

# Why Do Mothers Breastfeed Girls Less Than Boys? Evidence and Implications for Child Health in India\*

Seema Jayachandran  
Stanford University and NBER

Ilyana Kuziemko  
Princeton University and NBER

April 19, 2009

*Preliminary Draft*

## Abstract

Because breastfeeding inhibits fertility, a mother might limit (prolong) the nursing of an infant if she wants to continue (stop) having children. Similarly, if a mother becomes pregnant while still breastfeeding, she often weans the first child sooner than she otherwise would have. We develop a dynamic programming model of breastfeeding as a function of future fertility, and provide support for its predictions using survey data from India. First, breastfeeding increases with birth order, since mothers near or beyond their desired total fertility have greater demand for the contraceptive properties of nursing. Second, if parents have a preference for having sons, mothers with no or few sons are more likely to want to conceive again, so they will wean their current child sooner. Therefore, not only are girls breastfed less than boys in India, but the gender of older siblings also affects how long a child is breastfed. Furthermore, this sex-composition effect is strongest for medium birth-order children: at sufficiently low birth order, a mother will want to have more children (and thus limit breastfeeding) regardless of her children's gender composition, and at sufficiently high birth order she will always want to stop having children. In unsanitary environments, breastfeeding can protect against water- and food-borne disease, and we indeed find that mortality has many of the same relationships with the gender and birth-order interactions as does breastfeeding, and that these relationships are strongest in households without piped water. Our results suggest that the gender gap in breastfeeding leads to 26,000 to 44,000 "missing girls" in India each year.

---

\*We thank Kristin Plys and Luke Stein for excellent research assistance and Anne Case, Angus Deaton, Jeffrey Hammer, Adriana Lleras-Muney, Ben Olken, Eliot Regett and Sam Schulhofer-Wohl for helpful comments. Contact information: jayachan@stanford.edu; kuziemko@princeton.edu

# 1 Introduction

As medical and public health researchers have long documented, breastfeeding inhibits post-natal fertility. The converse also holds: the physical demands of another pregnancy generally cause a mother to stop nursing her current child. These biological constraints suggest a negative relationship between breastfeeding duration and future fertility.<sup>1</sup>

In developing countries, this relationship may be particularly important. First, many women lack access to modern forms of birth control, and thus would rely more on the contraceptive properties of breastfeeding. Second, mothers have more difficulty meeting the high caloric demands of breastfeeding while pregnant in environments characterized by high rates of malnourishment. Finally, breastfeeding is thought to have greater health benefits for infants who would otherwise consume unsafe drinking water and contaminated food.<sup>2</sup>

This paper studies the relationship between breastfeeding and future fertility in India. We test whether factors that likely affect a mother’s decision to have more children—such as her current number of children and their sex composition—affect the length of time she nurses. We also examine the health consequences for children who are breastfed less, e.g., low-birth-order children and girls.

Of course, many factors besides subsequent fertility could affect a mother’s propensity to breastfeed, such as her health, labor-market attachment, and education, or how expensive and available breast-milk substitutes are. To help separate our hypothesis from alternative ones, we specify a dynamic programming model of breastfeeding as a function of desired future fertility. The model makes a number of distinct predictions about breastfeeding patterns with respect to a child’s gender, siblings’ sex composition, and birth order. We then support these predictions using data from the 1992, 1998 and 2005 waves of the National Family Health Survey in India. While we do not try to eliminate every alternative hypothesis, our model predicts (and the data support) a number of very specific empirical relationships that any alternative hypothesis would also have to explain.

First, breastfeeding increases with birth order. As mothers reach their ideal family size, their demand for the contraceptive effects of breastfeeding grows. Second, so long as some preference for sons exists, boys are breastfed more than girls: after the birth of a daughter, parents are more likely to continue having children (and thus limit breastfeeding) in order to try for a son. Third,

---

<sup>1</sup>Research suggests that nursing suppresses the Gonadotropin-releasing hormone (GnRH) that regulates ovulation. Nursing can also cause weight loss, which can disrupt ovulation (Blackburn, 2007). While most doctors believe that there is no danger in breastfeeding while pregnant, many women in our data cite pregnancy as the reason they stopped nursing. See the next section for further discussion.

<sup>2</sup>See Feachem and Koblinsky (1984), Victoria et al. (1987), and Habicht, DaVanzo, and Butz (1988).

using the same logic, children with older brothers are breastfed more. Fourth, these gender effects are smallest for high and low values of birth order. For low (high) birth-order children, mothers will want to continue (stop) having children regardless of the sex of her children and thus breastfeed boys and girls equally. Finally, the peak gender effect should arise for birth order close to a couple's target family size, as at that point their decision to have another child is highly marginal and thus most dependent on considerations such as sex composition.

The contraceptive effects of breastfeeding coupled with its potential health benefits to children suggest a negative relationship between total fertility (quantity) and child health ("quality"): A mother who wants to have many children will wean her child sooner in order to conceive again quickly, with negative consequences for child health. Conversely, a mother who uses breastfeeding to limit her family size will confer health benefits to her children. Thus, breastfeeding represents a heretofore unexamined mechanism for the quantity-quality tradeoff in fertility (introduced by Becker (1960), Becker and Lewis (1973) and Becker and Tomes (1976), and documented in India by Rosenzweig and Wolpin (1980), among others). Through this mechanism, as fertility rates fall, mothers will, on average, breastfeed their children more.

The combination of the health benefits of breastfeeding and the gender gap in breastfeeding also might help explain child mortality patterns in India. Boys are breastfed more than girls, but the difference is small for first-borns and in the first six months of life, which coincides with excess female mortality patterns in India. We find that excess female mortality between ages one and three years—the age range where there is a large gender gap in breastfeeding—is driven by households without piped water, which is consistent with girls' lower rate of breastfeeding increasing their exposure to water-borne disease and, in turn, their mortality rate. Back-of-the-envelope calculations suggest that breastfeeding accounts for 27 percent of excess female deaths between ages one and three in India (or 26,500 to 44,400 "missing girls" each year).

Our results also have potential policy implications related to modern contraception. Expanding the availability of birth control could either increase or decrease breastfeeding. If mothers rely on breastfeeding when more effective forms of contraception are unavailable, then access to modern birth control might lead them to substitute away from breastfeeding. Obviously, the benefits of modern contraception may well swamp this potential cost, but one policy implication would be that campaigns to introduce modern contraception may need to be coupled with campaigns to encourage breastfeeding if policy makers wish to prevent breastfeeding rates from falling. Additionally, improving water quality may become a more urgent policy priority in communities with access to contraception, given the possible declines in nursing. Conversely, modern contraception could increase breastfeeding if it allows a mother to space births further apart or successfully stop

having children when she wants; in fewer cases will a mother need to suspend the breastfeeding of a child because she has become pregnant again. We find suggestive evidence that different forms of birth control have opposing effects on breastfeeding. Condoms and other reversible birth-control methods act as a substitute for breastfeeding, while access to sterilization seems to increase the average duration of breastfeeding.

The remainder of the paper is organized as follows. Section 2 provides background on how breastfeeding affects fertility and how pregnancy affects breastfeeding. Section 3 presents the model, and Section 4 describes the data. Section 5 tests the model’s predictions that breastfeeding should increase with birth order. Section 6 does the same for the predictions that breastfeeding depends on the gender composition of a mother’s children and that this effect interacts with birth order. Section 7 discusses the health effects related to these breastfeeding patterns. Section 8 examines how the availability of modern contraception affects breastfeeding, and Section 9 concludes.

## 2 Background on breastfeeding and subsequent fertility

### 2.1 Are women less fertile while they breastfeed?

The medical research suggests at least two mechanisms by which breastfeeding inhibits fertility. First, nursing affects certain hormones that regulate ovulation. Breastfeeding appears to interrupt the release of the Gonadotropin-releasing hormone (GnRH), which triggers the pituitary gland to release high levels of luteinizing hormone (LH). This so-called “LH surge” marks the beginning of ovulation. There is also some evidence that breastfeeding increases levels of the hormone prolactin, which inhibits ovulation (Blackburn, 2007). The degree to which ovulation is disrupted depends on how many times a day the mother breastfeeds and the intensity with which the child suckles (Rous, 2001).

Second, nursing diverts calories from the mother to the infant. For mothers who consume a limited number of calories, this diversion can lead to malnutrition, which shuts down ovulation. This channel likely plays a more important role in developing countries. Indeed, using India’s National Family Health Survey (NFHS), we regress mother’s weight on a dummy variable for whether she is currently breastfeeding and a rich vector of covariates and find that nursing is associated with a loss of 1.4 kilograms (three percent of body weight).<sup>3</sup>

An important question for our model is whether mothers actually know about the contraceptive

---

<sup>3</sup>Results available upon request. Covariates include a linear control for mother’s years of education, a linear and quadratic control for her age, a linear and quadratic control for her height, a linear and quadratic control of the child’s year-of-birth, and dummy variables for the survey wave, state of residence, urban/rural, sex of the child, and child’s age in months.

effects of breastfeeding. Typically, the nursing mother does not menstruate (a phenomenon known as “lactational amenorrhea”), which would presumably alert her to her temporary inability to conceive.<sup>4</sup> The NFHS survey we use directly asks non-pregnant women the reason they are not using birth control. Many respondents state they are currently trying for another child. Of the remainder, 34 percent cite breastfeeding as their reason for not using contraception.

## 2.2 Do women stop breastfeeding once they become pregnant again?

A related question is how subsequent conceptions or births affect the mother’s decision to continue breastfeeding the older child. According to the American Academy of Family Physicians (2008), there is no evidence that breastfeeding while pregnant is harmful to the fetus, but some speculate that it could increase the likelihood of miscarriages (Verd, Moll, and Villalonga, 2008). Similarly, the medical profession does not officially discourage breastfeeding two children at once (“tandem breastfeeding”), though there is some evidence that infant weight gain is slowed if a mother is simultaneously nursing an older sibling (Marquis et al., 2002).

Of course, this evidence speaks only to the risks associated with breastfeeding after another conception or birth, but the relevant question for this paper is whether mothers choose to stop breastfeeding after a conception or birth, for whatever reason. The mother’s decision might be driven not by the risks as perceived by the medical profession, but because, say, the opportunity cost of her time and energy rises after a subsequent pregnancy or birth. Indeed, in the NFHS, the most common reason women cite when they stop nursing is “became pregnant,” (over 32 percent of respondents cite this reason).

## 3 Model

This section presents a simple dynamic model in which a mother’s decision to breastfeed and her future fertility depend on each other. The model predicts that breastfeeding rates will have several distinctive features with respect to children’s birth order, gender, and the gender composition of older siblings.

### 3.1 Overview

We assume that a mother has a per-period utility determined by the number of children she currently has and their sex composition. After each birth, she must decide whether to breastfeed the child. If she does not breastfeed, she will have another child in the next period. If she does breastfeed, then she will not have another child in the next period. She makes this decision in

---

<sup>4</sup>Lactational amenorrhea is neither necessary nor sufficient for breastfeeding to prevent pregnancy. Even if a nursing mother resumes menstruation, she remains less fecund, and on rare occasions, an amenorrheic woman can conceive (Singh, Suchindran, and Singh, 1993).

order to maximize the infinite sum of discounted per-period utility.

Later in this section we discuss in greater detail many of the assumptions this simple framework entails, but we highlight one here. We assume the only function of breastfeeding is contraception and thus ignore any health benefits it might provide the child. Doing so allows us to demonstrate that our model can generate gender differentials in breastfeeding even when mothers value the health of daughters and sons equally. Including a health benefit would increase the level of the male breastfeeding advantage our model generates but would not qualitatively change the other, more distinct predictions regarding gender and birth-order interactions.

### 3.2 Setup of the model

A mother's utility depends on the quantity of children,  $n$  and the quantity of sons,  $s$  she has. Her period utility is  $u(n, s)$ , and she has an infinite horizon with a discount rate  $\beta$ . Time periods are denoted by  $t$ . In each period, a woman gives birth to either one child or no children.

A mother who gives birth to a child in period  $t$  decides whether to breastfeed the child,  $b_t \in \{0, 1\}$ . We assume breastfeeding perfectly inhibits fertility and has no ancillary costs or benefits. If  $b_t = 1$ , then  $n_{t+1} = n_t$  and  $s_{t+1} = s_t$ . If  $b_t = 0$ , then  $n_{t+1} = n_t + 1$  and, since the next child is equally likely to be a boy or a girl,  $s_{t+1} = s_t + 1$  or  $s_{t+1} = s_t$ , each with probability  $1/2$ . For the remainder of this section, we suppress the time subscript.

The breastfeeding decision, in essence, acts as a fertility stopping decision. For a mother who currently has  $n$  children of which  $s$  are sons, one option is to not breastfeed (continue having children), in which case she receives  $u(n, s)$  this period and the discounted expected value function in the next period. The value function in the next period is  $V(n + 1, s)$  or  $V(n + 1, s + 1)$ , with equal probability. The other option is to breastfeed (stop having children) and receive  $u(n, s)$  in this period and in the infinite subsequent periods (giving a lifetime discounted utility of  $u(n, s) + \beta \frac{u(n, s)}{1 - \beta} = \frac{u(n, s)}{1 - \beta}$ ). We assume that mothers also have access to another form of contraception and that they breastfeed when indifferent; they use the other form of contraception in subsequent periods.

The decision problem is therefore:

$$V(n, s) = \max\{V^{b=1}, V^{b=0}\} = \max\left\{\frac{u(n, s)}{1 - \beta}, u(n, s) + \beta \left(\frac{V(n + 1, s) + V(n + 1, s + 1)}{2}\right)\right\}.$$

We specify the period utility function as

$$u(n, s) = \underbrace{\phi f(n) - c(n)}_{\equiv q(n)} + \lambda g(s).$$

The term  $q(n) \equiv \phi f(n) - c(n)$  captures the net benefits (benefits,  $\phi f(n)$ , minus costs,  $c(n)$ ) of having  $n$  children, with  $\phi > 0$  parameterizing the demand for children. We assume that  $f'' < 0$  and  $c'' > 0$ , so  $q'' < 0$  for all  $n$ . In other words, with respect to the number of children  $n$ , the *marginal* net benefit of an additional child is strictly decreasing and the *total* net benefit  $q$  displays an inverted- $u$  shape.

The term  $\lambda g(s)$  represents the additional utility from sons, with  $\lambda > 0$  measuring the degree of son preference. We assume  $g' > 0$  and  $g'' < 0$ , with  $g(s) < G \forall s$ ; the value of having more sons is bounded above. Note that for convenience we consider smooth  $f$ ,  $c$ , and  $g$  defined over  $\mathbb{R}_+$ , despite the fact that in a woman's choice set  $n$  and  $s$  only take on integer values.

A useful quantity to define is the value of  $n$  up to which a mother would choose to have another child regardless of her son preference or the sex composition of her children. We call this quantity  $\hat{n}$ .

**Definition.** Let  $\hat{n} = \max\{n \mid q(n+1) - q(n) \geq 0\}$ .

Son preference factors into the breastfeeding decision once  $n > \hat{n}$ . Intuitively, mothers unambiguously gain from having up to  $\hat{n}$  children, after which point they weigh the net cost of having more children (which grows unboundedly) against the value of trying for more sons.<sup>5</sup>

### 3.3 Predictions of the model

The subsection presents the predictions of the model regarding the dependence of breastfeeding on a child's birth order, gender, and the sex composition of his older siblings.

**Proposition 1.** *Breastfeeding is increasing in birth order.*

*Proof.* A mother who stops at  $n$  children will have breastfed her  $n^{\text{th}}$  child and not breastfed her first  $n - 1$  children. This follows from the equation of motion for  $n$ . Once a mother chooses to stop having children, she will not resume having children. Suppose she did. She could increase her lifetime utility by shifting her childbearing earlier, given her positive discount rate. ■

In addition to depending on birth order, breastfeeding also depends on gender.

**Proposition 2.**

- (i) *A boy is more likely to be breastfed than a girl.*
- (ii) *A child is more likely to be breastfed if a larger number of his or her older siblings are male, all else equal.*

---

<sup>5</sup>Lemma 1 in the Appendix shows that there exists a unique  $\hat{n}$  and that it always falls in  $(n_{max} - 1, n_{max})$  where  $n_{max}$  is the  $n$  that maximizes  $q$ . Note that we are interested in the case where a mother wants to have at least one child, or  $\hat{n} > 0$  (equivalently,  $q(1) - q(0) > 0$ ). Also note that a mother with all sons might continue past  $\hat{n}$  because of the marginal benefit of more sons,  $\lambda g'(s)$ , but every mother continues up to at least  $\hat{n}$ .

*Proof.*

- (i) Using Lemma 2 from the Appendix, a mother will breastfeed iff  $u(n, s) \geq \frac{u(n+1,s)+u(n+1,s+1)}{2}$ . Keeping the terms in the utility function that depend on  $s$ , breastfeeding is increasing in  $g(s) - g(s+1)$ , which is increasing in  $s$  since  $g'' < 0$ . Holding the sex of the first  $n - 1$  children fixed,  $s$  is higher when the  $n^{\text{th}}$  child is a boy, so a boy is more likely to be breastfed than a girl.
- (ii) From the proof to part (i), breastfeeding is increasing in  $s$ . Holding the sex of the  $n^{\text{th}}$  child fixed,  $s$  is increasing in the number of boys among the first  $n - 1$  children, so an  $n^{\text{th}}$  child with more brothers among his or her siblings is more likely to be breastfed. ■

In the model, how breastfeeding affects the child who is nursed does not enter into the utility function, so there is no difference in how much the mother values breastfeeding her sons versus her daughters per se. Instead, the breastfeeding gender gap is caused by fertility stopping preferences. Moreover, through this mechanism, not just the gender of the child, but also the gender of his or her older siblings affects breastfeeding.

Perhaps the least obvious prediction is how the gender gap in breastfeeding depends non-monotonically on birth order.

**Proposition 3.** *The gap in breastfeeding of boys versus girls is most pronounced at middle birth order. In other words, the likelihood that a mother will breastfeed a son but not a daughter rises with birth order at low birth order, and then eventually decreases with birth order.*

*Proof.* See appendix. ■

The intuition behind this result is that at low  $n$ , mothers want to continue having children regardless of the sex composition of their existing children since  $\phi f(n) - c(n)$  is still increasing in  $n$ . Therefore they will breastfeed neither sons nor daughters. At high enough  $n$ , the increased cost  $c(n)$  becomes large enough (since  $c' > 0$  and  $c'' > 0$ ) that it outweighs any benefit of having another son. A mother will breastfeed both a son or a daughter in this case. However, at intermediate values of  $n$ , when weighing the costs of higher quantity with the benefit of having (in expectation) more sons, mothers want to stop having children if and only if their  $n^{\text{th}}$  child is male, so they will breastfeed a son but not a daughter.

### 3.4 Predictions regarding “distance from ideal family size”

The model also predicts that breastfeeding patterns change abruptly when  $n$  reaches  $\hat{n}$ . While a mother’s “ideal” quantity of children is ill-defined in the presence of son preference, one can think of  $\hat{n}$  as one measure of the mother’s preferred quantity; it is the quantity at which she wants

to stop having children in the limit that her son preference goes to zero. Empirically, we will use survey questions on ideal family size to test the predictions below, and we refer to  $\hat{n}$  as “ideal family size.”

**Proposition 4.**

- (i) *Breastfeeding is constant for birth order below the ideal family size and increases in birth order only after the ideal family size has been reached.*
- (ii) *There is no gender gap in breastfeeding for birth order below the ideal family size. The gender gap in breastfeeding only arises when the ideal family size has been reached.*

*Proof.*

- (i) A mother will not breastfeed for  $n < \hat{n}$ , by Lemma 1. Proposition 1 establishes the remainder of the proof. The mother will begin breastfeeding at some  $n \geq \hat{n}$ , the exact value depending on the gender composition of her children and the extent of her son preference.
- (ii) The proof to part (i) establishes that neither boys nor girls are breastfed for  $n < \hat{n}$ , so there is no gender gap. Proposition 2(i) establishes the remainder of the proposition. ■

**3.5 Discussion**

We close this section by revisiting some of the assumptions of the model, beginning with two implications arising from our decision to model breastfeeding as a “zero/one” decision. Though the model can easily be relabeled so that the decision is between a short period and long period of nursing to reflect that fact that almost all children in India are breastfed initially, the binary nature of the decision still requires further attention.

First, modeling the decision as binary makes the predictions of the model hold weakly for an individual mother. For example, a mother with negligible son preference who wants three children will breastfeed her third child but not the first two (or, equivalently, {short, short, long}). In line with Proposition 1, her breastfeeding choice is weakly increasing in birth order for all birth order; it is constant between birth order one and two, and then strictly increasing from birth order two to three.

Empirically, we examine a large population of mothers, and given heterogeneity in ideal family size and son preference, aggregating over a population should smooth out the discrete fertility-stopping decisions predicted by the model. In addition, patterns for a particular mother depend on the realized sex composition of her previous children, whereas for a population, the population-average sex composition is known. Thus, we conjecture that the results would hold strictly (e.g.,

breastfeeding is strictly increasing in birth order) at the population-level for most well-behaved joint distributions of  $\phi$  (demand for quantity) and  $\lambda$  (son preference).

Second, the binary specification does not take into account mothers' using breastfeeding to *space* births as opposed to merely prevent them. The essential assumption to incorporate birth-spacing into our framework is that mothers who want to continue having children might space their births to an extent, but they will still breastfeed less than mothers who want to stop having children entirely. This assumption follows naturally from a model (such as ours) where a mother receives flow utility from her children; she will begin receiving the positive flow of utility sooner if she has her children sooner.<sup>6</sup>

The final assumption we discuss is that breastfeeding perfectly prevents conception after birth, even though the studies cited in the previous section indicate it merely decreases fecundity. However, allowing mothers to conceive while breastfeeding only reinforces the model's predictions. Recall from the previous section that mothers who become pregnant while breastfeeding tend to wean the current child, which suggests a causal relationship *from* future fertility *to* breastfeeding in addition to the causal relationship we model from breastfeeding to future fertility. Both mechanisms predict a negative relationship between a mother's tendency to breastfeed and variables related to desired future fertility, and we make the choice to model only one in the interest of simplicity.<sup>7</sup>

Even with these simplifications, the model makes several very distinct empirical predictions. Other theories might be able to generate the prediction that breastfeeding depends on birth order (e.g., learning-by-doing models) or gender (e.g., a biological difference in the difficulty of weaning boys versus girls). However, alternative hypotheses would also have to explain why breastfeeding depends on older siblings' gender *and* why the breastfeeding gender gap has an inverted-*u* shape with respect to birth order. Similarly, they would have to explain why breastfeeding discontinuously increases once mothers reach their "ideal" family size and why the male breastfeeding advantage also increases just at this point. Thus, we feel that evidence supporting these predictions would collectively provide strong support for our claim that women take into account preferences over

---

<sup>6</sup>Stepping outside the model, fecundity declines with age, so delaying childbearing also carries the risk that the mother becomes less able to conceive. A related point is that in the model, a mother would not stop having children and then resume childbearing in a later period; she accrues more utility from her children if she has them sooner.

<sup>7</sup>For example, consider a modification of the model such that a mother chooses whether to breastfeed her child both in the period in which the child is born and also in the next period. Assume a mother is constrained to breastfeed only one child at a time, and she prefers to breastfeed the younger child (which follows directly if she perceives the health benefits for each child as concave in breastfeeding duration). Then a mother will breastfeed her child in the second period of his life if and only if the child does not have a younger sibling, so if and only if the child was breastfed in the first period of his life. Thus, the two channels reinforce each other. If there were unwanted pregnancies or mothers who wanted to conceive were sometimes unsuccessful, which would entail a stochastic element to the model, then one channel could operate without the other.

future fertility when deciding when to wean their children.

## 4 Data

Our empirical analysis uses the 1992, 1998 and 2005 waves of the National Fertility and Health Survey (NFHS) of India, a repeated cross-sectional data set based on the Demographic and Health Survey. The NFHS surveys a representative sample of ever-married women ages 15 to 49 across India.

The main advantage of the data set for our purposes is that it records the number of months all children under three, four or five years (depending on the wave) were nursed. Additionally, as basic demographic information is recorded for every child born to a survey mother, we can calculate variables such as birth order and the sex composition of siblings for each child in the survey. The survey also includes a variety of information on contraception, desired fertility, and child health, as well as standard demographic and household characteristics.

We make several sampling restrictions. First, our observations are at the child level, so we include only women who have given birth to at least one child. Second, we exclude observations with missing values for length of nursing period, which restricts the survey to relatively recent births since the survey does not collect retrospective breastfeeding information for older children; the data were collected for children up to age four for the first wave, up to age three for the second wave, and up to age five for the third wave. (Because the duration of breastfeeding is censored at 36 months in the second wave, we top-code the variable at 36 months for all the survey waves.) In addition, for about one percent of children whose breastfeeding information was actually solicited, the data are missing or were deemed “inconsistent” by the surveyors. Third, we exclude mothers who have eight or more children (the 95<sup>th</sup> percentile for this variable) to reduce composition bias from mothers with unusually large family size. Fourth, we exclude multiple births (e.g., twins) since their birth order is less well-defined and how long they are breastfed might systematically differ by virtue of their not being a singleton. Finally, for the breastfeeding analysis, we exclude children who have died, as otherwise the nursing period would be censored in such a manner that does not reflect mothers’ preferences regarding breastfeeding; this restriction results in a loss of about four percent of remaining observations. Our final sample includes just over 110,000 observations.<sup>8</sup>

Summary statistics for the sample used in the breastfeeding analysis are presented in Table 1. The mean length of breastfeeding is 14.8 months, with a standard deviation of 9.0 months. However, this variable is censored for children still being breastfed; the mean length of breastfeeding

---

<sup>8</sup>For the analysis of child mortality, we of course do not condition on being alive and we are not limited to births for which breastfeeding data are available, so the sample is slightly different, with about 163,000 observations; we describe the construction of that sample in section 7.

adjusted for censoring is 18.5 months. The mean (median) value of birth order is 2.6 (2). We observe women who may not have completed their fertility, so we cannot calculate total fertility in our data. However, based on Census data, total fertility conditional on having at least one child (the population from which our sample is drawn) is roughly 4 for the cohorts in our data.<sup>9</sup>

Although the data set covers a large sample of children and includes detailed breastfeeding and contraceptive information, it also presents several challenges with respect to our hypothesis-testing. For example, the breastfeeding variable is only recorded for relatively young children, so we can only observe breastfeeding duration for at most one or two children per mother.<sup>10</sup> As such, comparing breastfeeding across siblings with a mother-fixed-effect model is infeasible, and our analysis is subject to some of the standard criticisms of cross-sectional estimation, which we attempt to address in the next two sections.

## 5 Empirical analysis: Breastfeeding as a function of birth order

### 5.1 Estimation strategy

Our first set of predictions relates breastfeeding duration to a child’s birth order. We test the hypothesis that breastfeeding is a positive function of the desire to delay or inhibit future fertility and that both of these variables increase with birth order.

Essentially, all of our estimation equations in this section take the following form:

$$Outcome_i = F(BirthOrder_i) + X_i \cdot \gamma + \varepsilon_i \quad (1)$$

where *BirthOrder* is a vector of variables related to birth order,  $F()$  is a function from the appropriate vector space to  $\mathbb{R}$ ,  $X$  is a vector of covariates, and  $\varepsilon$  is an error term. *Outcome* is usually some measure of breastfeeding duration, though in later sections we also examine variables related to future fertility and health.

We begin exploring the effect of birth order on breastfeeding by imposing as little structure as possible, allowing birth order to enter non-parametrically as a vector of dummy variables, and estimate both OLS and hazard models. We estimate the following OLS model, which includes a vector of age-in-months fixed effects up to 36 months (the maximum value of the outcome) to mechanically account for the fact that recently-born children will appear to have shorter breastfeeding spells due to right-censoring:

---

<sup>9</sup>In the 1991 Census, mothers age 50-54 had a mean total fertility of 4.8. We assume fertility is falling at 1% a year. Disaggregated fertility data for the 2001 Census are not yet publicly available.

<sup>10</sup>A further problem that invalidates a mother-fixed-effect model is that there is a mechanical correlation between breastfeeding and inclusion in the sample: For a mother to have two children young enough for their breastfeeding data to be available, she must not have breastfed the older one much.

$$Breastfeed_i = \sum_j \beta_j \cdot \mathbb{1}(BirthOrder_i = j) + X_i \cdot \gamma + a_i + \varepsilon_i. \quad (2)$$

$Breastfeed_i$  is the number of months a mother reports having breastfed child  $i$ , and  $a_i$  is the vector of age-in-month fixed effects.

Although the OLS model has the advantage that the coefficients are easy to interpret, the right-censoring must be mechanically corrected with age-in-months fixed effects. We therefore also estimate the following proportional hazard model, which accounts for the fact that many children in our regression sample are still being nursed, imposes no conditions on the baseline hazard function, and models independent variables as having a proportional effect on the hazard rate:

$$h_i(t) = h_0(t) * exp \left( \sum_j \beta_j \cdot \mathbb{1}(BirthOrder_i = j) + X_i \cdot \gamma + \varepsilon_i \right). \quad (3)$$

We show the results from these and related estimations in the following graphs and tables.

## 5.2 Results on birth order

Figure 1 plots birth order on the horizontal axis and the estimated coefficient on the birth-order dummy variables on the vertical axis. We show the estimated coefficients from both an OLS model and a hazard model. To make the coefficients comparable, the hazard coefficients are for “survival” rather than “failure”; a larger coefficient implies that the variable is associated with a longer duration of breastfeeding. As the figure shows, the pattern of coefficient estimates are similar regardless of which specification is used. The figure also shows the distribution of birth order in the data.

The first three columns of Table 2 report regression results that summarize the effect of birth order on breastfeeding duration. The first column shows the coefficient on birth order from an OLS regression with no other covariates, which suggests that a one-unit increase in birth order is associated with a 0.46 month increase in breastfeeding duration.

Higher birth-order children are born to older mothers, are born more recently, and belong to a larger family. If breastfeeding has been trending over time or if mothers with higher fertility differ in their propensity to breastfeed, for example because they have less attachment to the labor force, then the coefficient on birth order could suffer from omitted variable bias. Thus, the second column includes a linear control for mother’s years of education, a linear and quadratic control for her age, a linear and quadratic control of the child’s year-of-birth, and dummy variables for the survey wave, state of residence, urban/rural, and sex of the child. The covariates directly address mother’s age and time trends as confounding factors, and include proxies for likely total fertility, such as mother’s education. With the controls added, the estimated coefficient falls to 0.21. Col.

(3) is the hazard estimation analogue to col. (2) and suggests that a one-unit increase in birth order is associated with a six-percent decrease in the probability of being weaned in any given month.

We have a number of concerns about the birth-order results. The fact that including a set of covariates shrinks the coefficient on birth-order by roughly fifty percent suggests that birth order is correlated with other variables that predict breastfeeding duration. While the rich set of controls in col. (2) partially addresses this problem, it is impossible to completely eliminate concern about omitted variable bias for the birth-order results. Any variable that has a roughly linear relationship with birth order could serve as a candidate confounding variable, and there are likely to be many. And even with the most comprehensive controls, it is difficult to account for the fact that, mechanically, birth-order is at least partly determined by family size. For reasons previously described, including mother fixed effects is not a viable option using our data.

Moreover, even a causal effect of birth order on breastfeeding would not be definitive evidence of our hypothesis that future fertility determines breastfeeding choices. With respect to the relationship between breastfeeding duration and birth order, our hypothesis is observationally equivalent to a “learning by doing” model in which a mother’s cost of breastfeeding declines as she gains experience with each subsequent birth.

For all of these reasons, while it is reassuring that the basic birth-order results fit the model’s predictions, we consider the results presented in the remainder of this section and in the next section to be stronger tests of the model.

### 5.3 Results using “distance from ideal family size”

Our model makes predictions regarding not only a child’s birth order but his birth order *relative to his mother’s “ideal” family size*. Specifically, the model implies that “breastfeeding is constant for birth order below the ideal family size and increases in birth order only after the ideal family size has been reached,” (*Proposition 4(i)*). Fortunately, our data set includes each mother’s self-reported “ideal” family size.<sup>11</sup> From this measure we generate a variable that measures the current distance from the mother’s ideal family size:  $\Delta Ideal_{ij} = BirthOrder_{ij} - Ideal_j$  (where  $BirthOrder_{ij}$  is the birth order of mother  $j$ ’s  $i$ th child and  $Ideal_j$  is mother  $j$ ’s ideal family size). This variable allows us to directly test this prediction of the model.

Before describing our estimation strategy and results, we highlight some potential problems with the “ideal” family size measure. First, the concept of ideal family size is not well-defined without reference to sex composition. For example, a mother with no children might say her ideal

---

<sup>11</sup>The survey question is, “If you could go back to the time you did not have any children and could choose exactly the number of children to have your whole life, how many would that be?”

family size is two, thinking her first two children would be boys; if she knew she would have two girls first, then her ideal family size might be larger.

Second, to avoid cognitive dissonance, mothers might self-report their “ideal” fertility preference to match their actual fertility outcome, and thus the variable might not reflect their preferences at the time they were having children and making breastfeeding choices (Pritchett, 1994). However, we suspect that such ex-post rationalization is limited. Table 1 indicates that actual fertility systematically exceeds self-reported ideal fertility; ideal family size is on average 2.7, but based on the Census, total fertility is about four for these cohorts. Furthermore, the histogram in Figure 2 shows that many children are born to mothers who have already reached their ideal family size (i.e.,  $\Delta Ideal_{ij} \geq 0$ ).<sup>12</sup>

Figure 2 plots the estimated coefficients on the  $\Delta Ideal$  dummy variables from a regression analogous to equation (2) that replaces *BirthOrder* with  $\Delta Ideal$ . The point estimates display a similar increasing pattern as those for birth order, which is not surprising as  $\Delta Ideal$  is increasing in birth order (though not merely a linear transformation, as  $Ideal_j$  varies for each mother). The figure also suggests there is a jump in the level of breastfeeding once ideal family size is reached ( $\Delta Ideal_{ij} = 0$ ), as predicted by the model.<sup>13</sup>

The last four columns of Table 2 show results from regressions relating breastfeeding duration and  $\Delta Ideal$ . In particular, we specify the  $\Delta Ideal$  effect as a level increase once mothers reach their ideal fertility. Col. (4) of Table 2 shows the coefficient on an indicator variable for  $\Delta Ideal \geq 0$  when no other covariates are included; once mothers reach their ideal fertility they breastfeed subsequent children an extra 1.07 months. Adding covariates in col. (5) yields a coefficient of 0.88 (an 18 percent drop). That adding covariates decreases the  $\Delta Ideal \geq 0$  coefficient much less than the birth-order coefficient suggests that the  $\Delta Ideal$  regressions are less vulnerable to composition bias since  $\Delta Ideal$  implicitly accounts for heterogeneity in fertility preferences.

The more exacting specification, though, is to test for a discrete increase in breastfeeding at  $\Delta Ideal = 0$  while allowing for an overall linear effect of  $\Delta Ideal$ . We estimate the following equation, where again the variable of interest is the indicator for  $\Delta Ideal \geq 0$ :

$$Breastfeed_i = \delta \cdot \mathbb{1}(\Delta Ideal_{ij} \geq 0) + \lambda \cdot \Delta Ideal_{ij} + X_i \cdot \gamma + a_i + \varepsilon_i. \quad (4)$$

Col. (5) shows the results without our set of covariates included. Once mothers reach their ideal fer-

<sup>12</sup>A full 65 percent of mothers in our sample who are 35 years or older have more children than their ideal number; we look at older women because they are more likely to have completed their fertility, but even these women might continue having more children.

<sup>13</sup>To explore more rigorously the discrete jump in breastfeeding when a mother reaches her desired number of children, we create a series of dummy variables corresponding to the conditions  $\Delta Ideal \geq x$  and then run separate regressions of breastfeeding on each dummy variable (as well the standard covariates). Indeed, the regression using the variable  $\Delta Ideal \geq 0$  yields the  $t$ -statistic with the largest magnitude. Results available upon request.

tility, the duration they breastfeed subsequent children increases by 0.40 months. With covariates included, the effect now becomes stronger: Above and beyond an overall linear effect of  $\Delta Ideal$ , breastfeeding duration goes up by 0.58 months once a mother reaches her ideal family size.

## 5.4 Discussion

There appears to be a strong positive correlation between breastfeeding and birth order, consistent with the model’s first prediction. As the average mother in India has about four children, the last child would be breastfed about 0.6 (4.5 percent, given a censoring-adjusted base of 18.5 months) longer than his oldest sibling. Similarly, children born once their mother has reached her “ideal” family size are breastfed 0.6 months longer than older siblings. While this effect size might not seem large, a breastfeeding differential of this magnitude has important consequences for child mortality, as we show in Section 7.

While we have already discussed potential biases in our estimates, especially those regarding birth-order, several pieces of evidence from this section point to a causal relationship between desired fertility and breastfeeding. First, while composition bias may indeed affect the coefficient on birth-order in Table 2, such bias alone cannot explain our finding in Figure 1 that breastfeeding increases between birth order one and two, since almost all mothers in India have at least two children. Moreover, the marked increase in breastfeeding just at the point when mothers reach their ideal family size suggests that they take into account their fertility preferences when deciding when to wean their children.

We now turn to testing our model’s distinctive predictions regarding how gender, sex composition of existing children, birth order and desired fertility interact to predict breastfeeding duration.

## 6 Breastfeeding as a function of gender and birth order

There are many reasons a mother may decide to breastfeed sons longer than daughters. Daughters might simply be harder to nurse, or sons might be harder to wean. Another explanation relates to parents’ tendency to allocate more resources to sons in India (Das Gupta, 1987; Pande, 2003; Mishra, Roy, and Retherford, 2004; Oster, 2009). If mothers perceive breastfeeding as superior to alternatives such as infant formula or solid food, then they will nurse sons longer.

Our model offers a different explanation, namely that a preference for having a *future* son causes a gender gap in breastfeeding the *current* child.<sup>14</sup> We argue that demand for an additional

---

<sup>14</sup>Sex-selective abortion is particularly high among those who want to have fewer children, suggesting an at-least-one-son preference (Arnold, Kishor, and Roy, 2002). Moreover, son preference appears to persist beyond the first son; researchers find that at any birth order, a couple is more likely to stop having children if they have just had a son (Das, 1987; Sen, 1992; Mutharayappa, Choe, Arnold, and Roy, 1997). Even when a family has subsequent

child is higher after the birth of a girl and thus mothers wean daughters sooner in the hopes of conceiving again (*Proposition 2(i)*). However, this prediction alone does not allow us to distinguish our hypothesis from those that argue that mothers value the health of sons more than daughters and thus nurse sons more in the belief that “breast milk is best” or that sons simply “take to the breast” more easily than do daughters. We focus on predictions of the model that allow greater separation of our hypothesis from others in the following subsection.

## 6.1 Testing our hypothesis

First, since parents’ preference for another son would depend on the gender composition of all previous children and not just the last one, there should be a separate effect on breastfeeding from variables such as “already has a son” and “percent of children that are male” (*Proposition 2(ii)*). While these variables are highly correlated with the sex of the current child, they should exhibit a separate effect in our model but not in other models of breastfeeding in which, say, mothers simply value the health of sons more than the health of daughters.

Second, the model suggests that the effect of a child’s gender and other sex-composition variables are strongest at “intermediate” birth order (*Proposition 3*). At very low (high) birth order, mothers want to continue (stop) having children regardless of sex composition. Candidate confounding variables would have to cause a similar, non-monotonic gender differential with respect to birth order.

Third, our model predicts that the male breastfeeding advantage should become most pronounced once mothers have reached their “ideal” family size (*Proposition 4(ii)*). Again, candidate confounding variables would have to conform to this specific pattern.

## 6.2 Results on breastfeeding as a function of gender and sex composition

Figure 3 plots breastfeeding duration (the survival function) separately for boys and girls. The graphs do not hold any other variables constant but do account for censoring of the breastfeeding duration variable due to some mothers still breastfeeding at the time of the survey. Though the heaping of observations at multiples of six months makes it difficult to see the differences across gender at those specific points, one can compare percentiles for which both distributions are far from a heaping point.<sup>15</sup> For example, the 30<sup>th</sup> percentile for boys is about one month (seven percent) more than that of girls. At the 70<sup>th</sup> percentile the difference is about 2.5 months (ten

---

children, birth intervals are longer after the birth of a son (Arnold, Choe, and Roy, 1998; Retherford and Roy, 2003).

<sup>15</sup>Some of the observed age-heaping may reflect true discontinuities in behavior and not merely rounding error. For example, the Qur’an calls for children to be nursed for 24 months (Yurdakok, 1988). Indeed, we find that Muslims mothers are more likely to report having weaned their child at exactly 24 months. When we exclude Muslims mothers from our regression sample, all of our relevant effect sizes grow, which is not surprising since following this religious dictate should make these mothers less responsive to the incentives described in our model.

percent).

Table 3 shows the effect of gender and sex composition of existing children on breastfeeding duration. The first column indicates that sons receive an additional 0.37 months of breastfeeding relative to daughters. This effect barely moves after adding our standard set of covariates from Table 2 plus birth-order fixed effects in the second column. Col. (3), the hazard-model analogue of col. (2), suggests that sons have a ten percent lower probability of being weaned in any given month relative to daughters.

In the next two columns we examine whether the sex composition of siblings has an independent effect on breastfeeding even after accounting for the gender of the current child. Col. (4) shows that a mother already having at least one son increases by 0.28 months the current child’s breastfeeding duration. In other words, the breastfeeding gap between two girls, one of whom has an older brother and the other of whom does not, is almost as large as the breastfeeding gap between boys and girls. Similarly, mothers breastfeed a child longer when the male share of her other children is high. Given the high degree of correlation between these variables and the male dummy variable ( $\rho = 0.73$  for “male” and “percent of children who are male” and  $\rho = 0.60$  for “male” and “at least one son”), these regressions provide strong evidence that the gender composition of *past* births affects the breastfeeding of the *current* child, a pattern not predicted by theories that assume mothers prefer to breastfeed sons or that sons “take to the breast” better than daughters.

Finally, the last column of Table 3 re-estimates the specification in col. (2), allowing the male coefficient to vary by survey wave (i.e., including interactions of male with survey-wave dummies). The main effect is the gender differential in breastfeeding for the most recent wave of data, collected in 2005-6; the coefficient of 0.46 is about 18 percent higher than the average pooled gender differential seen in column 2. The male-wave interactions are imprecisely estimated but suggest that the gender gap in breastfeeding has been increasing over time from 0.31 months in the first wave to 0.39 months in the second wave to 0.46 months most recently. This increase over time is consistent with previous evidence that India’s current fertility decline has intensified sex bias (Das Gupta and Bhat, 1997).<sup>16,17</sup>

### 6.3 Gender effects as a function of birth order

We now examine how the gender differences seen in the previous subsection vary with birth order. Recall that our model predicts that the effect of these variables are small for both high and

---

<sup>16</sup>A decrease in fertility will intensify sex bias if the desired number of children falls more rapidly than the desired number of sons. See, for example, Das Gupta and Bhat (1997). This effect seems to outweigh the fact that as sex selective abortion has become more available, girls should be born more often into families with less son preference.

<sup>17</sup>Note that our other empirical results, those presented in the previous section and in the remainder of the paper, are also similar when estimated separately by survey wave.

low birth order. Moreover, for the population as a whole, the peak effect of these variables should occur between the average for “ideal family size” (around three) and average total completed fertility (around four).<sup>18</sup> Therefore, our model not only predicts that the gender effect take an inverted- $u$  shape with respect to birth-order but also specifies the birth-order interval in which the gender effect peaks.

In order to investigate the effect of birth order on the gender coefficients in a flexible manner, we estimate the following equation:

$$\begin{aligned} Breastfeed_i = & \alpha \cdot Male + \sum_j \beta_j^p \cdot \mathbb{1}(BirthOrder_i = j) \\ & + \sum_j \beta_j^m \cdot Male \times \mathbb{1}(BirthOrder_i = j) + X_i \cdot \gamma + \varepsilon_i. \end{aligned} \quad (5)$$

The key addition is the vector of  $BirthOrder \times Male$  dummy variables, which allows each combination of gender and birth order to have its own fixed effect.

Figure 4 plots the estimated breastfeeding durations by birth order and gender from the above estimation (after accounting for right-censoring but including no other covariates). Sons are breastfed more than girls throughout the birth-order distribution, but the difference is not constant across birth order. The difference is increasing until birth order four, and then decreasing after that, in line with the model’s predictions. The male-female difference by birth order is also plotted to show the inverted- $u$  shape more clearly.

Based on the evidence in Figure 4, we specify the gender effect in the parametric regression analysis as a quadratic function of birth order and display results of such regressions in the first three columns of Table 4. Col. (1) shows the OLS estimate accounting for right-censoring and including dummy variables for male and birth order. The coefficients for the quadratic terms suggest that sons’ breastfeeding advantage peaks when birth order equals roughly 4.1. Cols. (2) to (3) suggest this peak is robust to including additional controls or using a hazard model.

Figure 5 plots breastfeeding duration by gender, but this time against  $\Delta Ideal$ , the distance between the birth order of the current child and the mother’s “ideal family size” (recall that the variable  $\Delta Ideal_{ij} = BirthOrder_{ij} - Ideal_j$ , where  $BirthOrder_{ij}$  is the birth order of the mother  $j$ ’s  $i$ th child and  $Ideal_j$  is mother  $j$ ’s ideal family size). The male-female difference appears to increase just as mothers reach their “ideal” family size and then slowly narrows. Indeed, allowing

---

<sup>18</sup>Recall the two mechanisms underlying the breastfeeding-fertility relationship. If the relationship depended entirely on women using breastfeeding as contraception, then the population-average effect would peak at the average value for “ideal” fertility. If the relationship depends entirely on subsequent pregnancies causing women to stop nursing the current child, then the effect would peak at the average completed fertility. As both channels appear to operate in our data, we predict the effect to peak somewhere between these two values.

a level increase once  $\Delta Ideal_{ij}$  reaches zero is the specification most favored by the data. (Results available upon request.) We read the evidence in Figure 5 as consistent with our prediction that gender preferences will have the largest effect on breastfeeding when the mother’s decision to have another child is most marginal.

The final two columns of Table 4 present the regression analogue to Figure 5. Even after controlling for our standard covariates and allowing for an interaction of *Male* and  $\Delta Ideal$ , there appears to be a discrete jump in the son breastfeeding advantage just when mothers reach their “ideal” family size. The son advantage jumps by 0.49 months when a mother reaches her ideal family size. The effect size is unchanged when covariates are added to the regression.

#### 6.4 Relating our results to fertility outcomes

This paper hypothesizes a negative relationship between breastfeeding and subsequent fertility, both because mothers who want to conceive again wish to limit nursing-induced postpartum amenorrhea and because mothers who conceive again generally stop breastfeeding. We believe that the breastfeeding patterns with respect to gender and birth-order that we have found lend considerable support to our hypothesis. However, we can also directly test whether the same variables that predict breastfeeding also predict fertility-stopping.

Table 5 shows the results when “no younger sibling” serves as the dependent variable and the right-hand-side takes on many of the specifications explored in this section and the previous one. For comparability with the previous results, we use the same sample as used for the breastfeeding analysis (though in principle, we could include older children). Col. (1) shows that high-birth-order children are more likely not to have a younger sibling at the time of the survey and col. (2) shows an increase in this probability once mothers reach their “ideal” family size.

The next two columns focus on gender and sex composition. Not only does being male increase the probability of not having a younger sibling, so does having at least one older brother—exactly the pattern as when breastfeeding served as the dependent variable.

Finally, the interactions between gender and birth order echo the patterns we found for breastfeeding. A son’s increased likelihood of being the youngest child peaks when birth-order equals about 3.5. In addition, the amount by which boys are more likely than girls to be the youngest child is higher for mothers who have reached or surpassed their ideal family size.

#### 6.5 Discussion

In this section we have presented evidence in support of the more demanding predictions of our hypothesis that desired future fertility affects breastfeeding duration. First, not only are boys breastfed for longer periods (a result consistent with mothers valuing boys’ health more than

girls or simply wanting to be more loving toward sons), but the gender of *older* children affects the breastfeeding duration of the *current* child. Second, this gender effect take the shape of an inverted-*u*: it is smallest at low birth order when a mother wants to continue having children (limit breastfeeding) regardless of her children’s gender and at high birth order when she wants to stop having children (prolong breastfeeding) regardless of gender. Third, the peak gender effect arises at a birth order between average ideal family size in our sample and realized family size according to the Indian census; at these values of birth order mothers’ decisions to conceive again are the most marginal. Fourth, the male breastfeeding advantage displays a discrete increase once mothers reach their self-reported ideal family size. Finally, the same specifications that predict breastfeeding duration in our data also predict fertility-stopping.

## 7 Breastfeeding and child health patterns in India

In this section, we examine the implications of our results for infant and child health patterns in India. We first review the medical evidence on the health benefits of breastfeeding, focusing on its effects in early childhood. We then review the patterns in child health and mortality across gender and birth order documented by past researchers to determine whether they appear plausibly related to the patterns we find in breastfeeding. Finally, we use our NFHS sample to directly test whether child survival patterns exhibit the same relationship with gender, birth-order and  $\Delta Ideal$  as does breastfeeding duration.

### 7.1 Breastfeeding and infant and child health in developing countries

Medical and public health researchers have suggested several mechanisms by which breastfeeding promotes health for infants and young children in developing countries. First, human milk has immunological benefits; for example, it contains glycans that are believed to play an anti-infective role in the gastrointestinal tract (Morrow et al., 2005). Second, breastfeeding allows infants to avoid contaminated food and water sources, a mechanism that can play an especially important role in environments with poor sanitation (Habicht, DaVanzo, and Butz, 1988).

Much of the empirical work has focused on breastfeeding’s effect on mortality. Victoria et al. (1987) and Betran et al. (1987) find that breastfeeding is associated with lower rates of infant mortality from diarrheal disease and acute respiratory infection in Latin America, and Chen, Yu, and Li (1988) find similar results in China. Finally, Retherford et al. (1989) find that controlling for breastfeeding largely eliminates the negative correlation between infant mortality and subsequent birth spacing in Nepal.

Of particular interest are the studies that examine how breastfeeding beyond infancy affects mortality. Briend, Wojtyniak, and Rowland (1988) report that in Bangladesh, children who have

been weaned have three times the mortality rate in the age range of 18 to 36 months as those still being breastfed; the authors calculate that one third of the deaths in this age range are attributable to lack of breastfeeding. The World Health Organization (2000) estimates that in developing countries, mortality risk between ages one and two is twice as high if a child is not being breastfed.

Finally, other work focuses on health status and morbidity. Som, Pal, and Bharati (2007) report that in India shorter periods of breastfeeding are associated with a higher risk of stunting. Perera et al. (1999) find that breastfeeding reduces respiratory and diarrheal illness in Sri Lanka. Feachem and Koblinsky (1984) review over 30 studies conducted in 14 countries and find overwhelming evidence that breastfeeding reduces the risk of diarrheal disease.

## **7.2 Documented variation in child health across birth order and gender**

### **Child health patterns with respect to birth order**

There appears to be no strong consensus in the medical literature on the causal relationship between birth order and child health. There are some proposed mechanisms that favor higher birth order children (e.g., negative effects on first-borns from higher levels of intrauterine estrogen levels), but the literature primarily focuses on mechanisms that favor lower birth-order children (Arad et al., 2001). Thus, the evidence appears to undermine our prediction that higher birth-order children, by virtue of nursing longer, would enjoy better health outcomes.

Many of the mechanisms cited by researchers relate to resource allocation. For example, Garg and Morduch (1998) find that in Ghana higher birth-order children experience more stunting and are more likely to be underweight than their older siblings, suggesting parents provide more calories during infancy to children born earlier. Behrman (1988) focuses on India, examining inputs rather than outcomes; he uses food consumption data to show that parents favor higher birth-order children in India. In the US, Price (2008) finds that parents spend more “quality time” with the first-born child. Thus, favoritism toward older children in terms of nutrition and attention may swamp any breastfeeding effects.

### **Child health patterns with respect to gender**

Several of the breastfeeding patterns we have documented coincide with previously established features of gender differentials in mortality in India. A distinctive feature of the missing women problem in India is that the sex ratio (ratio of girls to boys) drops considerably in childhood; in contrast, in China the problem is essentially fully realized at or shortly after birth (Das Gupta, 2005). In India, compared to boys, girls have a forty percent higher mortality rate between the ages of one and five but an equal mortality rate before age one (Acharya, 2004). This pattern

coincides with our finding, seen in Figure 3, that most of the gender gap in breastfeeding does not arise until about twelve months after birth.

Furthermore, researchers have documented that in India excess female mortality is muted for first births (Das Gupta, 1987; Retherford and Roy, 2003). Similarly, we find relatively little gender difference in breastfeeding for first-born children.

### 7.3 Testing for a mortality-breastfeeding relationship in our sample

#### Empirical strategy

While the results from the existing literature are consistent with breastfeeding contributing to observed child mortality patterns, we can also directly test whether the breastfeeding patterns we found in the previous sections correspond to mortality differentials in the NFHS. The survey records mortality data for all children ever born to a mother, and we test whether the same gender, birth-order and ideal-family-size interactions that predicted breastfeeding duration in the previous sections also predict infant survival.

We use results from the previous sections as well as past research to construct tests of the mortality-breastfeeding hypothesis. The gender differences in the cumulative-distribution function of breastfeeding duration presented in Figure 3 indicate the age range where breastfeeding differences (and thus any related mortality differences) should be largest.<sup>19</sup> The gender difference in breastfeeding is concentrated between age 12 and 36 months, and there is no apparent gender difference during the first few months of the child’s life. Hence, we test for breastfeeding effects on the probability a child dies between twelve and thirty-six months after birth and use immediate post-natal mortality between zero and three months as a placebo test.

Furthermore, as the medical literature stresses breastfeeding’s benefits in the presence of unsanitary water, we test whether the mortality patterns are most pronounced in households without piped water. Not only does this comparison shed light on the mechanism behind any health effects, it also allows us to separate the effects of breastfeeding from potentially confounding variables. For example, one important confounding factor is family size. The fertility-stopping patterns we observe in Table 5 imply that on average girls have more younger siblings, and having a larger family size could cause higher mortality rates for girls (Yamaguchi, 1989; Clark, 2000; Jensen, 2003; Rosenblum, 2008). We can make progress in separating the effects of family size from breastfeeding by comparing households with and without piped water, since family size should affect children in both types of households, but breastfeeding should primarily benefit children without piped water.

---

<sup>19</sup>There may also be long-term effects on the child’s health which lead to mortality at later ages; in essence we test only for contemporaneous effects.

## Mortality data from the NFHS

As we relate mortality to the gender, birth-order and  $\Delta Ideal$  interactions that predict breastfeeding and not to breastfeeding itself, we do not need breastfeeding variables in this estimation and are thus no longer restricted to children under the age of five. However, we might still want to exclude children born, say, fifteen years before the survey date as their mortality information may suffer from recall bias.<sup>20</sup>

Although we do not claim to eliminate recall bias, we try to hold it constant in all the regressions we run. Specifically, when we examine mortality rates between zero and three months, we include children such that  $survey\ date - birth\ date > 3$  months and  $survey\ date - birth\ date < 63$  months. In this case, all observations would have been fully “at risk” for the three-month mortality window and the recall period is five years ( $63 - 3 = 60$  months). Similarly, when we examine mortality between 12 and 36 months, we examine observations satisfying  $survey\ date - birth\ date > 36$  months and  $survey\ date - birth\ date > 96$  months (as well as conditioning on being alive at twelve months; otherwise the variable is not defined). This measure holds recall bias roughly constant across the two measures of mortality, and it also forces us to have similar sample sizes for both the 12-to-36-month mortality rate and the 0-to-3-month mortality rate that we use as a placebo test.<sup>21</sup>

The sample for the 12-to-36-month-mortality regressions has about 163,000 observations and a 12-to-36-month mortality rate of 0.0203. Twenty-two percent of the sample has piped water in their dwelling or on their plot, which we use as a proxy for access to clean water. The sample for the 0-to-3-month-mortality regressions is roughly the same size, with a 0-to-3-month mortality rate of 0.0455.

## Results

Consistent with past research, we find birth-order effects that tend to go “against” our breastfeeding hypothesis. In the interest of space we do not report these regressions, though note here that although there is no birth-order effect for mortality before three months, mortality between ages 12 and 36 months increases by 0.41 percentage points with every one-unit increase in birth order for households without piped water and by 0.13 percentage points for households with piped water. We tentatively conclude that the resource advantage of first-born and low-birth-order chil-

---

<sup>20</sup>Byass et al. (2007) contend that recall bias in the DHS leads researchers to underestimate infant and under-age-five mortality in Ethiopia by 14 to 27 percent.

<sup>21</sup>The results are robust to using the same sample for the two analyses, i.e., using the 12-to-36-month-mortality sample for the 0-to-3-month-mortality analysis. Obviously we cannot use the 0-to-3-month-mortality sample for both regressions as those who die in the first three months of life do not have well-defined values for 12-to-36-month mortality. The results are also robust to using mortality between 6 and 30 months as the dependent variable in order to avoid the age-heaping at multiples of 12 months.

dren documented in past work outweighs the health benefits high-birth-order children receive via breast milk.

The remainder of this section focuses on the effects of gender and its interactions with birth-order and  $\Delta Ideal$  on child mortality.

Table 6 shows the results when mortality between the ages of 12 and 36 months serves as the dependent variable, with the first four columns focusing on households without piped water. Col. (1) indicates that sons are 0.85 percentage points less likely to die between 12 and 36 months, consistent with a breastfeeding survival advantage since boys are also breastfed more than girls. The point estimates in the second column suggest that the male survival advantage, like the breastfeeding advantage estimated in the previous section, has an inverted- $u$  shape with respect to birth-order (although the coefficient on the squared term is not quite statistically significant, it has a  $p$ -value of 0.12). Col. (3) shows that once mothers reach their ideal number of children, the male advantage with respect to child mortality grows significantly. Col. (4) controls for the interaction of *Male* and  $\Delta Ideal$ , and the male advantage still appears to increase discontinuously once a mother has reached her ideal family size.

To help distinguish breastfeeding's effect from the effects of confounding variables such as family size that display similar patterns with respect to gender and its interaction with birth-order and  $\Delta Ideal$ , we estimate the exact same regression as in cols. (1) to (4) but this time include only households with piped water. For these households, the mortality effects of breastfeeding should not be as strong, so the mortality patterns with respect to gender, birth order, and desired fertility that we examine should be muted. As seen in cols. (5) to (8), the coefficients of interest are all smaller in magnitude or in fact wrong-signed. For example, in households with piped water there seems to be a son *disadvantage* once ideal family size is reached.

We formally test for the equality of the coefficients by estimating four fully-interacted models (i.e., every variable of interest and control variable is interacted with *Piped*) equivalent to jointly estimating columns 1 and 5, columns 2 and 6, columns 3 and 7, and columns 4 and 8. For the first specification, the male coefficients for the piped and unpiped households are statistically different from each other with a  $p$ -value less than 0.001. For the second specification, the relevant coefficients are the linear and quadratic birth-order terms interacted with the male dummy, and they differ between the two types of households with a  $p$ -value of 0.05. For the last two specifications, which test for a discrete increase in the survival advantage of boys once the ideal family size is reached, the coefficients for the male dummy interacted with  $\Delta Ideal \geq 0$  differ with  $p$ -values less than 0.001 and 0.05, respectively. Thus, households where breastfeeding would play an especially important role in protecting child health appear to drive the mortality results.

Our second robustness check is to use infant mortality immediately after birth as the outcome, as there is no gender differential in breastfeeding during this period. Finding the same patterns as we do in Table 6 would suggest that something correlated to breastfeeding duration, and not breastfeeding duration itself, is driving those results. Appendix Table 1 is the exact analogue to Table 6 except mortality in the first three months after birth serves as the dependent variable. The male coefficient in col. (1) has the wrong sign and moreover is essentially identical to that in col. (5), suggesting the result does not depend on water purity. As in Table 6, male survival advantage emerges in households without piped water once mothers reach their “ideal” family size; but, unlike in Table 6, the effect is essentially identical between households with and without piped water. For every 0-to-3-month-mortality specification, one cannot reject that the coefficients of interest are the same in households with and without piped water, with  $p$ -values ranging from 0.72 to 0.97.

In summary, in settings for which breastfeeding plays an especially important role in child health, we can directly map the relationship between breastfeeding and gender, birth-order and  $\Delta Ideal$  interactions onto mortality. In settings in which breastfeeding should play little role, the same patterns between mortality and these variables do not emerge.

#### 7.4 Breastfeeding and “missing girls”

Using our results from Sections 5 and 6 and estimates of the health benefits of breastfeeding from the medical and public health literature, we present back-of-the-envelope calculations of how much breastfeeding contributes to the gender gap in child mortality in India.

Based on existing literature, we assume that not breastfeeding a child between the ages of 12 and 36 months increases the probability she dies during this interval by 150 percent.<sup>22</sup> In our sample, between the ages of 12 and 36 months girls have a 0.75 percentage point higher mortality rate (2.42 versus 1.67 percent) and are breastfed eight percent less than boys.<sup>23</sup> We thus calculate that by virtue of being nursed less, girls’ mortality rates are 12 percent higher ( $8.0 \times 1.50$ ) than boys, or 0.2 percentage points higher, given the baseline male mortality rate of 1.67 ( $0.12 \times 1.67 = 0.201$ ). Therefore, breastfeeding appears to account for 27 percent ( $0.20/0.75$ ) of excess female mortality

---

<sup>22</sup>Breastfeeding between the ages of 18 and 36 months was found to reduce mortality risk by two-thirds in Bangladesh (Briend, Wojtyniak, and Rowland, 1988). An alternative estimate is that breastfeeding between ages one and two reduces mortality risk in developing countries by one half (World Health Organization, 2000). We use the midpoint of these estimates in our calculation, namely that the mortality risk for ages 12 to 36 months is 2.5 times as high in the absence of breastfeeding (or, equivalently, that not breastfeeding during this age interval increases mortality rates by 150 percent).

<sup>23</sup>We obtain the eight-percent differential by running a regression of breastfeeding duration in months identical to that in Table 3, col. (2), but condition on the child being at least one year old. The coefficient on male indicates boys are breastfed an additional 0.53 months, or roughly eight percent given that the average child is breastfed 6.7 months between the ages of one and three years ( $0.53/6.7 = 0.0791$ ). Regression results and relevant summary statistics available upon request.

rates between the ages of 12 and 36 months. Given that about 14 million girls are born each year in India, of whom 13.2 million survive to age one, earlier weaning accounts for about 26,500 ( $13,200,000 \times 0.002$ ) excess female deaths per year.

We can also use our results from the previous subsection to produce a second estimate of the excess female mortality that results from girls being breastfed less than boys. In households with unpiped water, girls' mortality is 0.85 percentage points greater than that of boys (Table 6, column 1). Using households with piped water to gauge the gender gap in mortality from other channels besides breastfeeding (0.38 percentage points, from Table 6, column 4) implies that the gender gap in breastfeeding leads to 0.47 ( $0.85 - 0.38$ ) percentage points of excess female mortality. Equivalently, in households without piped water, breastfeeding accounts for 55 percent ( $0.47 / 0.85$ ) of excess female mortality between the ages of 12 and 36 months. Noting that about 75 percent of Indian households lack piped water, the breastfeeding gender gap appears to account for roughly 44,400 ( $0.75 \times 0.0047 \times 13,200,000$ ) "missing girls" each year.<sup>24</sup>

Our differences-in-differences estimate is of course very rough. It might overstate the effect of breastfeeding since not all of the differential gender gap in mortality in households with unclean water may be due to breastfeeding. It also might understate the effect since it makes the extreme assumption that breastfeeding confers no health benefits in households with piped water. Nevertheless, whether we use our estimated treatment effects or those from the existing medical literature, our results suggest that breastfeeding may be an important factor behind the "missing girls" problem. Like other types of parental behavior that lead to excess female mortality, son preference is the underlying cause. However, in contrast to most other explanations, excess female mortality due to differential breastfeeding is an unintended consequence of parents' desire to have more sons rather than an explicit decision to allocate fewer resources to daughters.

## 8 How access to modern contraception affects breastfeeding

This section examines the effect of the availability of modern birth control on breastfeeding. The relationship is theoretically ambiguous. As we documented in Section 2, two phenomena underlie the negative relationship between breastfeeding duration and subsequent fertility: breastfeeding

---

<sup>24</sup>We can derive the implied breastfeeding-mortality relationship from Table 6 and compare it to those in the medical literature cited in footnote 22. In households without piped water, between the ages of one and three boys have a 1.93 percent mortality rate and are breastfed seven percent more than girls. (Again, regression results and relevant summary statistics available upon request.) Therefore, girls' seven percent breastfeeding disadvantage between the ages of 12 and 36 months appears to increase their mortality by 24.3 percent ( $0.47 / 1.93$ ), implying that not breastfeeding during this age interval increases mortality by 347 percent. While this estimated effect is larger than that in the past literature, it is estimated specifically for households with unclean water, where we expect the effect to be largest. If we scale our estimate to account for the fact that only three-quarters of households have unpiped water, our estimated effect falls to 260 percent, close to the 200 percent effect found in Bangladesh by Briend, Wojtyniak, and Rowland (1988).

the current child helps prevent or delay a subsequent pregnancy, and a subsequent (perhaps unwanted) pregnancy often causes mothers to wean the current child. The first channel suggests that by providing an alternative form of contraception, modern birth control would substitute for breastfeeding; the second suggests that by more reliably preventing or delaying pregnancies, modern birth control might prolong the period a mother can nurse her current child.

One approach to empirically examining the effect of modern birth control on breastfeeding is to identify exogenous supply-side variation in contraception. However, we know of no strategy to isolate plausibly exogenous variation in access to contraceptives in India during our study period.<sup>25</sup> The imperfect measure of “access” that we construct—which of course is a function of not only supply but also demand—is the average usage of contraception in the respondent’s primary sampling unit (PSU), excluding herself. A PSU is typically a village, and different PSUs are sampled in each survey wave, so this reference group is specific to the survey wave in which the respondent appears. We calculate the proportion of women in the PSU who use terminal birth-control (i.e., sterilization) and the proportion who use reversible birth-control (condoms, IUDs, and less commonly in India, the pill).<sup>26</sup> We cluster the regressions by PSU since the birth control supply measure is highly correlated within a PSU. (The only within-PSU variation is because we exclude the respondent herself when calculating usage in her PSU.)

The mean (standard deviation) of total modern contraception usage is 0.36 (0.22). Notably, sterilization (mean of 0.25 and standard deviation of 0.19) is more common than reversible methods (mean of 0.12 and standard deviation of 0.15).

Col. (1) of Table 7 shows that the local “supply” of total modern contraception seems to decrease breastfeeding. The coefficient is marginally significant. Different birth-control methods might have different effects, so col. (2) allows the supply of sterilization and reversible methods of birth control to enter separately. We find that access to condoms, IUDs, and the pill decreases breastfeeding, while access to sterilization seems to increase it. An interpretation of this result is that sterilization allows mothers to breastfeed their last child longer, as they are uninterrupted by a subsequent pregnancy. Reversible methods, more suitable for birth spacing, have the opposite effect. In a setting without contraception, mothers who want to space births have to do so by breastfeeding longer; when contraception is available, they substitute away from breastfeeding.

If the results in col. (2) indeed reflect a causal effect of different types of contraception on breastfeeding, then there is a further testable implication. Since sterilization permanently elimi-

---

<sup>25</sup>For example, there is very limited temporal or spatial variation in sterilization or IUD incentive payments over our study period. See Bharadwaj (2009).

<sup>26</sup>The survey question used to construct these measures asks a woman which method if any she is using, and the answers are mutually exclusive. The mean number of respondents used to construct this reference group is 17, with a standard deviation of 8.

nates fertility, it is obviously not used for birth spacing and thus its availability should have no effect for early birth-order children. Col. (3) examines the interaction of each type of contraception with low birth order (birth order less than three). As predicted, only at high birth order, when mothers begin to avail themselves of sterilization, does the sterilization supply in the area increase breastfeeding. For reversible birth-control methods, there is no equivalent interaction prediction, and as seen in col. (3), the negative effect of reversible methods on breastfeeding does not vary significantly across birth order.

Obviously, these results are subject to the criticism that the “supply” of modern contraception is not randomly assigned to sampling units. However, the fact that the sterilization supply measure has no effect at lower birth order and only has an effect for birth orders where there would be demand for sterilization among mothers does suggest that the results might not be spurious. We suggest areas for further work on this topic in the conclusion.

## 9 Conclusion

This paper began by arguing that whether a mother breastfeeds a child for a long duration should negatively correlate with whether she gives birth to a subsequent child. There are at least two mechanisms underlying this relationship. First, for physiological reasons, breastfeeding lowers a woman’s fertility. Second, women typically wean a child if they become pregnant again.

We then develop a dynamic programming model of fertility decisions that takes into account this negative covariance between breastfeeding duration and subsequent conception. The model makes a number of very specific predictions regarding how long children will be breastfed. First, breastfeeding increases with birth order. As mothers reach their ideal family size, their demand for contraception grows. They either breastfeed longer to suppress fertility or use other forms of birth control, which allows them to breastfeed longer as they will not be interrupted by another pregnancy. For the same reasons, breastfeeding increases discretely once women reach their “ideal” family size.

Second, if parents have a preference for sons, then boys are breastfed more than daughters: after the birth of a girl, parents are more likely to continue having children (and thus limit breastfeeding) in order to try for a boy. Third, using the same logic, children with older brothers are breastfed more. Fourth, these gender effects are smallest for high and low values of birth order. For low (high) birth-order children, mothers will want to continue (stop) having children regardless of the sex of her children and thus breastfeed boys and girls equally. Finally, the peak gender effect for the population should occur at a birth order somewhere between the average ideal family size (which is 2.7 in our data) and the average realized family size (about 4 for our sample). For birth

order values in this range, a mother’s joint decision about breastfeeding and further childbearing is highly marginal and thus most dependent on considerations such as sex composition.

Using data from the National Family Health Survey in India, we find strong support for each of these predictions. On average, the youngest child in India nurses five percent longer than his oldest sibling, and most of that difference comes from a discrete increase once mothers reach their “ideal” family size. Sons nurse two percent longer than daughters, and there are separate and statistically significant benefits to having at least one older brother and having a larger male share of older siblings. The son advantage is small for low and high birth-order and peaks around a birth order value of 4; it also displays a discrete jump when mother’s reach their ideal family size.

Given the well-documented health benefits of breastfeeding for infants in developing countries, each of the predictions should hold if we simply swap “child survival” with “breastfeeding duration.” We find no support for child survival increasing in birth order either in our sample or based on patterns documented by other researchers, though we note that the strong favoritism toward first-born children found in existing work likely swamps the benefits of breastfeeding.

The mortality patterns regarding gender and its interactions with birth-order and desired family size do support the model’s predictions. As we find that gender differentials in breastfeeding do not appear until after the first year of life, we focus on mortality between twelve and thirty-six months. Boys, especially those of intermediate birth order and those born to mothers who have already reached their “ideal” family size, have lower mortality rates between the ages of twelve and thirty-six months. These results are driven by children living in households that do not have piped water, for whom weaning means possible exposure to contaminated water. As an additional check, we find that these patterns do not hold for mortality within the first three months, when no gender gap in breastfeeding exists.

We calculate that the gender breastfeeding gap accounts for over a quarter of excess female deaths between age one and three years in India, or, equivalently, fifteen percent of excess female deaths under age five. The fact that mothers breastfeed daughters less than sons leads to 26,500 “missing girls” each year. Unlike many other proposed factors causing missing girls, our hypothesis does not require that parents value girls’ health less than boys’. Instead, excess female mortality arises because subsequent fertility decisions, by being intertwined with breastfeeding duration, have unintended health consequences.

Finally, we analyze the interaction between breastfeeding and artificial birth-control. Although we hesitate to draw definite conclusions as we use variation in birth control access that is potentially related to unobserved determinants of breastfeeding, we find some evidence suggesting that sterilization promotes breastfeeding while reversible methods such as condoms and IUDs discour-

age it. Our results are consistent with sterilization extending the nursing period of the last child, and reversible technologies being more conducive to birth-spacing and acting as a substitute for breastfeeding.

Future work might more fully investigate this relationship between breastfeeding and access to birth control, particularly reversible methods which are the focus of most policy initiatives. If contraception indeed crowds out breastfeeding, then policy makers might want to consider pairing contraceptive campaigns with promotion of breastfeeding. In addition, expanded access to birth control and improvements in water quality might be complementary policies if birth control causes breastfeeding rates to fall. We hope researchers investigating birth-control access, whether through randomized controlled trials, ethnographic investigations, or other research designs, will consider breastfeeding as an outcome of interest.

## References

- ACHARYA, A. K. (2004): "Gender Preference and India's Missing Girls: Evidence from Some Selected States of India," mimeo, UNAM.
- AMERICAN ACADEMY OF FAMILY PHYSICIANS (2008): "Breastfeeding, Family Physicians Supporting," Position paper.
- ARAD, I., M. BARAS, R. GOFIN, B. BAR-OZ, AND O. PELEG (2001): "Does Parity Affect the Neonatal Outcome of Very-Low-Birth-Weight Infants?," *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 94(2), 283 – 289.
- ARNOLD, F., M. K. CHOE, AND T. ROY (1998): "Son Preference, the Family-Building Process and Child Mortality in India," *Population Studies*, 52, 301–315.
- ARNOLD, F., S. KISHOR, AND T. ROY (2002): "Sex-Selective Abortions in India," *Population and Development Review*, 28(4), 759–785.
- BECKER, G. S. (1960): "An Economic Analysis of Fertility," in *Demographic and Economic Change in Developed Countries*, ed. by G. S. Becker. Princeton University Press, Princeton, N.J.
- BECKER, G. S., AND H. G. LEWIS (1973): "On the Interaction Between the Quantity and Quality of Children," *Journal of Political Economy*, 81(2, Part 2), S279–S288.
- BECKER, G. S., AND N. TOMES (1976): "Child Endowments and the Quantity and Quality of Children," *Journal of Political Economy*, 84(2, Part 2), S143–S162.
- BEHRMAN, J. R. (1988): "Nutrition, Health, Birth Order and Seasonality: Intrahousehold Allocation Among Children in Rural India," *Journal of Development Economics*, 28, 43–62.
- BETRAN, A. P., M. DE ONÍS, J. A. LAUER, AND J. VILLAR (2001): "Ecological Study of Effect of Breast feeding on Infant Mortality in Latin America," *British Medical Journal*, 323(303), 1–5.
- BHARADWAJ, P. (2009): "Fertility and Rural Labor Market Inefficiencies: Evidence from India," Mimeo, Yale University.
- BLACKBURN, S. T. (2007): "Postpartum Period and Lactation Physiology," in *Maternal, Fetal, and Neonatal Physiology: A Clinical Perspective*. Elsevier Health Sciences.
- 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.
- BYASS, P., A. WORKU, A. EMMELIN, AND Y. BERHANE (2007): "DSS and DHS: Longitudinal and Cross-sectional Viewpoints on Child and Adolescent Mortality in Ethiopia," *Population Health Metrics*, 5(12).
- CHEN, Y., S. YU, AND W.-X. LI (1988): "Artificial Feeding and Hospitalization in the First 18 Months of Life," *Pediatrics*, 81(1), 58–62.
- CLARK, S. (2000): "Son Preference and Sex Composition of Children: Evidence from India," *Demography*, 37(1), 95–108.
- DAS, N. (1987): "Sex Preference and Fertility Behavior: A Study of Recent Indian Data," *Demography*, 24(4), 517–530.

- DAS GUPTA, M. (1987): "Selective Discrimination Against Female Children in Rural Punjab, India," *Population and Development Review*, 13(1), 77–100.
- (2005): "Explaining Asia's 'Missing Women': A New Look at the Data," *Population and Development Review*, 31(3), 529–535.
- DAS GUPTA, M., AND P. M. BHAT (1997): "Fertility Decline and Increased Manifestation of Sex Bias in India," *Population Studies*, 51, 307–315.
- FEACHEM, R. G., AND M. A. KOBLINSKY (1984): "Interventions for the Control of Diarrheal Diseases Among Young Children: Promotion of Breastfeeding," *Bulletin of the World Health Organization*, 62, 271–291.
- GARG, A., AND J. MORDUCH (1998): "Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana," *Journal of Population Economics*, 11(4), 471–493.
- HABICHT, J.-P., J. DAVANZO, AND W. P. BUTZ (1988): "Mother's Milk and Sewage: Their Interactive Effects of Infant Mortality," *Pediatrics*, 81(3), 456–461.
- JENSEN, R. (2003): "Equal Treatment, Unequal Outcomes? Generating Sex Inequality Through Fertility Behavior," Mimeo, Harvard University.
- MARQUIS, G. S., M. E. PENNY, J. M. DIAZ, AND R. M. MARIN (2002): "Postpartum Consequences of an Overlap of Breastfeeding and Pregnancy: Reduced Breast Milk Intake and Growth During Early Infancy," *Pediatrics*, 109(4), e56.
- MISHRA, V., T. K. ROY, AND R. D. RETHERFORD (2004): "Sex Differentials in Childhood Feeding, Health Care and Nutritional Status," *Population and Development Review*, 30(2), 269–295.
- MORROW, A. L., G. M. RUIZ-PALACIOS, X. JIANG, AND D. S. NEWBURG (2005): "Human-Milk Glycans That Inhibit Pathogen Binding Protect Breastfeeding Infants against Infectious Diarrhea," *Journal of Nutrition*, 135, 1304–1307.
- MUTHARAYAPPA, R., M. CHOE, F. ARNOLD, AND T. ROY (1997): "Son Preference and Its Effect on Fertility in India," National Family Health Survey Subject Reports No. 3.
- 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. (2003): "Selective Gender Differences in Childhood Nutrition and Immunization in Rural India: The Role of Siblings," *Demography*, 40(3), 395–418.
- PERERA, B. J. C., S. GANESAN, J. JAYARASA, AND S. RANAWEERA (1999): "The Impact of Breastfeeding Practices on Respiratory and Diarrhoeal Disease in Infancy: A Study from Sri Lanka," *Journal of Tropical Pediatrics*, 45(2), 11518.
- PRICE, J. (2008): "Parent-Child Quality Time: Does Birth Order Matter?," *Journal of Human Resources*, 43(1), 240–265.
- PRITCHETT, L. H. (1994): "Desired Fertility and the Impact of Population Policies," *Population and Development Review*, 20(1), 1–55.
- RETFERFORD, R. D., M. K. CHOE, S. THAPA, AND B. B. GUBHAJU (1989): "To What Extent Does Breastfeeding Explain Birth-Interval Effects on Early Childhood Mortality?," *Demography*, 26(3), 439–450.

- RETFERD, R. D., AND T. K. ROY (2003): "Factors Affecting Sex-Selective Abortion in India," NFHS Bulletin No. 17.
- ROSENBLUM, D. (2008): "Fertility and Child Mortality: Unintended Consequences of Family Formation in India," Mimeo, Yale University.
- ROSENZWEIG, M. R., AND K. I. WOLPIN (1980): "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, 48(1), 227–240.
- ROUS, J. J. (2001): "Is Breast Feeding A Substitute for Contraception in Family Planning?," *Demography*, 38(4), 497–512.
- SEN, A. (1992): "Missing Women," *British Medical Journal*, 304, 586–87.
- SINGH, K. K., C. M. SUCHINDRAN, AND K. SINGH (1993): "Effects of Breastfeeding after Resumption of Menstruation on Waiting Time to Next Conception," *Human Biology*, 65(1), 71–86.
- SOM, S., M. PAL, AND P. BHARATI (2007): "Role of Individual and Household Level Factors on Stunting: A Comparative Study in Three Indian States," *Annals of Human Biology*, 34(6), 632–646.
- VERD, S., J. MOLL, AND B. VILLALONGA (2008): "Miscarriage rate in pregnancy-breastfeeding overlap," *Contraception*, 78(4), 348 – 349.
- VICTORIA, C. G., J. P. VAUGHAN, C. LOMBARDI, S. M. C. FUCHS, L. P. GIGANTE, P. G. SMITH, L. C. NOBRE, A. M. TEXIERA, L. B. MOREIRA, AND F. C. BARROS (1987): "Evidence for a Strong Protective Effect of Breastfeeding against Infant Death due to Infectious Diseases in Brazil," *Lancet*, 330(8554), 319–322.
- WORLD HEALTH ORGANIZATION (2000): "Effect of Breastfeeding on Infant and Child Mortality Due to Infectious Diseases in Less Developed Countries: A Pooled Analysis," *Lancet*, 355(9202), 451–455.
- YAMAGUCHI, K. (1989): "A Formal Theory for Male-Preferring Stopping Rules of Childbearing: Sex Differences in Birth Order and in the Number of Siblings," *Demography*, 26(3), 451–465.
- YURDAKOK, M. (1988): "Breast-Feeding in Islam," *Acta Paediatrica*, 77(6), 907–908.

## Appendix

**Definition.** Let  $\hat{n} = \max\{n \mid q(n+1) - q(n) \geq 0\}$ .

**Lemma 1.** *There exists a unique value of  $\hat{n} > 0$ , and for all  $n < \hat{n}$ , mothers will always have the  $(n+1)^{\text{st}}$  child, regardless of sex preference  $\lambda$  and regardless of the sex composition of existing children.*

*Proof.* Defining  $h(n) \equiv q(n+1) - q(n)$ , our assumption that  $q(1) > q(0)$  is equivalent to  $h(0) > 0$ . Strict concavity of  $q$  implies that  $h$  is a continuous, strictly decreasing function. Since  $q$  has a maximum at  $n_{\max}$  and is strictly decreasing to the right of  $n_{\max}$ , we know that  $h(n) < 0$  for all  $n \geq n_{\max}$ . Thus  $h$  crosses zero exactly once, and does so from above somewhere in  $(0, n_{\max})$ . We have that  $\hat{n} = \max\{n \geq 0 \mid h(n) \geq 0\}$  is equal to the unique  $n$  such that  $h(n) = 0$ .

For any  $n < \hat{n}$ , the benefit to having another child is positive regardless of the sex composition of existing children ( $q$  is positive in this region and  $\lambda g$  is always increasing), so mothers always choose to have the  $(n+1)^{\text{st}}$  child. ■

**Lemma 2.** *A mother will choose to breastfeed iff  $u(n, s) \geq \frac{u(n+1, s) + u(n+1, s+1)}{2}$  (assuming she breaks indifference in favor breastfeeding rather than childbearing).*

*Proof.* The lemma in words states that a mother will not breastfeed her  $n^{\text{st}}$  child and instead will have her  $(n+1)^{\text{st}}$  child iff the expected value of having *exactly* one more child is greater than the value of stopping at  $(n, s)$ .

To show necessity, assume that  $u(n, s) < \frac{u(n+1, s) + u(n+1, s+1)}{2}$ . The expression for the value function is

$$V(n+1, s) = \max \left\{ \frac{u(n+1, s)}{1-\beta}, u(n+1, s) + \beta \left( \frac{V(n+2, s) + V(n+2, s+1)}{2} \right) \right\}.$$

It follows that  $V(n+1, s) \geq \frac{u(n+1, s)}{1-\beta}$ , and, similarly,  $V(n+1, s+1) \geq \frac{u(n+1, s+1)}{1-\beta}$ . Therefore  $\frac{u(n, s)}{1-\beta} < \frac{V(n+1, s) + V(n+1, s+1)}{2}$ , which implies that  $V^{b=0} > V^{b=1}$  or that the mother will choose not to breastfeed.

To show sufficiency, suppose toward contradiction that  $u(n, s) \geq \frac{u(n+1, s) + u(n+1, s+1)}{2}$  but  $V^{b=0} > V^{b=1}$ . This gives  $\frac{V(n+1, s) + V(n+1, s+1)}{2} > \frac{u(n, s)}{1-\beta} > \frac{u(n+1, s) + u(n+1, s+1)}{2(1-\beta)}$ . The mother will have her  $(n+1)$ th child under these assumptions. Since utility is strictly increasing in the number of sons,  $V(n+1, s+1) > V(n+1, s)$  and  $u(n+1, s+1) > u(n+1, s)$ . It follows that either  $V(n+1, s+1) > \frac{u(n+1, s+1)}{1-\beta}$ , or  $V(n+1, s) > \frac{u(n+1, s)}{1-\beta}$ ; either the mother continues having children at  $(n+1, s)$  or at  $(n+1, s+1)$  (or both). Since the marginal value of continuing to have children is higher the lower  $s$  is, conditional on  $n$ , the mother continues having children (at least) at  $(n+1, s)$ .

That the mother continues at  $(n+1, s)$  implies that  $\frac{V(n+2, s) + V(n+2, s+1)}{2} > \frac{u(n+1, s)}{1-\beta}$ . However, if we can show that  $\frac{u(n+1, s)}{1-\beta} > \frac{u(n+2, s) + u(n+2, s+1)}{2(1-\beta)}$ , then we have back our original problem, with  $n$  replaced by  $n+1$ , implying that the mother will continue to have children if she reaches  $(n+2, s)$ . By induction, as  $n$  increases, the mother will keep having children if she remains at  $s$  sons, contradicting Lemma 3, which states that the number of children is bounded above. Therefore, if  $u(n, s) \geq \frac{u(n+1, s) + u(n+1, s+1)}{2}$ , then we have completed the proof of sufficiency.

The remaining piece, therefore, is to show that

$$\frac{u(n, s)}{1 - \beta} > \frac{u(n + 1, s) + u(n + 1, s + 1)}{2(1 - \beta)} \Rightarrow \frac{u(n + 1, s)}{1 - \beta} > \frac{u(n + 2, s) + u(n + 2, s + 1)}{2(1 - \beta)} \quad (6)$$

Rearranging terms, (6) holds if  $u(n + 1, s) - u(n, s) \geq \frac{1}{2}(u(n + 2, s) - u(n + 1, s) + u(n + 2, s + 1) - u(n + 1, s + 1))$ . The terms that depend on  $s$  drop out, so (6) holds if  $q(n + 1) - q(n) \geq q(n + 2) - q(n + 1)$ , which is true because  $q'' < 0$ . ■

**Lemma 3.** *The total number of children that a mother gives birth to is bounded above; that is,  $\exists \bar{n}$  such that she never has more than  $\bar{n}$  children, regardless of sex composition.*

*Proof.* Let  $\bar{n} = \min\{n \mid q(n + 1) - q(n) \leq -\lambda(G - g(0))\} = \min\{n \mid h(n) \leq -\lambda(G - g(0))\}$ . The proof to Lemma 1 established that  $h$  is a strictly monotonically decreasing function with  $h \rightarrow -\infty$  as  $n \rightarrow \infty$ , so  $\bar{n}$  is equal to the unique  $n$  such that  $h(n) = -\lambda(G - g(0))$ . The cost of having the  $(\bar{n} + 1)^{st}$  child is sufficiently large that even if doing so conferred the benefit  $\lambda(G - g(0))$ , which is the maximal utility from having an additional son (recall that  $s \geq 0$ ,  $g$  is strictly increasing and  $g(s) < G$  for all  $s$ ), it would not exceed the utility loss associated with an extra child,  $q(\bar{n}) - q(\bar{n} + 1)$ . Therefore, it is *never* optimal to have more than  $\bar{n}$  children, regardless of the sex composition of existing children. ■

**Proposition 3.** *The largest gap in breastfeeding of boys versus girls is at middle birth order. In other words, the gap rises with birth order at low birth order, and then eventually decreases with birth order.*

*Proof.* The first step of the proof is to show that, at low birth order, mothers never breastfeed because they always want to continue having children. This step is established by Lemma 1. The second step of the proof is to show that, at high birth order, mothers always breastfeed because they always want to stop having children. This step is established by Lemma 3. The final step of the proof, which remains to be shown, is that (1) there exists some  $(n, s)$  such that mothers have the  $(n + 1)^{st}$  child but then stop iff it is male; (2) there does not exist  $(n, s)$  such that mothers have the  $(n + 1)^{st}$  child but then stop only iff it is female.

First, we define a few terms. Define the interval  $\tilde{N} = (\max\{0, \hat{n} - 1\}, \hat{n})$ . Note that mothers with  $n \in \tilde{N}$  children will always have the  $(n + 1)^{st}$  child but not necessarily the  $(n + 2)^{nd}$ .

Now, define  $\hat{s}(n, \lambda) \equiv \max\{s \mid \frac{1}{2}(u(n + 2, s) + u(n + 2, s + 1)) > u(n + 1, s)\}$ . In other words,  $\hat{s}(n, \lambda)$  is the largest  $s$  such that at  $(n + 1, s)$  it is optimal for a mother with son preference  $\lambda$  to have the  $(n + 2)^{th}$  child.  $\hat{s}(n, \lambda)$  is well-defined for all  $n \in \tilde{N}$  and  $\lambda > 0$ . To see this, note that

$$\begin{aligned} \hat{s}(n, \lambda) &= \max\{s \mid \frac{1}{2}(u(n + 2, s) + u(n + 2, s + 1)) > u(n + 1, s)\} \\ &= \max\{s \mid \phi f(n + 2) + \frac{\lambda}{2}[g(s) + g(s + 1)] > \phi f(n + 1) + \lambda g(s)\} \\ &= \max\{s \mid g(s + 1) - g(s) > \frac{2\phi}{\lambda}[f(n + 1) - f(n + 2)]\}. \end{aligned}$$

Now,  $g(s + 1) - g(s)$  is a strictly positive, decreasing function of  $s$  that tends toward zero, so  $\hat{s}(n, \lambda)$  is equal to that unique  $s$  that solves  $g(s + 1) - g(s) = \frac{2}{\lambda}[f(n + 1) - f(n + 2)]$ .

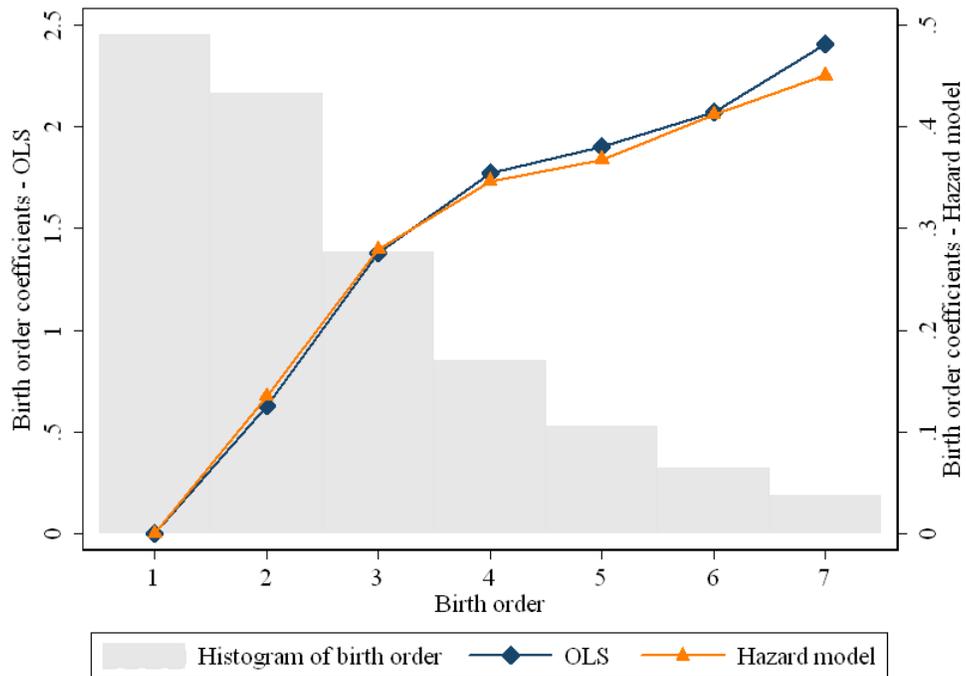
Now, consider a mother at  $(n, \hat{s}(n, \lambda))$ ,  $n \in \tilde{N}$ . She will have her  $(n+1)^{st}$  by definition of  $\tilde{N}$ . If her  $(n+1)^{st}$  child is female, she will now have  $(n+1, \hat{s}(n, \lambda))$  children and by definition of  $\hat{s}$  it is optimal to have another child. But if her  $(n+1)^{st}$  child is male, she will now have  $(n+1, \hat{s}(n, \lambda)+1)$  children. By definition  $\hat{s}(n, \lambda)$  is the *maximal* number of sons for which it is optimal to have the  $(n+2)^{nd}$  child, so she will not have another child. Thus, a mother at  $(n, \hat{s}(n, \lambda))$  will have her  $(n+1)^{st}$  child but iff it is female will she have the  $(n+2)^{th}$ .

Although we have shown that for any  $\lambda \geq 0$  and  $n \in \tilde{N}$  there exists some  $(n, \hat{s}(n, \lambda))$  such that mothers have the  $(n+1)^{st}$  child but continue iff it is female, we have yet to show that  $\hat{s}$  is in the “relevant” region of  $(0, n)$ .

For mothers with  $\lambda$  close to zero,  $\hat{s}$  goes to zero (having an arbitrarily small “amount” of sons is enough to satisfy the mothers son preference, after which she views sons and daughters equally). As son preferences goes to infinity, so does  $\hat{s}$ . As  $\hat{s} > n$  in such cases, mothers will have another child regardless of sex composition as the benefit of future sons always outweighs the cost of more kids regardless of the current number of sons. But for mothers with “moderate” son preference there exist some combinations of  $(n, s)$  such that she stops iff her last child was male.

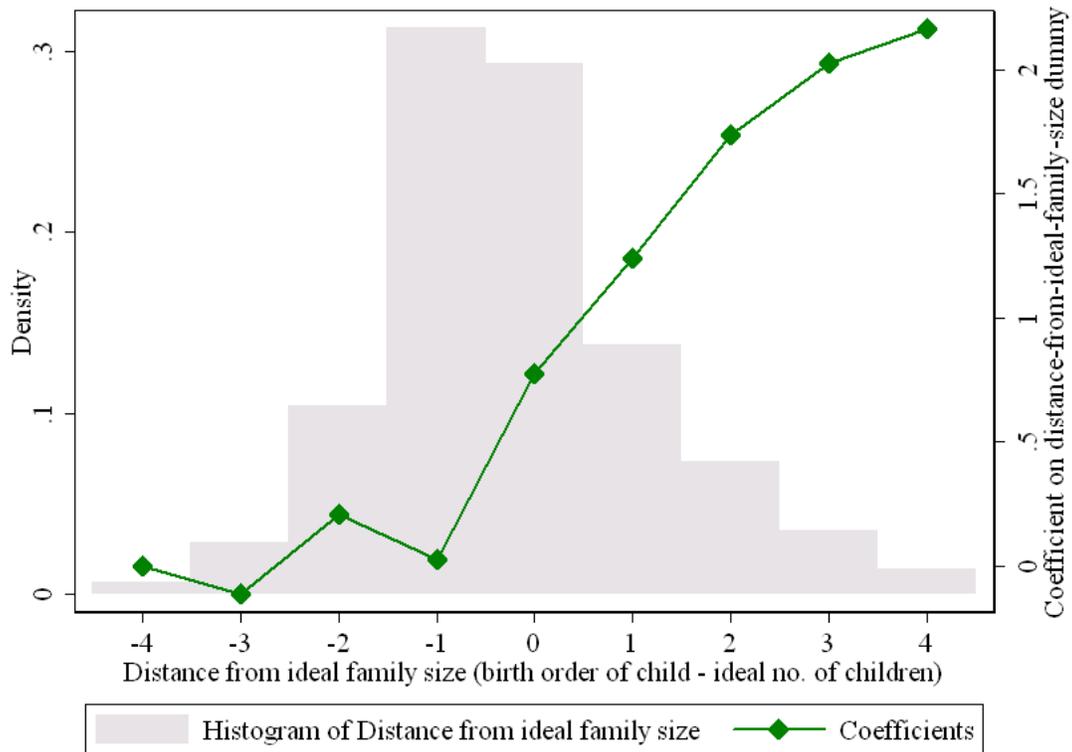
Finally, while we have shown that for “intermediate”  $n$  there exist conditions where mothers stop having children iff their last child was male, the final piece is to establish that no such region exists where mothers stop iff their last child was female. This follows from Proposition 2(i). ■

Figure 1: Breastfeeding duration, by birth order



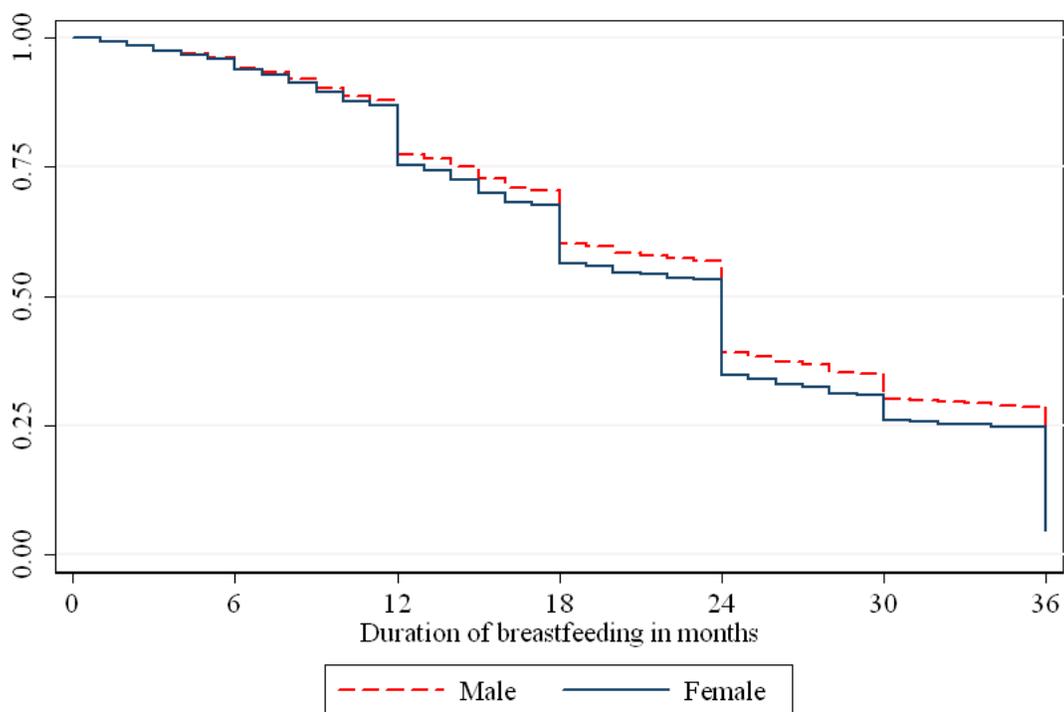
The figure plots the coefficient for birth-order dummies for a regression with breastfeeding duration in months as the dependent variable. The OLS model includes age-in-month fixed effects; no other control variables are included. The omitted category is birth order 1, for which the coefficient is normalized to zero. For the hazard rate, the coefficients are negated for comparability with the OLS coefficients; the hazard coefficients, thus, represent “survival” in breastfeeding rather than exit from breastfeeding. The histogram of birth order for the sample is displayed in the background.

Figure 2: Breastfeeding duration, by distance from ideal family size (birth order - ideal family size)



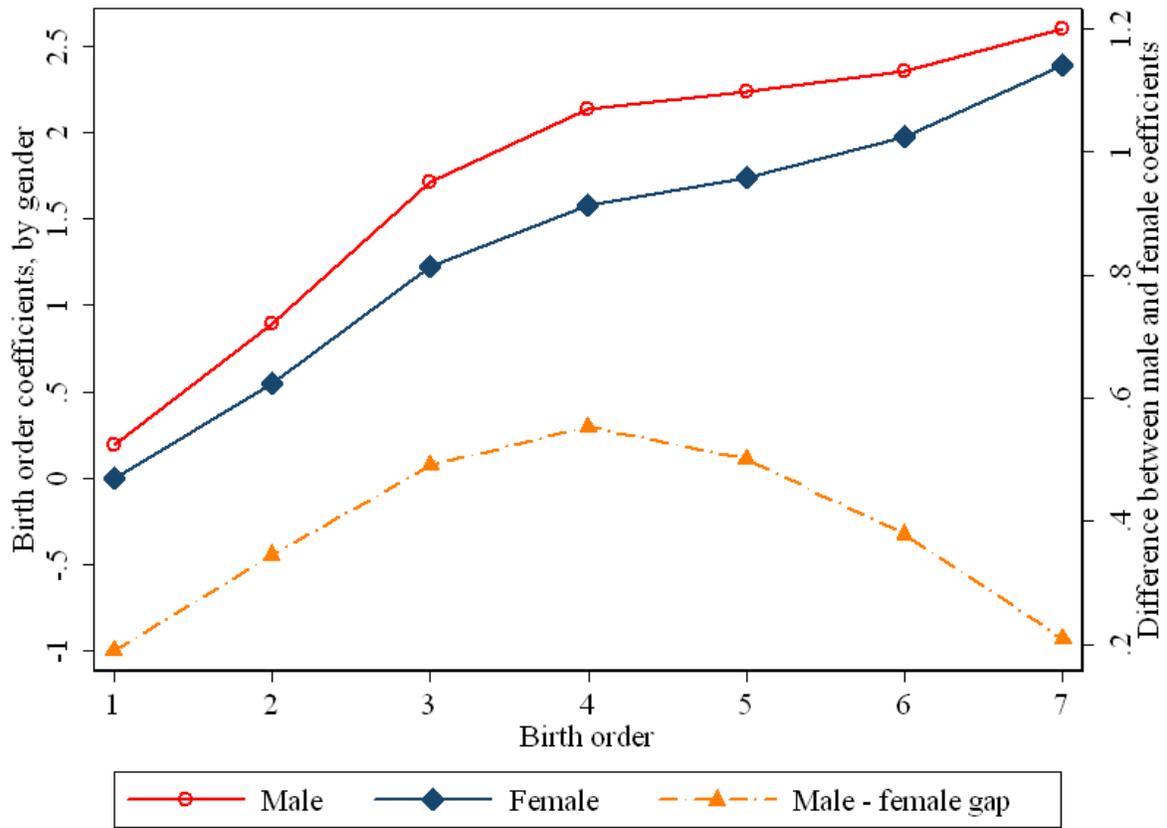
The figure plots the coefficient for “distance from ideal family size” dummies for a regression with breastfeeding duration in months as the dependent variable. Distance from ideal family size is defined as the child’s birth order minus the mother’s ideal family size. The omitted category is distance from ideal family size = -4, for which the coefficient is normalized to zero. The regression includes age-in-month fixed effects and no other control variables.

Figure 3: Survival function for breastfeeding, by gender



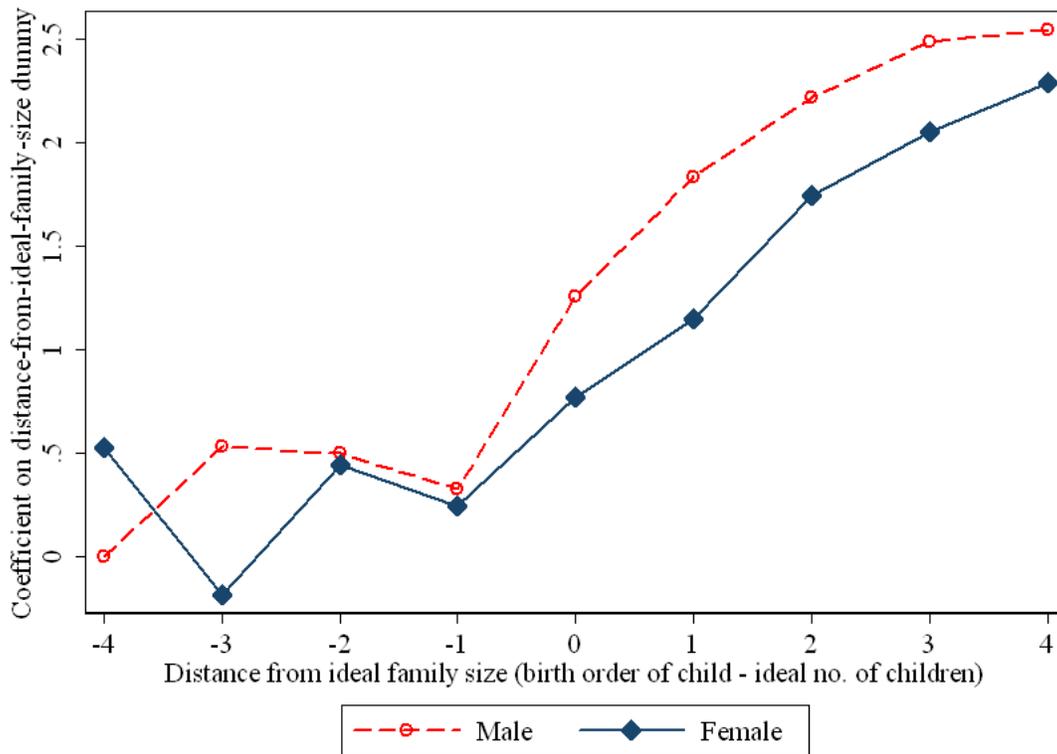
The figure plots the proportion of children, by gender, who are still being breastfed at the duration (age) given on the horizontal axis.

Figure 4: Gender differences in breastfeeding duration, by birth order



The solid lines plot the gender-specific coefficients for birth-order dummies for a regression with breastfeeding duration in months as the dependent variable, with the coefficient for birth order 1 for females normalized to 0. The regression includes age-in-month fixed effects and no other control variables. The dashed line is the difference between the male and female coefficients.

Figure 5: Gender difference in breastfeeding duration, by “distance from ideal family size” (birth order minus - ideal family size)



The figure plots the gender-specific coefficients for “distance from ideal family size” dummies for a regression with breastfeeding duration in months as the dependent variable. Distance from ideal family size is defined as the child’s birth order minus the mother’s ideal family size. The coefficient for males for distance from ideal family size = -4 is normalized to zero. The regression includes age-in-month fixed effects and no other control variables.

Table 1: Summary statistics (means and standard deviations)

	<i>All</i>	<i>Birth order 1 or 2</i>	<i>Birth order &gt;2</i>
Breastfeeding duration (months)	14.78 (8.993)	14.24 (8.739)	15.54 (9.287)
Birth order	2.565 (1.567)	1.469 (0.499)	4.109 (1.220)
Ideal number of children	2.715 (1.090)	2.405 (0.867)	3.170 (1.217)
$\Delta$ Ideal (birth order – ideal no. of children)	-0.190 (1.412)	-0.915 (0.894)	0.882 (1.354)
Child has no younger siblings	0.795 (0.403)	0.769 (0.422)	0.833 (0.373)
Male	0.516 (0.500)	0.513 (0.500)	0.522 (0.500)
Male share of mother's children	0.497 (0.373)	0.510 (0.436)	0.478 (0.258)
Mother has at least one son	0.749 (0.434)	0.631 (0.483)	0.915 (0.279)
Child's age in years	1.937 (1.258)	1.950 (1.262)	1.920 (1.252)
Child's year of birth	1997.2 (5.662)	1997.5 (5.684)	1996.8 (5.602)
Mother's age	25.76 (5.097)	23.72 (4.228)	28.64 (4.816)
Rural	0.681 (0.466)	0.637 (0.481)	0.743 (0.437)
Mother's years of education	4.282 (4.879)	5.597 (5.144)	2.429 (3.767)
Observations	110,183	64,439	45,744

Notes: Data drawn from 1992, 1998 and 2005 waves of National Family Health Survey in India. We include children for whom breastfeeding information is recorded (i.e., all children under the age of three, four or five, depending on the wave), who were alive at the time of the survey, who have values of parity less than eight (the 95<sup>th</sup> percentile), and who are singletons (i.e., not twins or triplets, etc.). The 1992, 1998 and 2005 waves account for 36.9%, 24.6%, and 38.5% of the observations, respectively. "Male share of mother's children" and "Mother has at least one son" includes the child him or herself.

Table 2: Effect of children's birth-order on the length of time they are breastfed

	<i>OLS estimation of months breastfed</i>		<i>Cox est. of hazard rate</i>	<i>OLS estimation of months breastfed</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Birth order	0.4640*** [0.0124]	0.2103*** [0.0179]	-0.0612*** [0.0042]				
$\Delta Ideal \geq 0$				1.0723*** [0.0399]	0.8757*** [0.0454]	0.4007*** [0.0624]	0.5763*** [0.0616]
$\Delta Ideal$						0.3060*** [0.0221]	0.1623*** [0.0236]
Male		0.3908*** [0.0373]	-0.1047*** [0.0087]		0.3750*** [0.0385]		0.3755*** [0.0385]
Rural		-0.1208*** [0.0050]	0.0289*** [0.0011]		-0.1358*** [0.0048]		-0.1320*** [0.0048]
Mother's years of education		0.8064*** [0.0478]	-0.1810*** [0.0102]		0.8336*** [0.0489]		0.8346*** [0.0490]
Observations	110183	110183	108616	104456	104456	104456	104456
R-squared	0.503	0.527	n/a	0.496	0.524	0.497	0.524

Notes: See notes to previous table. *Ideal* is the response of the child's mother to the question "What is your ideal number of children?"  $\Delta Ideal$  equals *Birth order* – *Ideal*. The unit of observation is the child and we cluster standard errors at the mother level to account for mothers who have more than one child in the sample. The specifications in columns 2, 3, 5, and 7 include linear and quadratic controls for mother's age and child's year of birth and dummy variables corresponding to the year of the survey wave and the child's state of residence. The breastfeeding duration variable ranges from 0 to 36 so we include child-age-in-months dummy variables up to 36 months in all OLS regressions to account for the fact that some children are still being breastfed at the time of the survey. The hazard estimation automatically accounts for such right-censoring. Note that the hazard regressions estimates the probability of being *weaned* at time  $t$  conditional on still being breastfed at time  $t-1$  and thus coefficient estimates should have the opposite sign of those of the OLS regressions. The reason the number of observations is not constant across specification is that (1) hazard estimations drop observations that immediately exit (i.e., duration of breastfeeding = 0) and (2) we exclude observations where  $\Delta Ideal$  is not in the interval  $[-4, 4]$  in cols. 4 to 7.

Table 3: The effect of children's gender and their siblings' sex composition on breastfeeding duration

	<i>OLS estimation of months breastfed</i>		<i>Cox est. of haz. rate</i>	<i>OLS estimation of months breastfed</i>		
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.3681*** [0.0384]	0.3887*** [0.0373]	-0.1034*** [0.0087]	0.2444*** [0.0486]	0.2623*** [0.0546]	0.4581*** [0.0675]
Mother has at least one son				0.2796*** [0.0623]		
Male share of mother's children					0.2306*** [0.0751]	
Male x Wave 2						-0.1444 [0.0895]
Male x Wave 1						-0.0654 [0.0929]
Covariates included?	No	Yes	Yes	Yes	Yes	Yes
Observations	110183	110183	108616	110183	110183	110183
R-squared	0.497	0.527	n/a	0.527	0.527	0.527

Notes: See notes to previous tables. "Additional covariates" include linear and quadratic controls for mother's age and child's year of birth and dummy variables corresponding to the year of the survey wave, the child's state of residence, and household's urban vs. rural status, as well as birth-order dummies. We include child-age-in-months dummy variables up to 36 months in all OLS regressions to account for the fact that some children are still being breastfed at the time of the survey. In calculating "male share of children" and "mother has at least one son," we include the child associated with the observation. Standard errors are adjusted for clustering by mother.

Table 4: The gender difference in breastfeeding as a function of birth order

	<i>OLS estimation of months breastfed</i>		<i>Cox estimation of hazard rate</i>	<i>OLS estimation of months breastfed</i>	
	(1)	(2)	(3)	(4)	(5)
Male	-0.0839 [0.1347]	-0.0661 [0.1312]	0.0203 [0.0320]	0.0633 [0.0858]	0.0993 [0.0835]
Male x Birth order	0.2990*** [0.0944]	0.3114*** [0.0923]	-0.0847*** [0.0236]		
Male x Birth order squared	-0.0365*** [0.0135]	-0.0381*** [0.0132]	0.0100*** [0.0035]		
Male x ( $\Delta Ideal \geq 0$ )				0.4922*** [0.1262]	0.4951*** [0.1226]
Male x $\Delta Ideal$				-0.0330 [0.0440]	-0.0172 [0.0429]
Covariates included?	No	Yes	Yes	No	Yes
Max. effect of “Male” when birth order equals...	4.09	4.09	4.25	n/a	n/a
Observations	110183	110183	108616	104456	104456
R-squared	0.504	0.527	n/a	0.498	0.524

Notes: See notes to previous tables. The maximal effect is calculated by setting the derivative of the predicting equation with respect to birth order to zero and solving for birth order.  $\Delta Ideal$  equals  $Birth\ order - Ideal$  where  $Ideal$  is the response of the child’s mother to the question “What is your ideal number of children?” Columns 1 to 3 include birth order fixed effects and columns 4 and 5 include fixed effects for each value of  $\Delta Ideal$ . The covariates are linear and quadratic controls for mother’s age and child’s year of birth and dummy variables corresponding to the year of the survey wave and the child’s state of residence. We include child-age-in-months dummy variables up to 36 months in all OLS regressions to account for the fact that some children are still being breastfed at the time of the survey. Standard errors are adjusted for clustering by mother.

Table 5: Fertility stopping rules as a function of birth order, gender, and ideal fertility

	<i>Dependent variable = Child has no younger siblings</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
Birth order	0.0175*** [0.0009]					
$\Delta Ideal \geq 0$		0.1159*** [0.0032]				
$\Delta Ideal$		0.0043*** [0.0012]				
Male			0.0300*** [0.0020]	0.0154*** [0.0025]	-0.0012 [0.0071]	-0.004 [0.0044]
Mother has at least one son				0.0283*** [0.0033]		
Male * Birth order					0.0241*** [0.0049]	
Male * (Birth order) <sup>2</sup>					-0.0034*** [0.0007]	
Male * ( $\Delta Ideal \geq 0$ )						0.0538*** [0.0063]
Male * $\Delta Ideal$						-0.0074*** [0.0022]
Fixed effects?	None	None	Birth order	Birth order	Birth order	$\Delta Ideal$
Observations	110183	104456	110183	110183	110183	104456
R-squared	0.331	0.344	0.338	0.338	0.338	0.346

Notes: See notes to previous tables. Dependent variable is an indicator variable for whether the child is currently the youngest of all his siblings (i.e., mother has not given birth since his birth). All regressions include linear and quadratic controls for mother's age and child's year of birth and dummy variables corresponding to the year of the survey wave and the child's state of residence.  $\Delta Ideal$  is defined as  $Birth\ order - Ideal$  where  $Ideal$  is the response of the child's mother to the question "What is your ideal number of children?" All regressions are estimated by OLS (so coefficients represent percentage-point changes in the probability of being the youngest child), and standard errors are adjusted for clustering by mother. The regression sample is the same as that used when breastfeeding is the dependent variable.

Table 6: Child mortality as a function of birth order, gender, and ideal fertility

	<i>Dependent variable = Child died between ages 12 and 36 months</i>							
	<u>Household lacks piped water</u>				<u>Household has piped water</u>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Male	-0.0085*** [0.0009]	0.0039 [0.0029]		-0.0064*** [0.0019]	-0.0038*** [0.0010]	0.0006 [0.0037]		-0.0063*** [0.0021]
Male * Birth order		-0.0063*** [0.0022]				-0.0030 [0.0031]		
Male * (Birth order) <sup>2</sup>		0.0005 [0.0003]				0.0003 [0.0005]		
Male * ( $\Delta Ideal \geq 0$ )			-0.0083*** [0.0017]	-0.0039 [0.0028]			0.0007 [0.0021]	0.0042 [0.0030]
Male * $\Delta Ideal$				-0.0018* [0.0010]				-.0017 [0.0014]
Difference in coeffs, unpipd minus piped HHs	-0.0047	-0.0034, 0.0002	-0.0090	-0.0081				
F-test between unpiped and piped HHs	p = .0004	p = .0496	p = .0008	p = .0470				
Observations	127639	127639	118630	118630	35703	35703	34366	34366
R-squared	0.010	0.010	0.009	0.009	0.007	0.007	0.008	0.008

Notes: See notes to previous tables. All regressions include linear and quadratic controls for mother’s age and child’s year of birth and dummy variables corresponding to the year of the survey wave and the child’s state of residence. In columns 1, 2, 5, and 6, birth order dummies are included. In columns 3, 4, 7, and 8, dummies for each value of  $\Delta Ideal$  are included. The sample includes children born between 36 and 96 months before the survey date. The lower bound accounts for the fact that children younger than 36 months do not have a well-defined value for the dependent variable. The upper bound excludes children born far before the survey date in order to limit recall bias (see discussion in Section 7). All regressions are estimated by OLS (so coefficients represent percentage-point changes in the probability of death), and standard errors are adjusted for clustering by mother. The F-test results reported test for the equality of the coefficients of interest between piped and unpiped households. They are based on fully interacted models equivalent to jointly estimating columns 1 and 5 (where the p-value is for the male coefficient); columns 2 and 6 (where the p-value is for the joint test of the linear and quadratic terms); columns 3 and 7; and columns 4 and 8 (where in the last two cases the p-value is for the male interaction with  $\Delta Ideal \geq 0$ ). If instead of estimating a fully interacted model, we constrain the control variables such as state dummies to be the same for piped and unpiped households, the four analogous p-values are 0.0004, 0.0380, 0.0003 and 0.0357.

Table 7: Relationship between breastfeeding and contraception availability

	<i>Dependent variable = breastfeeding duration in months</i>		
	(1)	(2)	(3)
Prevalence of modern birth control	-0.2492* [0.1422]		
Prevalence of sterilization		0.1450 [0.1593]	1.0623*** [0.2071]
Prevalence of other modern birth control (besides sterilization)		-0.9792*** [0.2073]	-0.8245*** [0.2899]
Prevalence of sterilization * (Birth order < 3)			-1.4890*** [0.2141]
Prevalence of other modern birth control * (Birth order < 3)			-0.3173 [0.2935]
Observations	110163	110163	110163
R-squared	0.53	0.53	0.53

Notes: See notes to previous tables. All of the regressions include linear and quadratic controls for mother's age and child's year of birth and dummy variables corresponding to the year of the survey wave and the child's state of residence, as well as birth-order dummies and a dummy for male. We include child-age-in-months dummy variables up to 36 months to account for the fact that some children are still being breastfed at the time of the survey. The "local prevalence" measure for each type of birth control is equal to the share of women who report using that type of contraception in the respondent's primary sampling unit (PSU) for the survey wave in which she appears, excluding herself. The mean number of respondents used to construct this reference group for the respondent is 17, with a standard deviation of 8. The survey question used to construct these measures asks a woman which method if any she is using, and the answers are mutually exclusive, so the sum of the sterilization and other methods is equal to the total modern birth control usage in the area. The mean of total modern contraception usage is 0.364 (sterilization accounts for 0.246 and other methods for 0.118). All regressions are estimated by OLS, and standard errors adjusted for clustering by PSU.

Appendix Table 1: Neonatal mortality as a function of birth order, gender, and ideal fertility

	<i>Dependent variable = Child died between ages 0 and 3 months</i>							
	<u>Household lacks piped water</u>				<u>Household has piped water</u>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Male	0.0057*** [0.0012]	0.0191*** [0.0045]	0.0131*** [0.0019]	0.0193*** [0.0027]	0.0051*** [0.0019]	0.0132** [0.0067]	0.0133*** [0.0030]	0.0075* [0.0044]
Male * Birth order		-0.0073** [0.0030]				-0.0036 [0.0051]		
Male * (Birth order) <sup>2</sup>		0.0006 [0.0004]				0.0000 [0.0008]		
Male * ( $\Delta Ideal \geq 0$ )			-0.0137*** [0.0025]	-0.0076* [0.0039]			-0.0139*** [0.0038]	-0.0051 [0.0060]
Male * $\Delta Ideal$				-0.0026* [0.0014]				-0.0046* [0.0028]
Difference in coeffs, unpiped minus piped HHs	0.0007	-0.0036, 0.0006	0.0002	-0.0026				
F-test between unpiped and piped HHs	p = .7574	p = .7823	p = .9694	p = .7170				
Observations	130276	130276	121841	121841	35516	35516	34354	34354
R-squared	0.009	0.009	0.009	0.009	0.008	0.008	0.008	0.008

Notes: See notes to main tables. All regressions include linear and quadratic controls for mother's age and child's year of birth and dummy variables corresponding to the year of the survey wave and the child's state of residence. In columns 1, 2, 4, and 5, birth order dummies are included. In columns 1, 2, 5, and 6, birth order dummies are included. In columns 3, 4, 7, and 8, dummies for each value of  $\Delta Ideal$  are included. The sample includes children born between three and sixty-three months before the survey date. The lower bound accounts for the fact that younger children do not have a well-defined value for the dependant variable. The upper bound excludes children born far before the survey date in order to limit recall bias (see discussion in Section 7). All regressions are estimated by OLS (so coefficients represent percentage-point changes in the probability of death), and standard errors adjusted for clustering by mother. The F-test results reported test for the equality of the coefficients of interest between piped and unpiped households. They are based on fully interacted models equivalent to jointly estimating columns 1 and 5 (where the p-value is for the male coefficient); columns 2 and 6 (where the p-value is for the joint test of the linear and quadratic terms); columns 3 and 7; and columns 4 and 8 (where in the last two cases the p-value is for the male interaction with  $\Delta Ideal \geq 0$ ). If instead of estimating a fully interacted model, we constrain the control variables such as state dummies to be the same for piped and unpiped households, the four analogous p-values are 0.8224, 0.7554, 0.9486 and 0.6660.