 0-1 years (N=1,947)		**Age 2-3 years (N=2,233)**		**Age 4-5 years (N=1,582)**	
Dependent variable:	Male=1		Male=1		Male=1	
	beta	se	beta	se	beta	se
Scheduled Tribe	-0.035	[0.023]	-0.020	[0.021]	-0.007	[0.025]
Scheduled Caste	-0.024	[0.020]	0.016	[0.020]	0.002	[0.024]
Hindu	0.005	[0.015]	0.038	[0.014]***	0.028	[0.016]*
Land Owned (in acres)	0.336	[0.452]	-0.163	[0.394]	1.47	[0.515]***
Pvalue (Joint Test)		0.559		0.047		0.012

Standard errors (in brackets). Coefficients reported from separate linear regressions, where a dummy for the gender of the youngest child is regressed on a characteristic and a constant. The coefficient on each characteristic is reported here. The p-value for the joint test comes from regressing the youngest child's gender on all the charateristics and testing whether they are jointly significant. ** p<0.05, * p<0.1

APPENDIX TABLE 2 . EFFECT OF CHILD GENDER ON HOUSEHOLD CHILD CARE TIME BY YOUNGEST AGE, TIME USE SURVEY (1998-1999)

Sample:	Youngest children under 6				
Model:	OLS	OLS	Logit	OLS	Tobit
Dependent variable:	Number of minutes per day, including 0s	Number of minutes per day, including 0s	Any care? (Beta reported)	Number of minutes per day>0	Number of minutes per day
	(1)	(2)	(3)	(4)	(5)
Male age 0=1	32.772	31.316	0.613	24.226	36.064
	[17.656]*	[17.501]*	[0.396]	[17.333]	[18.796]*
Male age 1=1	17.486	16.237	-0.2	24.236	16.202
	[9.297]*	[9.226]*	[0.201]	[9.350]***	[10.361]
Male age 2=1	-7.284	-7.369	-0.114	-5.258	-9.057
	[10.089]	[10.081]	[0.189]	[10.721]	[11.844]
Male age 3=1	3.584	2.477	-0.122	8.924	1.531
	[10.887]	[10.827]	[0.190]	[12.139]	[13.721]
Male age 4=1	-7.478	-8.108	0.005	-11.296	-8.171
	[9.823]	[9.852]	[0.213]	[12.303]	[14.894]
Male age 5=1	11.82	10.444	0.226	10.882	22.789
	[6.261]*	[6.249]*	[0.155]	[8.882]	[12.193]*
Controls?	no	yes	no	no	no
Observations	5762	5762	5762	4512	5762

Robust standard errors in brackets. All the regression include age dummies. The dependent variable in all columns except (3) is the number of minutes per day spent with child care by all household members. The dependent variable in columns (3) is an indicator variable for positive childcare time.Results for households whose youngest child was under 6 years old. The controls include dummies for household caste (2 dummies), a dummy for whether the household was Hindu and the area of the land owned and possessed by the household. Survey weights are used in estimation. *** p<0.01, ** p<0.05, * p<0.1

APPENDIX TABLE 3. EFFECT OF CHILD GENDER ON HOUSEHOLD CHILD CARE TIME, TIME USE SURVEY (1998-1999)

Sample:	Youngest children under 2					Youngest children 2-5			
Model:	OLS	OLS	Logit	OLS	Tobit	OLS	Logit	OLS	Tobit
Dependent variable:	Number of minutes per day, including 0s	Number of minutes per day, including 0s	Any care? (Beta reported)	Number of minutes per day>0	Number of minutes per day	Number of minutes per day, including 0s	Any care? (Beta reported)	Number of minutes per day>0	Number of minutes per day
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Male=1	18.689	16.602	-0.052	21.951	18.647	-1.745	-0.013	-1.888	-2.205
OLS se	[8.643]**	[8.593]*	[0.182]	[8.629]**	[9.509]**	[5.086]	[0.089]	[5.991]	[6.905]
svy se	[9.790]*	[9.718]*	[0.186]	[10.411]**	[10.651]*	[7.179]	[0.116]	[7.537]	[9.928]
Controls?	no	yes	no	no	no	no	no	no	no
Observations	1947	1947	1947	1747	1947	3815	3815	2765	3815
Mean Y	196.90	196.90	0.90	219.17	196.90	107.19	0.72	149.79	107.19

The table presents OLS standards errors and standard errors that are estimated taking the sampling scheme into account (in the row labeled "svy se"). The Indian Time Use Survey used a sophisticated sampling scheme with three levels of stratification and clustering. As a consequence, there are multiple strata with a single sampling unit. In order to compute the standard errors, we replace these sampling units to another stratum. For example, if a household is the single household in a given household-level stratum, we replace it to another household-level stratum in the same village. If the household is the single household in the village, then we replace it to another village in the same district. We employ a similar strategy for villages and districts (which are the other two levels of clusters). The procedure is explained in more details in Appendix 2. See footnote of Table 4 for more details. *** p<0.01, ** p<0.05, * p<0.1

APPENDIX TABLE 4: EFFECT OF CHILD GENDER ON VACCINATIONS, LOGIT, DHS (1992 AND 1998)

		From vaccination cards or mothers							
	controls?	BCG	DPT 1st dose	DPT 2nd dose	DPT 3rd dose	Polio 1st dose	Polio 2nd dose	Polio 3rd dose	Measles
Panel A: Youngest kids 0-12 months old									
Male = 1	no	0.158	0.185	0.152	0.177	0.152	0.131	0.163	0.194
		[0.035]***	[0.036]***	[0.038]***	[0.044]***	[0.035]***	[0.036]***	[0.042]***	[0.061]***
Male = 1	yes	0.151	0.197	0.154	0.159	0.147	0.123	0.124	0.127
		[0.035]***	[0.036]***	[0.037]***	[0.042]***	[0.034]***	[0.034]***	[0.039]***	[0.050]**
Obs		17325	17304	17299	17299	17355	17354	17354	17151
Mean of Y		0.465	0.461	0.337	0.233	0.524	0.381	0.24	0.1
Panel B: Youngest kids 0-23 months old									
Male = 1	no	0.147	0.184	0.157	0.149	0.162	0.14	0.123	0.118
		[0.026]***	[0.027]***	[0.027]***	[0.029]***	[0.027]***	[0.027]***	[0.028]***	[0.031]***
Male = 1	yes	0.154	0.195	0.162	0.15	0.164	0.138	0.117	0.106
		[0.029]***	[0.029]***	[0.029]***	[0.032]***	[0.028]***	[0.028]***	[0.029]***	[0.033]***
Obs		30860	30771	30759	30759	30922	30912	30912	30384
Mean of Y		0.539	0.546	0.448	0.347	0.609	0.504	0.369	0.242
		From vaccination card only							
Panel C: Youngest kids 0-12 months old									
Male = 1	no	-0.018	0.055	0.023	0.092	0.042	0.036	0.114	0.136
		[0.107]	[0.127]	[0.075]	[0.067]	[0.121]	[0.076]	[0.068]*	[0.090]
Male = 1	yes	-0.028	0.052	0.025	0.1	0.032	0.038	0.115	0.09
		[0.102]	[0.126]	[0.073]	[0.064]	[0.118]	[0.073]	[0.064]*	[0.074]
Obs		5201	5201	5201	5201	5201	5201	5201	5201
Mean of Y		0.885	0.925	0.702	0.509	0.916	0.705	0.51	0.179
Panel D: Youngest kids 0-23 months old									
Male = 1	no	-0.072	-0.025	-0.008	0.034	-0.005	0.005	0.049	0.02
		[0.084]	[0.116]	[0.064]	[0.053]	[0.106]	[0.063]	[0.053]	[0.051]
Male = 1	yes	-0.048	-0.02	0.009	0.053	-0.005	0.02	0.069	0.041
		[0.086]	[0.118]	[0.064]	[0.054]	[0.107]	[0.064]	[0.054]	[0.051]
Obs		9250	9250	9250	9250	9250	9250	9250	9250
Mean of Y		0.908	0.956	0.811	0.668	0.95	0.811	0.667	0.4

Standard errors [in brackets] are computed taking survey design into account. Each coefficient corresponds to a separate logit of the dependent variable on a dummy variable equal to one if the child is a boy. Controls include all variables in Table 3: # of brothers, # of sisters, birth month, mother's age, mother's caste (2 dummies), mother's religion (3 dummies), mother's years of education, whether mother was born in rural area, mother's age at first marriage, mother's age at first birth, and whether mother speaks Hindi. Survey weights are used for estimation. *** p<0.01, ** p<0.05, * p<0.1

APPENDIX TABLE 5: ADDITIONAL RESULTS ON EFFECT OF GENDER ON PARENTAL INVESTMENTS. CHILDREN AGES 0-23 MONTHS

Dependent variable:	Model:	Live schildren			Estimated betas, values for dead imputed			Urban children			
		Coefficient on I(male=1), [s.e]		Mean Y	Upper bound	Lower bound	Obs	Coefficient on I(male=1), [s.e]		obs	Mean Y
Was child ever breastfed?	OLS	0.003	[0.002]	0.97	-0.002	-0.006	32814	0.002	[0.005]	11508	0.96
# months breastfed	cens. reg.	0.705	[0.336]**	10.77	1.42	-1.055	31670	0.496	[0.393]	11038	10.27
log(# months breastfed)	cens. reg.	0.068	[0.035]*	2.20	0.147	-0.113	31670	0.051	[0.042]	11038	2.07
	AFT	0.128	[0.055]**		0.078	-0.008	31740	0.091	[0.053]*	11038	
# months breastfed	PH	-0.13	[0.056]**		-0.104	0.022	31740	-0.103	[0.054]*	11038	
Vitamin A supplement?	OLS	0.011	[0.005]**	0.15	0.04	-0.017	31735	0.012	[0.010]	11102	0.23
Did mother have vaccination card?	OLS	0.031	[0.006]***	0.29	0.055	0.003	32836	0.009	[0.011]	11529	0.43
BCG	OLS	0.037	[0.007]***	0.54	0.058	0.008	32734	0.015	[0.010]	11517	0.75
DPT 1st dose	OLS	0.046	[0.007]***	0.54	0.066	0.016	32645	0.003	[0.010]	11495	0.73
DPT 2nd dose	OLS	0.039	[0.007]***	0.44	0.06	0.01	32633	0.004	[0.011]	11485	0.64
DPT 3rd dose	OLS	0.034	[0.007]***	0.35	0.056	0.005	32633	0.007	[0.012]	11485	0.54
Polio 1st dose	OLS	0.039	[0.006]***	0.61	0.06	0.009	32796	-0.001	[0.010]	11515	0.76
Polio 2nd dose	OLS	0.035	[0.007]***	0.50	0.056	0.006	32786	0.003	[0.011]	11508	0.67
Polio 3rd dose	OLS	0.028	[0.006]***	0.37	0.051	0.001	32786	0.001	[0.012]	11508	0.54
Measles	OLS	0.022	[0.006]***	0.24							

Each coefficient corresponds to a separate estimation, and survey weights are used. No controls are included. The number of observations for each age group varies from outcome to outcome because there are a few missing values. Cens. reg. is a censored regression, AFT is an accelerated failure time model, and PH is a proportional hazard model. Upper bounds assume that all dead girls would have not received inputs (for dummy variables) or would have been given the 25th percentile of the girls' outcomes distribution. For boys we assume that had they lived they would all have been given inputs (for dummye variables) or given the 75th percentile of boys' outcome distribution. For upper bounds we assume the opposite. The urban sample was constructed using the same restrictions as our main estimation sample--we dropped twins and individuals with missing predetermined covariates.

APPENDIX TABLE 6: RESULTS ON EFFECT OF GENDER ON PARENTAL INVESTMENTS BY AGE. CHILDREN AGES 0-47 MONTHS IN 1992

Sample:		Children 0-11 months			Children 12-23 months			Children 24-35 months			Children 36-47 months		
Dependent variable:	Model:	Coefficient on I(male=1), [s.e]		mean	Coefficient on I(male=1), [s.e]		mean	Coefficient on I(male=1), [s.e]		mean	Coefficient on I(male=1), [s.e]		mean
Was child ever breastfed?	OLS	0.008	[0.006]	0.94	0.003	[0.002]	0.99	-0.004	[0.003]	0.98	0.001	[0.005]	0.98
# months breastfed	cens. reg.	0.849	[0.938]	5.21	1.861	[0.579]***	16.03	1.764	[0.480]***	22.23	1.681	[0.436]***	24.17
log(# months breastfed)	cens. reg.	0.186	[0.210]	1.46	0.141	[0.046]***	2.73	0.101	[0.028]***	2.99	0.063	[0.021]***	3.04
	AFT	0.211	[0.257]	1.46	0.211	[0.068]***	2.73	0.148	[0.036]***	2.99	0.114	[0.028]***	3.04
# months breastfed	PH	-0.216	[0.256]	1.46	-0.213	[0.069]***	2.73	-0.158	[0.037]***	2.99	-0.115	[0.030]***	3.04
Vitamin A supplement?	OLS	0.003	[0.007]	0.09	0.02	[0.011]*	0.20	0.001	[0.011]	0.22	0.042	[0.011]***	0.21
Has vaccination card?	OLS	0.04	[0.011]***	0.27	0.047	[0.012]***	0.29	0.016	[0.011]	0.20	0.033	[0.009]***	0.12
BCG	OLS	0.027	[0.013]**	0.39	0.05	[0.013]***	0.58	0.03	[0.013]**	0.57	0.061	[0.014]***	0.52
DPT 1st dose	OLS	0.037	[0.013]***	0.41	0.063	[0.013]***	0.63	0.034	[0.013]***	0.62	0.067	[0.014]***	0.56
DPT 2nd dose	OLS	0.014	[0.011]	0.28	0.061	[0.013]***	0.55	0.031	[0.013]**	0.55	0.059	[0.014]***	0.50
DPT 3rd dose	OLS	0.012	[0.010]	0.18	0.054	[0.013]***	0.47	0.023	[0.013]*	0.48	0.046	[0.014]***	0.43
Polio 1st dose	OLS	0.038	[0.013]***	0.41	0.066	[0.013]***	0.63	0.039	[0.013]***	0.62	0.068	[0.014]***	0.57
Polio 2nd dose	OLS	0.019	[0.011]*	0.29	0.063	[0.013]***	0.57	0.028	[0.013]**	0.57	0.062	[0.014]***	0.52
Polio 3rd dose	OLS	0.015	[0.010]	0.19	0.048	[0.013]***	0.49	0.015	[0.013]	0.49	0.054	[0.014]***	0.45
Measles	OLS	0.01	[0.006]	0.06	0.045	[0.013]***	0.38	0.026	[0.013]**	0.42	0.049	[0.014]***	0.38

Standard errors [in brackets] are computed taking survey design into account. Child ever breastfed is equal to zero if mother reports that child was not breastfed or if breastfeeding duration was less than a month. Each coefficient corresponds to a separate estimation, and survey weights are used. No controls are included. Cens. reg. is a censored regression, AFT is an accelerated failure time model, and PH is a proportional hazard model. The number of observations for each age group varies from outcome to outcome because there are a few missing values. C*** p<0.01, ** p<0.05, * p<0.1

APPENDIX TABLE 7: EFFECT OF SURVEY DESIGN ON STANDARD ERRORS. CHILDREN AGES 0-23 MONTHS

Dependent variable:	Model:	SE corrected for survey design		SE not corrected		SE clustered at the state level	
		Coefficient on I(male=1), [s.e]		Coefficient on I(male=1), [s.e]		Coefficient on I(male=1), [s.e]	
Was child ever breastfed?	OLS	0.003	[0.002]	0.003	[0.002]	0.003	[0.002]
# months breastfed	cens. reg.	0.705	[0.336]**	0.705	[0.316]**	0.705	[0.217]***
log(# months breastfed)	cens. reg.	0.068	[0.035]*	0.068	[0.033]**	0.068	[0.024]***
Vitamin A supplement?	OLS	0.011	[0.005]**	0.011	[0.005]**	0.011	[0.005]**
Did mother have vaccination card?	OLS	0.031	[0.006]***	0.031	[0.006]***	0.031	[0.008]***
BCG	OLS	0.037	[0.007]***	0.037	[0.007]***	0.037	[0.010]***
DPT 1st dose	OLS	0.046	[0.007]***	0.046	[0.007]***	0.046	[0.011]***
DPT 2nd dose	OLS	0.039	[0.007]***	0.039	[0.007]***	0.039	[0.009]***
DPT 3rd dose	OLS	0.034	[0.007]***	0.034	[0.006]***	0.034	[0.006]***
Polio 1st dose	OLS	0.039	[0.006]***	0.039	[0.007]***	0.039	[0.010]***
Polio 2nd dose	OLS	0.035	[0.007]***	0.035	[0.007]***	0.035	[0.011]***
Polio 3rd dose	OLS	0.028	[0.006]***	0.028	[0.006]***	0.028	[0.008]***
Measles	OLS	0.022	[0.006]***	0.022	[0.006]***	0.022	[0.005]***

Each coefficient corresponds to a separate estimation, and survey weights are used. No controls are included. The number of observations for each age group varies from outcome to outcome because there are a few missing values. Cens. reg. is a censored regression.

APPENDIX TABLE A8: EFFECTS OF SEX-SELECTIVE ABORTION. CHILDREN AGES 0-23 MONTHS

		By group, based on sex ratio at birth in 92 or 98 in the DHS was greater than 110				**By group, based on sex ratio at birth in 92 and 98 in the DHS was greater than 110**				**By group, based on Retherford and Roy's (2003) categorization**			
Dependent variable:	Model:	High SSA states (12 states)		Low SSA states (13 states)		High SSA states: Arunachal Pradesh, Haryana, Himachal Pradesh, Punjab.		Low SSA states (other 21 states)		High SSA states: Delhi, Haryana, Himachal Pradesh, Marahastra,		Low SSA states (other 20 states)	
Was child ever breastfed?	OLS	0.007	[0.004]	0.001	[0.003]	0.004	[0.007]	0.003	[0.002]	0.009	[0.007]	0.002	[0.003]
# months breastfed	cens. reg.	1.022	[0.617]*	0.604	[0.399]	0.362	[0.803]	0.757	[0.357]**	0.637	[0.880]	0.733	[0.364]**
log(# months breastfed)	cens. reg.	0.09	[0.064]	0.062	[0.042]	0.002	[0.091]	0.075	[0.037]**	0.047	[0.091]	0.074	[0.038]*
Vitamin A supplement?	OLS	0.016	[0.010]	0.009	[0.005]*	0.036	[0.016]**	0.01	[0.005]*	0.016	[0.020]	0.01	[0.005]**
Did mother have vaccination card?	OLS	0.056	[0.011]***	0.022	[0.007]***	0.061	[0.017]***	0.03	[0.006]***	0.04	[0.019]**	0.029	[0.006]***
BCG	OLS	0.062	[0.011]***	0.026	[0.008]***	0.067	[0.019]***	0.034	[0.007]***	0.064	[0.017]***	0.031	[0.007]***
DPT 1st dose	OLS	0.069	[0.011]***	0.036	[0.008]***	0.092	[0.019]***	0.043	[0.007]***	0.068	[0.017]***	0.041	[0.007]***
DPT 2nd dose	OLS	0.061	[0.012]***	0.029	[0.008]***	0.098	[0.020]***	0.035	[0.007]***	0.073	[0.019]***	0.032	[0.007]***
DPT 3rd dose	OLS	0.047	[0.012]***	0.027	[0.008]***	0.077	[0.019]***	0.031	[0.007]***	0.054	[0.020]***	0.029	[0.007]***
Polio 1st dose	OLS	0.06	[0.011]***	0.03	[0.008]***	0.082	[0.019]***	0.036	[0.007]***	0.055	[0.015]***	0.035	[0.007]***
Polio 2nd dose	OLS	0.059	[0.011]***	0.026	[0.008]***	0.094	[0.020]***	0.032	[0.007]***	0.069	[0.018]***	0.029	[0.007]***
Polio 3rd dose	OLS	0.045	[0.012]***	0.022	[0.008]***	0.076	[0.019]***	0.026	[0.007]***	0.044	[0.019]**	0.025	[0.007]***
Measles	OLS	0.026	[0.011]**	0.019	[0.007]***	0.054	[0.018]***	0.019	[0.006]***	0.02	[0.020]	0.02	[0.006]***
P value, joint F test that predetermined covariates predict gender at birth		0.0527			0.4258		0.023		0.1918		0.0799		0.2018

Each coefficient corresponds to a separate estimation, and survey weights are used. No controls are included. The number of observations for each age group varies from outcome to outcome because there are a few missing values. Cens. reg. is a censored regression. Sex ratios at birth are reported by state for both rounds of the DHS by Arnold, Kishor and Roy (1998).