NBER WORKING PAPER SERIES

WHY DO MOTHERS BREASTFEED GIRLS LESS THAN BOYS? EVIDENCE AND IMPLICATIONS FOR CHILD HEALTH IN INDIA

Seema Jayachandran
Ilyana Kuziemko

Working Paper 15041
http://www.nber.org/papers/w15041

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2009

We thank Kristin Plys and Luke Stein for excellent research assistance and Anne Case, Monica Das Gupta, Angus Deaton, Pascaline Dupas, Claudia Goldin, Jeffrey Hammer, Larry Katz, Alan Krueger, Adriana Lleras-Muney, Ben Olken, Emily Oster, Eliot Regett, Shanna Rose, Sam Schulhofer-Wohl, Chris Woodruff, and conference and seminar participants at Princeton, Stanford, the Pacific Conference for Development Economics, and the NBER Children's meeting for helpful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2009 by Seema Jayachandran and Ilyana Kuziemko. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

Seema Jayachandran and Ilyana Kuziemko

NBER Working Paper No. 15041
June 2009
JEL No. I1,J13,O12,O15

ABSTRACT

Medical research indicates that breastfeeding suppresses post-natal fertility. We model the implications for breastfeeding decisions and test the model's predictions using survey data from India. First, we find that breastfeeding increases with birth order, since mothers near or beyond their desired total fertility are more likely to make use of the contraceptive properties of nursing. Second, given a preference for having sons, mothers with no or few sons want to conceive again and thus limit their breastfeeding. We indeed find that daughters are weaned sooner than sons, and, moreover, for both sons and daughters, having few or no older brothers results in earlier weaning. Third, these gender effects peak as mothers approach their target family size, when their decision about future childbearing (and therefore breastfeeding) is highly marginal and most sensitive to considerations such as ideal sex composition.

Because breastfeeding protects against water- and food-borne disease, our model also makes predictions regarding health outcomes. We find that child-mortality patterns mirror those of breastfeeding with respect to gender and its interactions with birth order and ideal family size. Our results suggest that the gender gap in breastfeeding explains 14 percent of excess female child mortality in India, or about 22,000 "missing girls" each year.

Seema Jayachandran
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305-6072
and NBER
jayachan@stanford.edu

Ilyana Kuziemko
Princeton University
361 Wallace Hall
Princeton, NJ 08544
and NBER
kuziemko@princeton.edu

1 Introduction

As medical and public health researchers have long documented, breastfeeding inhibits postnatal fertility. The converse also holds: the physical demands of another pregnancy often cause a mother to stop nursing her current child. These biological constraints suggest a negative relationship between breastfeeding duration and future fertility.[1]

In developing countries, this relationship may be particularly important. First, many women lack access to modern forms of birth control, and thus rely more heavily 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.[2]

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 aside from subsequent fertility could affect a mother's propensity to breastfeed, such as her health, labor-market attachment, and education, or the price and availability of breast-milk substitutes. To help separate our hypothesis from these alternatives, 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 test 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 patterns that any alternative hypothesis would also have to explain.

First, breastfeeding increases with birth order. As mothers reach their ideal family size, their desire for more children falls and thus their demand for breastfeeding rises. Second, so long as some preference for sons exists, boys are breastfed more than girls; after the birth of a daughter, mothers are more likely to continue having children (and thus limit breastfeeding) in order to try for a son. Third, by the same logic, children with older brothers are breastfed more. Fourth,

[1] 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.

[2] See Feachem and Koblinsky (1984), Victoria et al. (1987), and Habicht, DaVanzo, and Butz (1988).

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 gender effect should be largest when a couple approaches its target family size, as at that point their decision to have another child is highly marginal and thus most sensitive to sex composition.

Our results have several potentially important implications for researchers and policymakers. First, breastfeeding's contraceptive effects, coupled with its potential health benefits to children, suggest a negative relationship between total fertility and child health. That is, a mother who wants to have many children will wean her child sooner in order to conceive again quickly, with negative consequences for the child's 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.

Second, the combination of the health benefits of breastfeeding and the gender gap in breastfeeding may help explain the "missing girls" problem in India (Sen, 1990). 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 14 percent of the gender gap in child mortality (deaths between ages one and five) in India, or 22,000 missing girls each year. Son preference is the underlying cause of this excess female mortality, but in a subtle way: Rather than resulting from parents' explicit decisions to allocate more resources to sons, the missing girls are mainly an unintended consequence of parents' desire to have more future sons.

Third, our results 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. Although the benefits of modern contraception may well swamp this potential cost, its introduction may need to be coupled with campaigns to encourage breastfeeding if policy makers wish to prevent a decline in breastfeeding rates. Moreover, 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, by giving mothers greater control over the timing and number of pregnancies, it reduces their need to suspend breastfeeding due to a new pregnancy.

The remainder of the paper is organized as follows. Section 2 provides background on the relationship between breastfeeding and fertility. Section 3 presents the model, and Section 4 describes the data. Section 5 tests the model's predictions that breastfeeding increases with birth order. Section 6 tests the predictions that breastfeeding depends on the sex 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 discusses how access to modern contraception might affect breastfeeding and offers concluding remarks.

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).

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 an 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).[3]

The degree to which fertility is suppressed depends on how many times a day the mother breastfeeds and the intensity with which the child suckles (Rous, 2001). Furthermore, while the World Health Organization warns that breastfeeding reliably prevents pregnancy only during the first sixth months after delivery, many studies argue that this window is considerably longer in developing countries. For example, Weis (1993) and Thapa (1987) find that breastfeeding inhibits

[3]Results available upon request. The analysis uses the third wave of the NFHS, the only wave with data on respondents' weight. 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 state of residence, urban/rural, sex of the child, and child's age in months.

fertility for 12 to 24 months in Bangladesh and Nepal, respectively. The longer duration of post-partum infertility in developing countries is likely related to mothers' underlying malnutrition (Huffman et al., 1987; Yadava and Jain, 1998).

An important question for our model is whether mothers actually know about the contraceptive effects of breastfeeding. Typically, a nursing mother does not menstruate (a phenomenon known as "lactational amenorrhea"), which would presumably alert her to her temporary inability to conceive. The NFHS survey we use directly asks non-pregnant women the reason they are not using birth control. Excluding those who report they are currently trying to conceive, 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, 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. Moreover, meeting the caloric requirements of tandem breastfeeding is significantly more challenging in developing countries than it would be for the mothers in the above studies. 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 current number and sex composition of her children. 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, she will not have another child in the next period. She makes this decision in order to maximize the infinite sum of discounted per-period expected utility.

Later in this section we discuss in greater detail many of the assumptions underlying this simple framework, but highlight one here: we assume that breastfeeding's only function 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. Introducing a health benefit would increase the magnitude of the male breastfeeding advantage our model generates but would not qualitatively change its other, more distinct predictions, for example regarding the interaction of gender and birth order.

3.2 Setup of the model

A mother's utility depends on the number of children, n and of sons, s she has. Her period utility is $u(n, s)$, and she has an infinite horizon with a discount rate β. 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 in period t decides whether to breastfeed the child, $b_t \in \{0, 1\}$. We assume breastfeeding perfectly inhibits fertility in the subsequent period but 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 not to breastfeed (continue having children), in which case she receives $u(n, s)$ this period and the discounted expected value function over subsequent periods. 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 and each of the infinite subsequent periods (giving a future lifetime discounted utility of $u(n, s) + \beta\frac{u(n,s)}{1-\beta} = \frac{u(n,s)}{1-\beta}$). We also assume that mothers have access to sterilization, but that even if they choose to get sterilized, they breastfeed their last child.[4]

The decision problem is therefore:

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

[4]The purpose of including the sterilization option is so that a mother can prevent further pregnancies after her last child is past the age of breastfeeding (one period in the model). Our assumption that mothers choose to breastfeed during this period instead of relying completely on sterilization does not seem unreasonable as mothers may still wish to pass on potential health benefits of breastfeeding. Alternatively, they may wish to delay their decision about sterilization given its irreversible effect and thus rely as long as possible on the contraceptive effects of breastfeeding.

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, and further that $\lim_{n\to\infty} q'(n) = -\infty$. In other words, with respect to the number of children n, the *marginal* net benefit of an additional child is strictly decreasing (and falls without bound) and the *total* net benefit displays an inverted-u shape.

The term $\lambda g(s)$ represents the additional utility from sons, with $\lambda \geq 0$ measuring the degree of son preference. We assume $g' > 0$ and $g'' < 0$ for all s. 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 $\widehat{n}$.

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

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

3.3 Predictions of the model

This subsection presents several 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^{th} child but 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 (weakly) on birth order, breastfeeding also depends on gender.

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

Proposition 2. *At any birth order, a child is more likely to be breastfed if*

(i) the child is a son rather than a daughter (holding fixed the gender of older siblings); or

(ii) more of his or her older siblings are male (holding fixed the child's gender).

Proof.

(i) By Lemma 2 (in the Appendix), a mother will breastfeed if and only if $u(n,s) \geq \frac{u(n+1,s)+u(n+1,s+1)}{2}$. 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^{th} child is a boy, so a son is more likely to be breastfed than a daughter. For sufficiently large λ, there exist integer values of n and $s \leq n$ for which the mother will choose to breastfeed a son but not a daughter. ■

(ii) From the proof to part (i), breastfeeding is increasing in s. Holding the sex of the n^{th} child fixed, s is increasing in the number of boys among the first $n-1$ children, so an n^{th} child with more brothers among his or her siblings is more likely to be breastfed. For sufficiently large λ, there exist integer values of n and $s \leq n$ for which the proposition holds strictly. ■

In the model, breastfeeding does not enter the utility function through its effects on the child who is nursed, 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 that the gender gap in breastfeeding depends non-monotonically on birth order.

Proposition 3. *The largest gap in breastfeeding of boys versus girls is at middle birth order. In other words, the gap is increasing with birth order for sufficiently low birth order, and decreasing in birth order for sufficiently high birth order.*

Proof. At sufficiently low birth order, breastfeeding is the same for sons and daughters: Lemma 1 establishes that mothers never breastfeed any child before their $\widehat{n}^{\text{th}}$. At sufficiently high birth order, breastfeeding is again the same for sons and daughters: by Lemma 3, mothers eventually breastfeed because they eventually want to stop having children. By Proposition 2(i), the only form of gender gap that can obtain is when a mother would breastfeed a son but not a daughter. If a mother's preferences are such that this would never obtain, the proposition holds trivially. If there are some birth orders (and gender compositions of older siblings) for which a mother would

breastfeed a son but not a daughter, the gender gap increases when the first such child is born and decreases when the last such child is born. ■

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 net cost of increasing n becomes large enough that it outweighs any benefit of having another son. A mother will breastfeed both a son or a daughter in this case. When weighing the costs of higher quantity with the benefit of having (in expectation) more sons at intermediate values of n, however, a mother may want to stop having children if and only if her n^{th} child is male, and will therefore breastfeed a son but not a daughter.

3.4 Predictions regarding "ideal family size"

The model also predicts that breastfeeding patterns can change abruptly when n reaches $\widehat{n}$. While a mother's "ideal" quantity of children is ill-defined in the presence of son preference, one can think of $\widehat{n}$ as one measure of the mother's preferred quantity: it is the quantity at which she would want to stop having children if she had no son preference ($\lambda = 0$). Empirically, we will use survey questions on ideal family size to test the predictions below, and we refer to $\widehat{n}$ as "ideal family size."

Proposition 4.

(i) Breastfeeding is constant for birth order below the ideal family size and can strictly increase 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 after the ideal family size has been reached.

Proof.

(i) A mother will not breastfeed for $n < \widehat{n}$, by Lemma 1. Proposition 1 establishes the remainder of the proof. The mother will begin breastfeeding at some $n \geq \widehat{n}$, 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 $n < \widehat{n}$, so there is no gender gap. Proposition 2(i) establishes the possibility of a gender gap at $n \geq \widehat{n}$. ■

3.5 Discussion

We close this section by revisiting some of the model's assumptions, beginning with our decision to model breastfeeding as a binary choice. The model can easily be relabeled so that the decision

is between short and long periods of nursing, to reflect that fact that almost all children in our sample are breastfed initially, but the binary nature of the model still deserves further attention.

First, modeling the decision as binary makes the predictions 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 orders; it is constant between birth order one and two, and then strictly increases 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 ϕ (demand for children) and λ (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.[6]

It is also worth reviewing the assumption that breastfeeding perfectly prevents conception, even though the studies cited in the previous section indicate it merely decreases fecundity. Allowing mothers to conceive while breastfeeding would only reinforce 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 her desired future fertility, and we make the choice to model only one in the interest of simplicity.[7]

[6]Stepping outside the model, fecundity declines with age, so delaying childbearing also carries the risk that the mother becomes less able to conceive.

[7]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 in the next period. Assume a mother is constrained to breastfeed only one child at a time, and she prefers to breastfeed a 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 or her life if and only if the child does not have a younger sibling, or equivalently if and only if the child was breastfed in the first period. Thus, the two channels reinforce each other.

Even with these simplifications, the model makes several 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., mothers place more value on the health of sons). 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, alternative hypotheses would have to explain why breastfeeding sharply 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 of these effects would collectively provide strong support for our claim that women's preferences over future fertility affect the decision of 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. 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 duration of nursing, 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, age three for the second wave, and 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 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^{th} 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.[8]

Summary statistics for the sample used in the breastfeeding analysis are presented in Table 1. The mean duration of breastfeeding, adjusted for censoring due to children still being breastfed, is 22.8 months. The mean value of birth order is 2.6. 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 four children per mother for the cohorts in our data.[9]

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.[10] 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 prediction relates breastfeeding duration to a child's birth order. We test the hypothesis that breastfeeding is a positive function of the mother's desire to inhibit her future fertility, which increases with birth order.

We begin by imposing as little structure as possible, allowing birth order to enter non-parametrically as a vector of dummy variables. We estimate the following OLS model:

$$Breastfeed_i = \sum_k \beta_k \cdot \mathbb{1}(BirthOrder_i = k) + X_i \cdot \gamma + a_i + \varepsilon_i. \quad (1)$$

$Breastfeed_i$ is the number of months a mother reports having breastfed child i. The term a_i is a vector of age-in-months fixed effects up to 36 months (the maximum value of the outcome)

[8] For the analysis of child mortality, we obviously do not condition on being alive and are not limited to births for which breastfeeding data are available; we describe the construction of the mortality sample in section 7.

[9] 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.

[10] 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 for very long.

that mechanically accounts for the fact that recently-born children will appear to have shorter breastfeeding spells due to right-censoring. X is a vector of covariates, and ε is an error term.

The OLS model has the advantage that the coefficients are easy to interpret, but 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:

$$h_i(t) = h_0(t) * exp\left(\sum_k \beta_k \cdot \mathbb{1}(BirthOrder_i = k) + X_i \cdot \gamma + \varepsilon_i\right). \tag{2}$$

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 coefficients on the birth-order dummy variables on the vertical axis. We show the coefficients from both an OLS 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 coefficients is 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 linear 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 are less attached to the labor force, then the coefficient on birth order could suffer from omitted variable bias. Thus, col. (2) includes a linear control for mother's years of education, a linear and quadratic control for her age, a linear and quadratic control for the child's birth year, and dummy variables for the survey wave, state of residence, urban/rural, and sex of the child. These covariates directly address mother's age and time trends as confounding factors and also 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 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 be a potential 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 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 prediction, 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 on "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.[11] 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 ith child and $Ideal_j$ is mother j's ideal family size). This variable allows us to 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 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 pref-

[11] 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?"

erences at the time they gave birth and made 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., $\Delta Ideal > 0$).[12]

Figure 2 plots the estimated coefficients on the $\Delta Ideal$ dummy variables from a regression analogous to equation (1) 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 ($\Delta Ideal = 0$), as predicted by the model.[13]

The last four columns of Table 2 show results from regressions relating breastfeeding duration and $\Delta Ideal$. 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 $\Delta Ideal \geq 0$:

$$Breastfeed_{ij} = \delta \cdot \mathbb{1}(\Delta Ideal_{ij} \geq 0) + \lambda \cdot \Delta Ideal_{ij} + X_i \cdot \gamma + a_i + \varepsilon_{ij}. \tag{3}$$

Col. (5) shows the results without our set of covariates included. Once mothers reach their ideal fertility, 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 increases by 0.58 months once a mother reaches her ideal family size.

[12] Two thirds of mothers in our sample age 35 years or older have more children than their ideal number; these older mothers are more likely to have completed their fertility, but even they might continue having more children.

[13] 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 largest t-statistic and R^2. Results available upon request.

5.4 Discussion

There appears to be a strong positive correlation between breastfeeding and birth order, consistent with the model's prediction. As the average mother has about four children, the last child would be breastfed about 0.6 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 the child's sex, the 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. Alternatively, parents in India may choose to allocate more resources to sons (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 by this logic 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.[14] Demand for an additional 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 the explanation that mothers value the health of sons more than daughters and nurse sons more in the belief that "breast milk is best," or that sons simply "take to the breast" more easily than do daughters. In the next subsection, we discuss the predictions of the model that allow

[14]Researchers find that in India, at any birth order, a couple is more likely to stop having children if they have just had a son (Das, 1987; Mutharayappa, Choe, Arnold, and Roy, 1997). Birth intervals are also longer after the birth of a son (Arnold, Choe, and Roy, 1998; Retherford and Roy, 2003).

greater separation of our hypothesis from these alternatives.

6.1 Testing our hypothesis

First, since parents' preference for another son depends on the gender composition of all previous children and not just the last one, there should be a separate effect on breastfeeding of 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. Any confounding variable 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 reach their "ideal" family size (*Proposition 4(ii)*). Again, potential 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.[15] For example, the 30^{th} percentile for boys is about one month more than that for girls. At the 70^{th} percentile the difference is about 2.5 months.

Table 3 shows the effect of the child's sex and the 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.

[15] 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 Muslim 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.

In the next two columns we examine whether the sex composition of siblings has an independent effect on breastfeeding even after accounting for the sex of the current child. Col. (4) shows that a mother already having at least one son increases the current child's breastfeeding duration by 0.28 months. 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. Thus, there is strong evidence that the gender composition of *past* births affects the breastfeeding of the *current* child—a pattern not easily 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. The main effect is the gender differential in breastfeeding for the most recent wave of data, collected in 2005; the coefficient of 0.46 is about 18 percent higher than the pooled gender differential seen in col. (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 recent fertility decline has intensified the country's sex bias (Das Gupta and Bhat, 1997).[16,17]

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 gender effects are small for both high and low birth order. Moreover, for the population as a whole, the peak effect of these variables should occur between the average "ideal family size" (around three) and average total completed fertility (around four).[18] Therefore, our model not only predicts that the gender effect takes an inverted-u shape with respect to birth order but also specifies the birth-order interval at which it peaks.

In order to investigate the effect of birth order on the gender coefficients in a flexible manner,

[16] 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) and Arnold, Kishor, and Roy (2002). This effect seems to outweigh the fact that as sex selective abortion becomes more available, girls should be born more often into families with less son preference.

[17] 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.

[18] 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.

we estimate the following equation:

$$\begin{aligned} Breastfeed_i &= \alpha \cdot Male + \sum_k \beta_k \cdot \mathbb{1}(BirthOrder_i = k) \\ &\quad + \sum_k \delta_k \cdot Male \times \mathbb{1}(BirthOrder_i = k) + a_i + X_i \cdot \gamma + \varepsilon_i. \end{aligned} \tag{4}$$

The key addition is the vector of $Male \times BirthOrder$ 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 without the additional control variables X. Sons are breastfed more than girls at every birth order, 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 parametrically as a quadratic function of birth order. These regression results are reported in the first three columns of Table 4. Col. (1) shows the OLS estimate, excluding the control variables X. The coefficients for the quadratic terms suggest that sons' breastfeeding advantage peaks when birth order equals roughly 4.1. Cols. (2) and (3) show that this peak is robust to adding in the control variables 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 mother j's ith 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 a level increase once $\Delta Ideal$ 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 male breastfeeding advantage just when mothers reach their "ideal" family size. This 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

Our model predicts 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. In this section, we verify these channels by testing whether the same variables that predict breastfeeding also predict fertility-stopping.

Table 5 shows 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 before. 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 columns focus on the child's gender and the sex composition of the child's siblings. Being male increases the probability of not having a younger sibling, as does having at least one older brother—the same 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 greater for mothers who have reached or surpassed their ideal family size.

6.5 Heterogeneity in son preference

We next test whether gender patterns in breastfeeding vary with the intensity of the mother's son preference. Our first measure of son preference is the male-to-female ratio of births in the respondent's state (using data from 2001, the most recent census year), which has a sample average of 1.07. Given sex-selective abortion, this statistic should be positively correlated with regional differences in son preference, and indeed we find that it is systematically lower in south India, where son preference is recognized to be less intense.

Table 6, col. (1) shows that the son advantage in breastfeeding is significantly larger in states with stronger son preference. The estimates imply that in states such as Kerala with sex ratios near one, mothers breastfeed sons only 0.2 months longer than daughters; in places such as Haryana or Punjab where the sex ratio is about 1.16, boys are breastfed 0.6 months longer. This basic result is consistent with both a stop-after-a-boy fertility preference and with parents caring more about sons' health.

We next examine how the breastfeeding gap increases discretely when the mother has reached her ideal family size, which is a more distinctive prediction of son-biased stopping rules. As seen in

col. (2), our predicted patterns are stronger in states where the sex ratio is higher. That is, in states where having sons is especially valued, the mother's decision to continue having children beyond her ideal number, and thus her decision about how long to breastfeed, depends more heavily on the gender of her children.

Another measure of son preference is how many sons the mother says she ideally would like to have. The survey asked the mother to break down her ideal number of children into her ideal number of sons (sample mean of 1.4), ideal number of daughters (mean of 1.0) and ideal number for which she was indifferent about their gender (mean of 0.3). The first prediction we test is that, just as a mother is more likely to stop having children once she reaches her ideal family size, she is more likely to stop once she reaches her ideal number of sons. We construct a variable $\Delta IdealSons$ that is analogous to $\Delta Ideal$; $\Delta IdealSons$ is the mother's current number of sons minus her ideal number of sons. (For each child-level observation, the mother's current number of sons is the number she has at the time of the child's birth, and in the case of a boy, it is inclusive of the child himself.) As seen in col. (3), the duration of breastfeeding increases by 0.33 months once the mother reaches her ideal number of sons, controlling for a linear effect of $IdealSons$. Once a mother has reached *both* her ideal number of sons and her ideal family size, her children are breastfed 0.83 months longer.

We next test a second prediction about $IdealSons$. The scenario in which a child's gender is most pivotal to the breastfeeding-cum-fertility decision is when the mother, by giving birth to a son, has just reached her ideal number of sons. Because the child was male, her son preference is satisfied; had she given birth to a girl instead, her son preference would not have been met. Thus, empirically, the male advantage in breastfeeding should be most pronounced when the mother has exactly reached her ideal number of sons, compared to when she has either fewer or more than her ideal number of sons. As seen in col. (4), we find precisely this pattern: relative to all other values of $\Delta IdealSons$, the case where $\Delta IdealSons = 0$ is when boys are breastfed most, relative to girls.[19] In short, we find strong evidence that the son advantage in breastfeeding varies as predicted with heterogeneity across regions in son preference, and breastfeeding duration varies across mothers who differ in their ideal number of sons, with quite specific predictions born out in the data.

6.6 Decomposing the male breastfeeding advantage

Thus far, we have provided a variety of evidence that boys are breastfed longer than girls, and that the gender interactions with birth order and ideal family size are consistent with distinct

[19]The results in columns (3) and (4) are unchanged when we control for a set of fixed effects for the number of sons the mother has and the identification comes from variation across mothers in their ideal number of sons.

predictions of our model in which mothers decide whether to breastfeed based on their future fertility. These predictions allow us to separate our mechanism from the more standard son-preference explanation that mothers simply give fewer resources—including breast milk—to daughters.

However, we have yet to calculate just what share of the male breastfeeding advantage our hypothesis explains. Overall, there is a 0.39-month male breastfeeding differential (Table 3, col. 2), which could be driven by our hypothesis, the more standard "feed the boys, starve the girls" explanation, or something else. One approach to gauging the importance of our mechanism is to assume that it does not enter until mothers reach their ideal family size. This estimate is provided by the coefficient on $Male * \Delta Ideal \geq 0$ in col. (5) of Table 4, which is 0.495. As the mean of $\Delta Ideal \geq 0$ is 0.55, we estimate that our mechanism explains 0.272 months (0.495*0.55) or 70 percent (0.27/0.39) of the male breastfeeding advantage. Another approach is to estimate the "feed the boys" effect and all other explanations besides ours as the male coefficient conditional on the mother's total number of children and total number of sons; in our model, these variables (n and s) fully determine breastfeeding. In unreported results, we find the male coefficient is then 0.146, suggesting that our hypothesized mechanism explains 62 percent (1 - 0.15/0.39) of the breastfeeding gender gap.

As further evidence that the conventional son-preference story is unlikely to explain our results, we find no evidence that vaccinations or doctor visits exhibit the same distinct patterns we find for breastfeeding. Table 7 shows the results for our main specifications when "child received at least one vaccination" serves as the dependent variable. While there is indeed a strong male advantage, none of the other patterns found for breastfeeding hold. Col. (1) shows vaccinations are in fact decreasing with parity. Col. (3) shows that the sex of older siblings does not affect vaccinations, and col. (5) shows that the male advantage seems to actually decrease once the mother reaches her ideal family size. We find similar results (available upon request) using "total number of vaccinations" or "received medical attention conditional on having a fever" as the dependent variable.

6.7 Discussion

In this section we have presented evidence in support of the more demanding predictions of our hypothesis that 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' health 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 takes an inverted-u shape: 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 occurs at a birth order

between average ideal family size in our sample and realized family size in 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.

We also find that future fertility exhibits the same patterns, and that the results vary with measures of son preference. Finally, we find that other inputs to child health such as vaccinations or doctor visits also show a son advantage, but they exhibit none of the more distinct patterns we find for breastfeeding that are predicted by our model. Parents wanting to provide more health inputs to sons no doubt plays a key role in children's health outcomes in India. However, we estimate it accounts for only about a third of the gender differential in breastfeeding.

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 discuss the medical evidence on the health benefits of breastfeeding. We then review the patterns for 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 for breastfeeding. Finally, we use our NFHS sample to directly test whether child survival exhibits the same relationships 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, 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. 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 between the ages of 18 and 36 months who have been weaned have three times the mortality rate of those still

being breastfed; they attribute one third of the deaths in this age range 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

Previous work on the relationship between birth order and child health appears to undermine our prediction that higher birth-order children, by virtue of nursing longer, enjoy better health outcomes. 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 medical literature primarily has identified mechanisms that favor lower birth-order children (Arad et al., 2001).

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) uses food consumption data to show that parents favor lower birth-order children in India. Our own results from Table 7 suggest that parents vaccinate lower birth-order children significantly more than later children. Finally, the mechanical negative correlation between birth order and family size may impart a health advantage to low-parity children. Thus, the favorable circumstances of low-parity children in terms of resource allocation and family size 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 (the ratio of boys to girls) increases considerably in childhood; in China, by contrast, the problem is almost fully realized at or shortly after birth (Das Gupta, 2005). In India, girls have a forty percent higher mortality rate, relative to boys, between the ages of one and five but an equal mortality rate before age one (Acharya, 2004). This pattern coincides with our finding that most of the gender gap in breastfeeding does not arise until about twelve months after birth (Figure 3).

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 the 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.[20] The gender gap in breastfeeding is concentrated between the ages of 12 and 36 months, and there is no apparent gender gap during the first few months of the child's life. Hence, we test for breastfeeding effects on the probability a child dies between 12 and 36 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 evidence suggests that having a larger family size could lead to higher mortality rates for girls (Yamaguchi, 1989; Clark, 2000; Jensen, 2003; Rosenblum, 2008). We can make progress in separating the effects of family size and 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.

[20] 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.[21] 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.[22]

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 use 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 outcome is not defined). This approach 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.[23]

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 go against our breastfeeding hypothesis (results available upon request). We tentatively conclude that the resource advantage of first-born and low-birth-order children documented in past work and in our Table 7 outweighs the health benefits that high-birth-order children receive via breast milk. Furthermore, birth order is mechanically linked to family size and, as children from larger families tend to have worse outcomes, composition bias works against finding any breastfeeding benefit for high-parity children.

The remainder of this section examines the effects on child mortality of gender and its interac-

[21] We do not regress mortality on breastfeeding duration because breastfeeding duration is censored at the age when the child dies and the results would suffer from reverse causality.

[22] 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.

[23] 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.

tions with birth order and ideal family size. Table 8 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, which is 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. 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, we estimate the same regressions as in cols. (1) to (4) but this time use 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) through (8), the coefficients of interest are all smaller in magnitude or even 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, not merely variables of interest, is interacted with $Piped$) that are 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; 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 in which breastfeeding should 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 8 would suggest that something correlated with breastfeeding duration, and not breastfeeding duration itself, is driving those results. Appendix Table 1 is the exact analogue to Table 8 with the exception that 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 8, male survival advantage emerges in households without piped water once mothers reach their "ideal" family size; but, unlike in Table 8, the effect is essentially identical between households with and without piped water. For each of our 0-to-3-month-mortality regressions, 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.

Of course, whether a household has piped water could be correlated with other factors that might independently affect gender-specific mortality patterns. We thus estimate regressions in which every variable is also interacted with an indicator variable for whether the household is in a rural area. As the first four columns in Appendix Table 2 show, for 12-to-36 month mortality the differences between households with and without piped water generally remain statistically significant. The magnitude of the coefficients does fall in some cases, suggesting some role for omitted variables bias. For example, the difference in the male coefficient between unpiped and piped households was -0.0047 with a p-value of 0.0004 without the rural interactions (see Table 8, col. 1) and is -0.0034 with a p-value of 0.0245 with the rural interactions. The final four columns of Appendix Table 2 show that for mortality in the first three months of life, as in Appendix Table 1, the gender differences between households with and without piped water are not significant and in some cases wrong-signed.

In summary, in settings in which breastfeeding should play an especially important role in child health (households without piped water) and in which breastfeeding gender differences are most pronounced (between the ages of one and three), the variation in mortality with respect to gender, birth order and $\Delta Ideal$ is remarkably similar to the variation in breastfeeding with respect to these same variables. In settings in which breastfeeding should play little role or in which no breastfeeding gender gap exists, these mortality patterns 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 lack of breastfeeding between the ages of 12 and 36 month increases the probability the child dies during this interval by a factor of 2.5.[24] About

[24]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.

58.6 percent of the children aged 12 to 36 months in our sample are being breastfed, as (roughly) seen in Figure 3, and about 2.03 percent die during this age interval. Thus, the implied mortality rate for breastfed children solves the equation $0.586x + 2.5(1 - 0.586)x = 2.03$. This calculation yields a mortality rate of 1.25 percent for breastfed children and 3.13 percent ($1.25 * 2.5 = 3.13$) for non-breastfed children in this age interval.

Between the ages of 12 and 36 months, girls in our sample have 0.75 percentage points of excess mortality (2.42 percent versus 1.67 for boys) and we now calculate the share of this gap that is explained by breastfeeding differences. In this age range, 56 percent of girls and 60 percent of boys are breastfed (again, see Figure 3). Therefore, if breastfeeding differences were the only source of mortality variation, boys would be expected to have a mortality rate of $0.60 * 1.25 + 0.40 * 3.13 = 2.00$ percent compared to $0.56 * 1.25 + 0.44 * 3.13 = 2.08$ percent for girls, and the mortality gap would equal 0.08 percentage points. Therefore, breastfeeding differences explain 11 percent ($0.08/0.75 = 0.11$) of the mortality gender gap between the ages of 12 and 36 months, or, given that each year in India about 13.2 million girls survive to age one, about 10,600 ($13,200,000 * 0.0008$) excess female deaths per year.

We can also use our results from the previous subsection to perform a similar calculation. Our comparison of gender gaps in mortality between households with and without piped water implies that breastfeeding differences account for about 0.34 percentage points of excess female mortality in households without piped water (taking the more conservative estimate given in Appendix Table 2, col. 1). As roughly 75 percent of Indian households lack piped water, the breastfeeding gender gap appears to explain $0.34 * 0.75 = 0.255$ percentage points (34 percent) of excess female mortality between the ages of one and three, or about $0.00255 * 13,200,000 = 33,600$ missing girls.

Taking the mid-point of these estimates suggests breastfeeding accounts for roughly 22,000 excess female deaths, which is equivalent to about 22 percent of excess female mortality between the ages of one and three or about 14 percent of excess female child mortality (defined as deaths between ages one and five).[25] Of course, readers should consider these estimates with caution. Our difference-in-differences estimate is a rough approximation. 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

[25] Past work has shown that the gender gap in child mortality is 1.2 percent (Pandey et al., 1998). Averaging the results from our two estimates suggests that breastfeeding accounts for 0.168 percentage points (or $0.168/1.2 = 14.0$ percent) of excess female mortality. Any immunities or other protections breastfeeding confers on children after weaning would negatively bias this estimate.

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, our results suggests that in contrast to conventional explanations, excess female mortality due to differential breastfeeding is largely an unintended consequence of parents' desire to have more sons rather than an explicit decision to allocate fewer resources to daughters.

8 Conclusion

This paper began by arguing that the duration of breastfeeding negatively correlates with the mother's likelihood of a subsequent birth. 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 incorporates 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 that allow them to breastfeed longer without being 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, by 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 their 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 0.6 months longer than his oldest sibling, and most of that difference comes from a discrete increase once mothers reach their "ideal" family size. Sons are breastfed 0.4 months longer than daughters, and having older brothers also significantly increases how long a child is breastfed. The son advantage is small for low and high

birth order and peaks around a birth order value of four; it also displays a discrete jump when mothers reach their ideal family size.

Parents valuing sons' health more than daughters' health is unlikely to explain these results. We estimate that, instead, two thirds of the son advantage in breastfeeding is due to the value parents place on having future sons. As further support for this interpretation, we find that while other inputs to child health such as vaccinations show a son advantage, they exhibit none of the more distinct patterns our model predicts for breastfeeding.

Given the well-documented health benefits of breastfeeding for children in developing countries, we test whether mortality patterns with respect to gender, birth order and ideal family size mirror the breastfeeding patterns. Indeed, boys, especially those of intermediate birth order and those born to mothers who have reached their "ideal" family size, have a lower mortality rate between the ages of 12 and 36 months, the age range where we find a gender gap in breastfeeding. These results are driven by children living in households without 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 twenty percent of excess female deaths between the ages of one and three years, or, equivalently, 14 percent of excess female child mortality (deaths between the ages of one and five years) in India. We also estimate that the fact that mothers breastfeed daughters less than sons leads to 22,000 "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 decisions, have unintended health consequences.

The relationship between breastfeeding and access to birth control poses an interesting direction for future research. This relationship is theoretically ambiguous. As we documented earlier, two phenomena underlie the negative relationship between breastfeeding duration and subsequent fertility: breastfeeding 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 over which a mother can nurse her current child.[26]

[26] We know of no strategy to isolate exogenous variation in access to contraceptives in India during our study period. As a suggestive test, we constructed an imperfect measure of "access" in our data, the average usage of contraception in the respondent's primary sampling unit (village), excluding herself. We find some evidence suggesting that access

If contraception crowds out breastfeeding, then policy makers might consider pairing contraceptive campaigns with promotion of breastfeeding or improvements in water quality. Conversely, if contraception enables a mother to breastfeed her children longer because she can space them further apart, then policies that expand access to contraception might have an added benefit of encouraging breastfeeding. 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.

to condoms, IUDs, and the pill decreases breastfeeding, while access to sterilization increases it. Our results are consistent with sterilization allowing mothers to breastfeed their last child longer, as they are uninterrupted by a subsequent pregnancy, and reversible technologies being more conducive to birth-spacing and acting as a substitute for breastfeeding.

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.

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.

Huffman, S. L., F. K. A. H. A., and P. Streble (1987): "Nutrition and Fertility in Bangladesh: Breastfeeding and Post-Partum Amenorrhoea," *Population Studies*, 41, 447462.

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.

Pandey, A., M. K. Choe, N. Y. Luther, D. Sahu, and J. Chand (1998): "Infant and Child Mortality in India," National Family Health Survey Subject Report.

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), 115–118.

Pritchett, L. H. (1994): "Desired Fertility and the Impact of Population Policies," *Population and Development Review*, 20(1), 1–55.

Retherford, 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.

Retherford, 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. (1990): "More Than 100 Million Women Are Missing," *New York Review of Books*, 37(20).

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.

Thapa, S. (1987): "Determinants of Fertility in Nepal: Applications of an Aggregate Model," *Journal of Biosocial Science*, 19(3), 351–365.

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.

Weis, P. (1993): "The Contraceptive Potential of Breastfeeding in Bangladesh," *Studies in Family Planning*, 24(2), 100–108.

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.

Yadava, K., and S. Jain (1998): "Postpartum Amenorrhoea in Rural Eastern Uttar Pradesh, India," *Journal of Biosocial Science*, 30, 227–243.

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

Lemma 1. *$\widehat{n}$ exists, and $\widehat{n} > 0$. For all $n < \widehat{n}$, a mother will have the $(n+1)^{st}$ child regardless of sex preference λ and regardless of the sex composition of existing children.*

Proof. Given strict concavity of q, that $\lim_{n\to\infty} q'(n) = -\infty$, and our assumptions that $q(1) > q(0)$, we know $q(n)$ must achieve a maximum at some unique $n_{\max} > 0$. 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 $\widehat{n} = \max\{n \mid h(n) \geq 0\}$ is equal to the unique n such that $h(n) = 0$.

For any $n < \widehat{n}$, the benefit to having another child is strictly positive regardless of the sex composition of existing children (q is strictly greater than zero in this region and λg is always increasing), so mothers always choose to have the $(n+1)^{\text{st}}$ child. ∎

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

Proof. In words, the lemma states that a mother will not breastfeed her n^{th} child and instead will have her $(n+1)^{\text{st}}$ child if and only if the expected value of having *exactly* one more child exceeds the value of stopping now.

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 the mother has her $(n+1)^{\text{st}}$ child, so $V^{b=0} > V^{b=1}$. This gives $\frac{V(n+1,s)+V(n+1,s+1)}{2} > \frac{u(n,s)}{1-\beta} \geq \frac{u(n+1,s)+u(n+1,s+1)}{2(1-\beta)}$. 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 $u(n+1,s) > \frac{u(n+2,s)+u(n+2,s+1)}{2}$, then we have back our original problem with $n+1$ replacing n, implying that the mother will continue to have children if she reaches $(n+2,s)$. By induction, as long as the mother had daughters she would continue having children indefinitely, contradicting Lemma 3 which states that the number of children is bounded above. Therefore, all

that remains to prove sufficiency is to show that

$$u(n,s) > \frac{u(n+1,s)+u(n+1,s+1)}{2} \Rightarrow u(n+1,s) > \frac{u(n+2,s)+u(n+2,s+1)}{2}.$$

Rearranging terms and substituting in the definition of u gives,

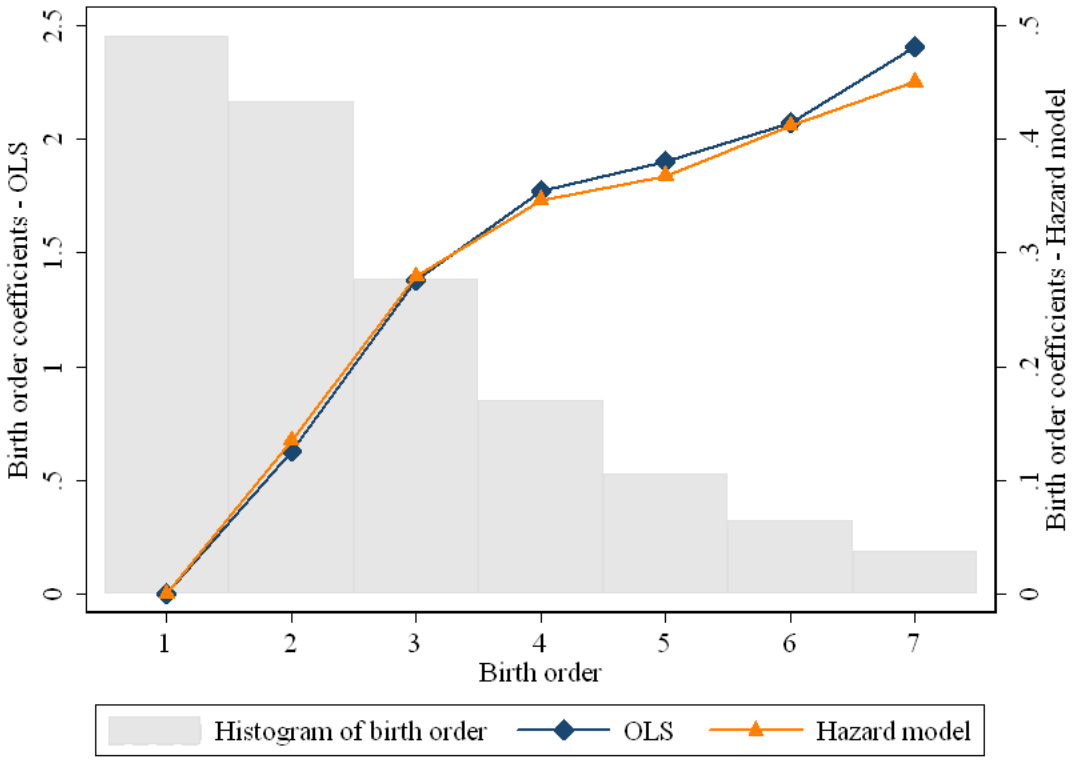
$$-\lambda\frac{g(s+1)-g(s)}{2} > q(n+1)-q(n) \Rightarrow -\lambda\frac{g(s+1)-g(s)}{2} > q(n+2)-q(n+1)$$

which holds by the concavity of q. ■

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

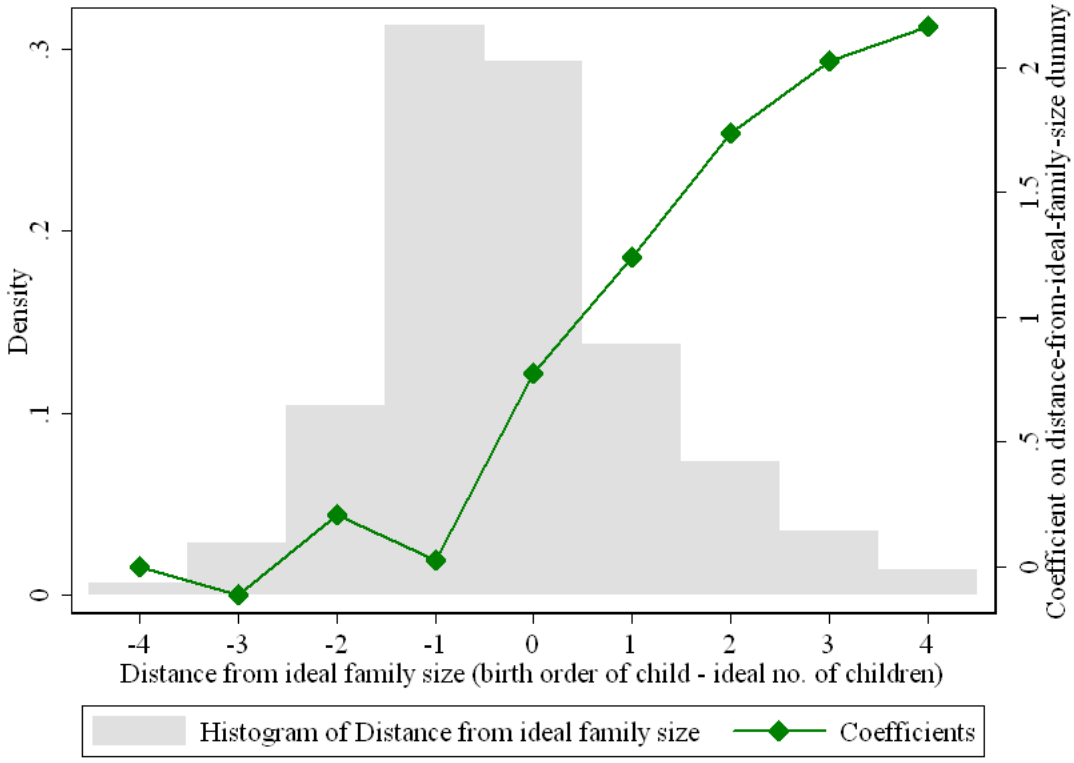
Proof. The proof of Lemma 1 established that $h(n) \equiv q(n+1)-q(n)$ is a continuous, strictly decreasing function. Since $\lim_{n\to\infty} q'(n) = -\infty$, we have that $\lim_{n\to\infty} h(n) = -\infty$ by the Fundamental Theorem of Calculus. Let $\bar{n}$ be equal to the unique n such that $h(n) = -\lambda\big[g(1)-g(0)\big]$. Consider a woman who has $n \geq \bar{n}$ children; the marginal net cost of having the $(n+1)^{\text{st}}$ child is sufficiently large that even if doing so conferred the maximal marginal benefit associated with an additional son ($\lambda\big[g(1)-g(0)\big]$, which arises if the child is her first son), it would not exceed the utility loss associated with the extra child, $q(n)-q(n+1) \geq q(\bar{n})-q(\bar{n}+1)$. Thus having any child beyond the $\bar{n}^{\text{th}}$ will necessarily decrease a mother's period utility, and it is *never* optimal to have more than $\bar{n}$ children, regardless of the sex composition of existing children. ■

Figure 1: Breastfeeding duration, by birth order



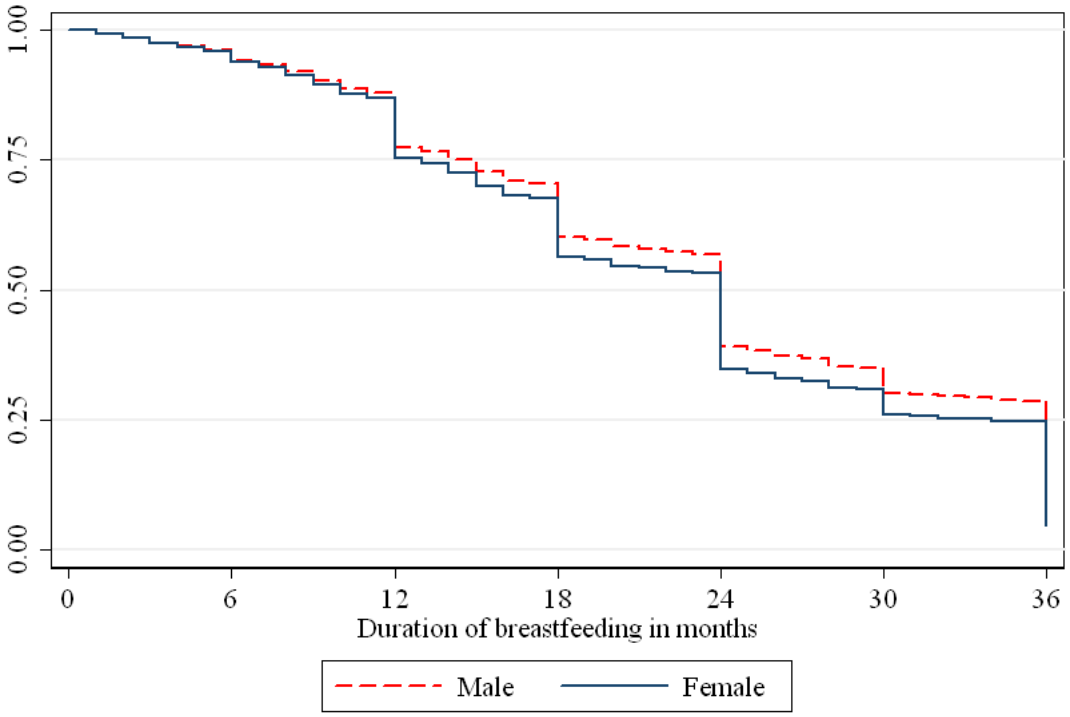
The figure plots the coefficients for birth-order dummies from a regression with breastfeeding duration in months as the dependent variable. The OLS model includes age-in-month fixed effects; no other control variables are included. The omitted category is birth order 1, for which the coefficient is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

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



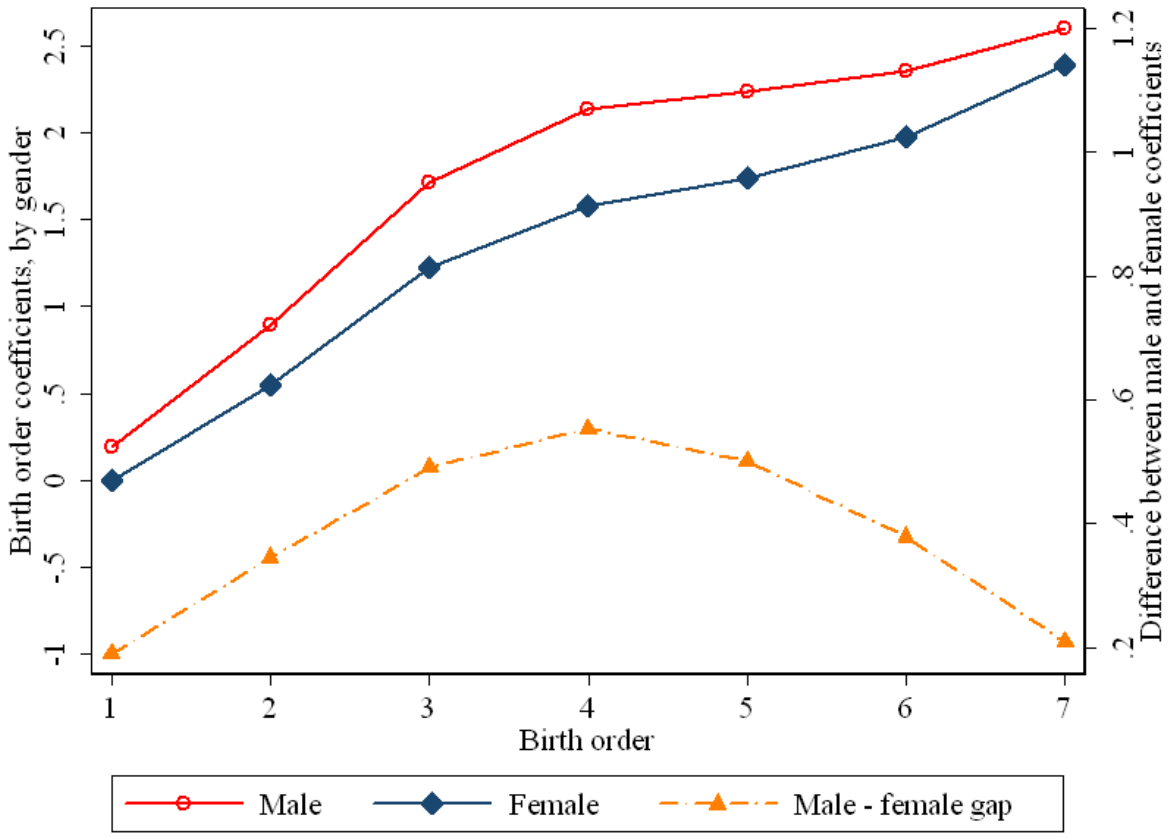
The figure plots the coefficients for "distance from ideal family size" dummies from a regression with breastfeeding duration in months as the dependent variable. Distance from ideal family size is defined as the child's birth order minus the mother's ideal family size. The omitted category is distance 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



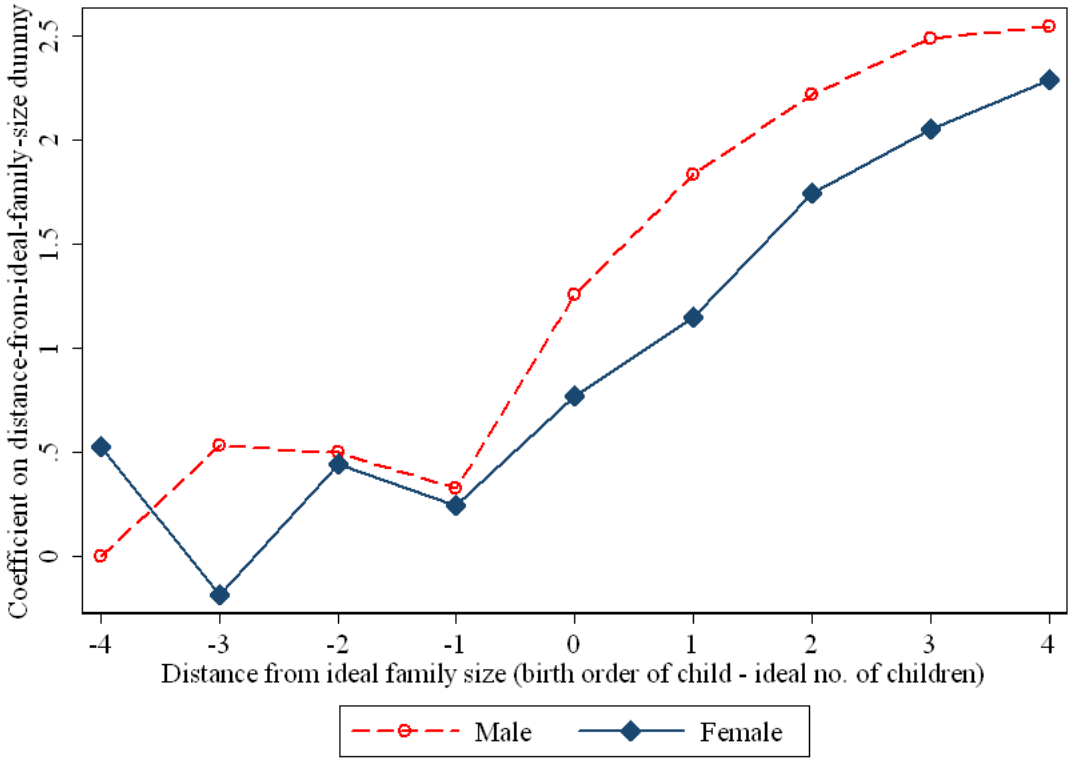
The figure plots the proportion of children, by gender, who are still being breastfed at the duration (age) given on the horizontal axis.

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



The solid lines plot the gender-specific coefficients for birth-order dummies from 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 from 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)

	All	*Birth order 1 or 2*	*Birth order >2*
Breastfeeding duration (months)	22.815 (12.978)	21.554 (12.864)	24.583 (12.846)
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)
ΔIdeal (birth order – ideal no. of children)	-0.190 (1.412)	-0.915 (0.894)	0.882 (1.354)
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 has no younger siblings	0.795 (0.403)	0.769 (0.422)	0.833 (0.373)
Child was vaccinated	0.807 (0.395)	0.851 (0.356)	0.744 (0.436)
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 the 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, and who are singletons. The 1992, 1998 and 2005 waves account for 36.9%, 24.6%, and 38.5% of the observations, respectively. Breastfeeding duration is adjusted for censoring due to children still being breastfed; the unadjusted sample mean (standard deviation) is 14.8 (9.0) months. "Male share of mother's children" and "Mother has at least one son" include the child him or herself. "Child was vaccinated" is a dummy for whether the child received any of the standard childhood immunizations for polio, DPT (diphtheria/pertussis/tetanus), and measles.

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

	OLS estimation of months breastfed		*Cox est. of hazard rate*	*OLS estimation of months breastfed*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Birth order	0.4640*** [0.0124]	0.2103*** [0.0179]	-0.0612*** [0.0042]				
ΔIdeal ≥ 0				1.0723*** [0.0399]	0.8757*** [0.0454]	0.4007*** [0.0624]	0.5763*** [0.0616]
ΔIdeal						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: *** p<0.01, ** p<0.05, * p<0.1. The unit of observation is the child and we cluster standard errors (in brackets) by mother to account for mothers who have more than one child in the sample. We define the variable *Ideal* as the response of the child's mother to the question "What is your ideal number of children?" We define *ΔIdeal* as *Birth order – Ideal.* The specifications in columns 2, 3, 5, and 7 include linear and quadratic controls for mother's age and child's year of birth, as well as 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 *ΔIdeal* 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

	OLS estimation of months breastfed		*Cox est. of haz. rate*	*OLS estimation of months breastfed*		
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.3681***	0.3887***	-0.1034***	0.2444***	0.2623***	0.4581***
	[0.0384]	[0.0373]	[0.0087]	[0.0486]	[0.0546]	[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: *** p<0.01, ** p<0.05, * p<0.1. Standard errors, in brackets, are adjusted for clustering by mother. See notes to previous tables. "Additional covariates" include linear and quadratic controls for mother's age and child's year of birth, as well as 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.

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

	OLS estimation of months breastfed		*Cox estimation of hazard rate*	*OLS estimation of months breastfed*	
	(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 (ΔIdeal $\geq$ 0)				0.4922*** [0.1262]	0.4951*** [0.1226]
Male x ΔIdeal				-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: *** p<0.01, ** p<0.05, * p<0.1. Standard errors, in brackets, are adjusted for clustering by mother. 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. We define the variable *Ideal* as the response of the child's mother to the question "What is your ideal number of children?" We define *ΔIdeal* as *Birth order – Ideal.* Columns 1 to 3 include birth-order fixed effects and columns 4 and 5 include fixed effects for each value of *ΔIdeal*. The covariates are linear and quadratic controls for mother's age and child's year of birth, as well as 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.

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

	Dependent variable = Child has no younger siblings				
	(1)	(2)	(3)	(4)	(5)
Male	0.0302*** [0.0020]	0.0272*** [0.0020]	0.0154*** [0.0025]	-0.0012 [0.0071]	-0.004 [0.0044]
Birth order	0.0175*** [0.0009]				
ΔIdeal ≥ 0		0.1159*** [0.0032]			
Mother has at least one son			0.0283*** [0.0033]		
Male * Birth order				0.0241*** [0.0049]	
Male * (Birth order)^2				-0.0034*** [0.0007]	
Male * (ΔIdeal ≥ 0)					0.0538*** [0.0063]
Fixed effects?	None	None	Birth order	Birth order	ΔIdeal
Observations	110183	104456	110183	110183	104456
R-squared	0.331	0.344	0.338	0.338	0.346

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors, in brackets, are adjusted for clustering by mother. 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, as well as dummy variables corresponding to the year of the survey wave and the child's state of residence. We define the variable *Ideal* as the response of the child's mother to the question "What is your ideal number of children?" We define *ΔIdeal* as *Birth order – Ideal.* Column 2 also includes a linear term for *ΔIdeal*, and column 5 includes *Male*ΔIdeal*. All regressions are estimated by OLS (so coefficients represent percentage-point changes in the probability of being the youngest child at the time of the survey). The regression sample is the same as that used when breastfeeding is the dependent variable.

Table 6: Heterogeneity in son preference

	Dependent variable = Duration of breastfeeding in months			
	Regional heterogeneity		Individual heterogeneity	
	(1)	(2)	(3)	(4)
Male	-1.9225**	1.7095		0.3664*
	[0.8424]	[1.8901]		[0.2024]
Male * Sex ratio	2.1449***			
	[0.7813]			
Male * (ΔIdeal ≥ 0) * Sex ratio		6.0733**		
		[2.5212]		
ΔIdealSons ≥ 0			0.3287***	
			[0.0670]	
ΔIdeal ≥ 0			0.5026***	
			[0.0628]	
Male * (ΔIdealSons = 0)				0.2047**
				[0.0956]
Male * (ΔIdealSons > 0)				-0.0603
				[0.1286]
Fixed effects included	Birth order	Birth order	None	Male * Birth order
Other covariates	None	Male * ΔIdeal	ΔIdealSons, ΔIdeal	ΔIdealSons = 0, ΔIdealSons > 0
Observations	110183	104456	104333	105054
R-squared	0.527	0.525	0.524	0.525

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors, in brackets, are adjusted for clustering by mother. See notes to previous tables. We define *IdealSons* as the mother's answer to the question, "What is your ideal number of sons?" We define *ΔIdealSons* as *total number of sons – IdealSons*. 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 addition to those listed. The results in columns 3 and 4 are essentially identical if we also include fixed effects for the total number of sons.

Table 7: Vaccinations as a function of birth order, gender, and ideal fertility (placebo test)

	Dependent variable = Child received any vaccination				
	(1)	(2)	(3)	(4)	(5)
Male	0.0192*** [0.0021]	0.0193*** [0.0021]	0.0215*** [0.0029]	0.0023 [0.0071]	0.0242*** [0.0045]
Birth order	-0.0149*** [0.0011]				
ΔIdeal ≥ 0		-0.0059* [0.0032]			
Mother has at least one son			-0.0047 [0.0037]		
Male * Birth order				0.0075 [0.0055]	
Male * (Birth order)^2				-0.0003 [0.0008]	
Male * (ΔIdeal ≥ 0)					-0.0076 [0.0065]
Fixed effects?	None	None	Birth order	Birth order	ΔIdeal
Observations	109770	104083	109770	109770	104083
R-squared	0.238	0.231	0.238	0.238	0.233

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are adjusted for clustering by mother. See notes to previous tables. Dependent variable is an indicator variable for whether the child received a vaccination for polio, DPT (diphtheria/pertussis/tetanus), or measles. All regressions include linear and quadratic controls for mother's age and child's year of birth, as well as dummy variables corresponding to the year of the survey wave and the child's state of residence We define the variable *Ideal* as the response of the child's mother to the question "What is your ideal number of children?" We define *ΔIdeal* as *Birth order – Ideal* Column 2 also includes a linear term for *ΔIdeal*, and column 5 includes *Male*ΔIdeal*. All regressions are estimated by OLS (so coefficients represent percentage-point changes in the probability of receiving at least one vaccination). The regression sample is the same as that used when breastfeeding is the dependent variable.

Table 8: Child mortality as a function of birth order, gender, and ideal fertility

	Dependent variable = Child died between ages 12 and 36 months							
	Household lacks piped water				Household has piped water			
	(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)^2		0.0005 [0.0003]				0.0003 [0.0005]		
Male * (ΔIdeal $\geq$ 0)			-0.0083*** [0.0017]	-0.0039 [0.0028]			0.0007 [0.0021]	0.0042 [0.0030]
Male * ΔIdeal				-0.0018* [0.0010]				-.0017 [0.0014]
Difference in coeffs, unpiped 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: *** p<0.01, ** p<0.05, * p<0.1. Standard errors, in brackets, are adjusted for clustering by mother. All regressions include linear and quadratic controls for mother's age and child's year of birth, as well as 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 *ΔIdeal* 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). The F-test results reported test for the equality of the coefficients of interest between households with and without piped water. 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 *ΔIdeal* $\geq$ *0*). If instead of estimating a fully interacted model, we constrain the control variables such as state dummies to be the same for households with and without piped water, the four analogous p-values are 0.0004, 0.0380, 0.0003 and 0.0357.

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

	Dependent variable = Child died between ages 0 and 3 months							
	Household lacks piped water				Household has piped water			
	(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)^2		0.0006 [0.0004]				0.0000 [0.0008]		
Male * (ΔIdeal ≥ 0)			-0.0137*** [0.0025]	-0.0076* [0.0039]			-0.0139*** [0.0038]	-0.0051 [0.0060]
Male * ΔIdeal				-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: *** p<0.01, ** p<0.05, * p<0.1. Standard errors, in brackets, are adjusted for clustering by mother. See notes to main tables. All regressions include linear and quadratic controls for mother's age and child's year of birth, as well as 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 *ΔIdeal* 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). 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 *ΔIdeal ≥ 0*). If instead of estimating a fully interacted model, we constrain the control variables such as state dummies to be the same for households with and without piped water, the four analogous p-values are 0.8224, 0.7554, 0.9486 and 0.6660.

Appendix Table 2: Mortality results controlling for interactions with rural status

	Dependent variable = Child died between ages 12 and 36 months				*Dependent variable = Child died between ages 0 and 3 months*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Male * Lacks piped water	-0.0034**	0.0041	0.0004	0.0016	0.0015	0.0031	-0.0009	0.0042
	[0.0015]	[0.0051]	[0.0023]	[0.0032]	[0.0025]	[0.0087]	[0.0038]	[0.0057]
Male * Birth order * Lacks piped water		-0.0045				-0.0016		
		[0.0040]				[0.0040]		
Male * (Birth order)^2 * Lacks piped water		0.0005				0.0004		
		[0.0006]				[0.0010]		
Male * (ΔIdeal $\geq$ 0) * Lacks piped water			-0.0068**	-0.0086*			0.0031	-0.0047
			[0.0031]	[0.0046]			[0.0051]	[0.0079]
Male * ΔIdeal * Lacks piped water				0.0010				0.0041
				[0.0019]				[0.0034]
Observations	163342	163342	152996	152996	165792	165792	156195	156195
R-squared	0.011	0.012	0.011	0.011	0.010	0.011	0.011	0.011

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors, in brackets, are adjusted for clustering by mother. See notes to main tables. Columns 1 and 5 include the main effects of *Male* and *Lacks piped water*, as well as the interaction term, *Male*Rural*. Similarly, columns 2 through 4 and 6 through 8 include all of the relevant main effects. In addition, columns 2 and 6 control for *Male*Rural*, *Male*Birth order*Rural*, and *Male*(Birth order)^2*Rural*; columns 3 and 7 control for *Male*Rural* and *Male*(ΔIdeal ≥ 0)*Rural*; and columns 4 and 8 control for *Male*Rural*, *Male*(ΔIdeal ≥ 0)*Rural* and *Male*ΔIdeal*Rural*. Columns 1, 2, 5 and 6 include birth order fixed effects and their interactions with *Lacks piped water* and *Rural*. Columns 3, 4, 7, and 8 include *ΔIdeal* fixed effects and their interactions with *Lacks piped water* and *Rural*. All specifications include our standard covariates (linear and quadratic controls for mother's age and child's year of birth, as well as dummy variables corresponding to the year of the survey wave and the child's state of residence), plus their interactions with *Lacks piped water* and with *Rural*.