

# Why Do Mothers Breastfeed Girls Less Than Boys?

## Evidence and Implications for Child Health in India\*

Seema Jayachandran

Stanford University and NBER

Ilyana Kuziemko

Princeton University and NBER

September 2010

### Abstract

Breastfeeding is negatively associated with future fertility both because nursing temporarily reduces fecundity and because mothers generally stop nursing upon a subsequent pregnancy. If parents have a fertility preference for sons, as is common in developing countries such as India, they will wean daughters more quickly to “try again” for a boy, potentially increasing girls’ exposure to contaminated food and water as an unintended consequence. We develop a model of breastfeeding that predicts that breastfeeding duration increases with birth order, and rises significantly once a mother reaches her ideal family size; sons are nursed longer, as are children with older brothers; and these gender differences peak at birth order near ideal family size, when the decision to have another child is most marginal. We confirm each of these predictions using data from India. Moreover, gender differences in child survival exhibit similar patterns, especially in households without piped water where the health benefits of breastfeeding should be greatest.

---

\*We thank Kristin Plys and Luke Stein for excellent research assistance and Anne Case, Monica Das Gupta, Angus Deaton, Esther Duflo, Pascaline Dupas, Fred Finan, Claudia Goldin, Jeffrey Hammer, Larry Katz, Alan Krueger, Adriana Lleras-Muney, Ben Olken, Emily Oster, Rohini Pande, Eliot Regett, Shanna Rose, Sam Schulhofer-Wohl, Chris Udry, Chris Woodruff, several conference and seminar participants, and the referees for helpful comments. Contact information: jayachan@stanford.edu; kuziemko@princeton.edu

# 1 Introduction

Medical and public health research has established that breastfeeding inhibits post-natal fertility.<sup>1</sup> In addition, the physical demands of another pregnancy often cause a mother to stop nursing her current child.<sup>2</sup> These two biological constraints induce a negative relationship between breastfeeding duration and future fertility, especially in developing countries with limited access to artificial contraception.

In many developing countries such as India and China, son preference is common and results in a stop-after-a-son fertility pattern: When a daughter is born, parents are more likely to want to have more children to “try again” for a son (Das, 1987).

The central insight of this paper comes from relating these two results: If son preference exists, girls will be weaned earlier than boys since parents want to try again for a son. Given the large health benefits of breastfeeding in environments with contaminated food and water (Feachem and Koblinsky, 1984; Palloni and Millman, 1986; Habicht, DaVanzo, and Butz, 1988), such behavior can lead to gender disparities in child health. In contrast to the standard explanation of gender health disparities—that parents actively invest more in sons’ health—we show that if parents merely prefer to have sons, then mothers will breastfeed daughters less, causing a gender gap in health even when parents equally value the health of all their children. Our results highlight how health disparities can arise “passively” from fertility preferences, and relate to earlier work (Yamaguchi, 1989; Clark, 2000; Jensen, 2003) showing that a “try until you have a son” fertility rule results in girls, on average, having more siblings and thus more competition for household resources.

The paper develops a dynamic programming model of breastfeeding that incorporates its contraceptive properties and yields several predictions regarding when mothers wean their children, including the above gender result. First, breastfeeding duration increases with birth order, as demand for breastfeeding’s contraceptive properties should rise as couples approach or exceed their ideal family size, and in fact should increase discontinuously once ideal family size is met. Second, as explained above, as long as some son preference exists, girls are breastfed less than boys. Third, by the same logic, children with older brothers are breastfed more. Fourth, the gender effect is largest when a couple approaches its target family size

---

<sup>1</sup>Research suggests that nursing suppresses the Gonadotropin-releasing hormone that regulates ovulation. Nursing can also cause weight loss, which can disrupt ovulation (Blackburn, 2007). See the next section for further discussion.

<sup>2</sup>In our data, many women cite pregnancy as the reason they stopped nursing. As discussed in the next section, there are many reasons for this effect, e.g., pregnancy can cause a decrease in breast-milk production (Feldman, 2000).

since, at that point, their decision to have another child is highly marginal and thus most sensitive to sex composition; the effect is smallest for children whose birth order is either significantly below or above ideal family size as couples will want to, respectively, wean or nurse such children regardless of gender.

We confirm each prediction using data from the 1992, 1998 and 2005 waves of the National Family Health Survey in India. Moreover, we find similar gender and birth-order patterns for child survival as we do for breastfeeding, especially in households without piped water where the health benefits of nursing are likely to be most important. These results suggest that in India, the decision about when to wean a child depends heavily on whether or not the mother wants to become pregnant again, and moreover, that this decision rule disadvantages daughters as well as other children born to mothers whose demand for children is unmet.

These results have several implications for researchers and policymakers. First, they offer a new mechanism for gender disparities in child health and thus may partly explain the “missing girls” problem (Sen, 1990). Back-of-the-envelope calculations suggest that differences in breastfeeding could account for 8,000 to 21,000 missing girls each year in India, explaining roughly nine percent of the gender gap in child mortality (deaths between ages one and five).

Second, breastfeeding’s contraceptive effects coupled with its health benefits to children living in unsanitary environments suggest a negative relationship between total fertility and child health in developing countries. A mother who wants 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.

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.

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 prediction 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, 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.

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 fertility for 12 to 24 months in Bangladesh and Nepal, respectively. The longer duration of postpartum infertility in developing countries is likely related to mothers’ underlying malnutrition (Huffman et al., 1987).

For a mother to use breastfeeding as birth control, she first must know about breastfeeding’s contraceptive effects. 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 current 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”), but research finds that infant weight gain is slowed if a mother is simultaneously nursing an older sibling (Marquis et al., 2002). There is also some evidence that breast milk production declines during pregnancy and the taste of the milk changes, both of which might hasten weaning (Feldman, 2000). In developing countries, meeting the caloric requirements of tandem breastfeeding or breastfeeding while pregnant is likely much more challenging than it would be for the mothers in the above studies.

Of course, the relevant question for this paper is whether mothers choose to stop breastfeeding after a conception or birth, for whatever reason, e.g., the opportunity cost of her time has risen. 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).

## **2.3 Connecting fertility preferences, breastfeeding and child health**

This paper examines the causal chain from desired fertility, and in particular son preference, to breastfeeding and, eventually, to child health and survival. While to our knowledge we are the first to systematically look at this question either theoretically or empirically, previous work has looked at parts of this causal chain.

First, several papers show empirically that in settings with son preference, a couple that just had a son is more likely to stop having children (Das, 1987) or wait longer to have the next child (Rindfuss et al. 1982; Trussell et al., 1985; Arnold, Choe, and Roy, 1998; Retherford and Roy, 2003). Hemochandra, Singh, and Singh (2010) show that birth intervals are shorter if parents have not reached their self-reported ideal number of sons. Second, Nemeth and Bowling (1985) examine whether having no sons shortens the duration of breastfeeding.

Third, a few papers on related topics discuss the mechanism that we focus on (or part of it), though they do not test for it. Muhuri and Menken (1997) show that birth spacing and

the sex of older siblings affect mortality and state that shorter birth intervals after girls are one potential cause of excess female mortality. In unreported results, Arnold, Choe, and Roy (1998) examine whether the gender gap in mortality in India is due to birth spacing; they do not find evidence to support the hypothesis. Finally, Mutharayappa et al. (1997) show that there is a gender gap in breastfeeding in India and conjecture that it is related to son-biased fertility patterns.

We hope to advance this body of work by deriving a number of quite specific testable predictions of the fertility-breastfeeding hypothesis, and then using them to empirically separate the causal chain we propose from alternative hypotheses.

### 3 Model

In the appendix we present a simple dynamic model of a mother’s decision to breastfeed and her future fertility. The model predicts that breastfeeding duration will have several distinctive patterns with respect to children’s birth order, gender, and the gender composition of older siblings. Here we give an overview of the model, focusing on the empirical predictions that we will test in later sections.

#### 3.1 Assumptions of the model

The model relies on two main assumptions. First, breastfeeding inhibits fecundity. In our model, the breastfeeding decision is, in essence, a fertility stopping decision, and the predictions about breastfeeding we generate are fundamentally predictions about stopping rules.<sup>3</sup> Second, utility depends on both the quantity and the sex composition of one’s children. Utility is increasing and then decreasing in quantity; mothers want some target number of children and are worse off if they have more or fewer than this number, all else equal. In addition, there is son preference: Conditional on the number of children, utility is increasing in the number of sons at a diminishing rate. These preferences mean that at intermediate birth order, around the target family size, further childbearing entails a tradeoff between disutility from having another child and positive utility from potentially having another son.

---

<sup>3</sup>The model, for simplicity, treats breastfeeding as perfectly inhibiting conception. Allowing mothers to conceive while breastfeeding would reinforce the model’s predictions. Mothers who become pregnant while nursing 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. Another simplification is that we do not model breastfeeding being used to space births. The model’s predictions would be the same if we allowed mothers who want to continue childbearing to breastfeed in order to space births but assumed they still breastfeed less than mothers who want to stop having children entirely.

### 3.2 Predictions of the model

The model makes several testable predictions. The fact that a mother’s desired future fertility declines as she has more children gives us the following first result.

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

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

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

*(i) the child is male; or*

*(ii) more of his or her older siblings are male.*

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.*

The intuition behind this result is that at low birth order, mothers want to continue having children regardless of the sex composition of their existing children since utility is still increasing in the quantity of children. At high enough birth order, the net cost of increasing family size 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. At intermediate birth order, however, a mother who just had a daughter may find that the benefit of trying for a son is larger than the cost of higher quantity; in contrast, for a mother who just had a son, that benefit is smaller and, thus, less likely to outweigh the cost.<sup>4</sup>

The model also predicts that breastfeeding patterns change non-linearly when a mother reaches her “ideal” quantity of children. Ideal family size is an ambiguous concept, but we use

---

<sup>4</sup>While several papers have theoretically explored how fertility stopping rules depend on sex composition, e.g., Leung (1991), to our knowledge, none has identified or tested this non-monotonic pattern.

the following definition: the family size at which a mother would want to stop having children if she had no son preference. Recall that the marginal utility from having more children goes from positive to negative, so there is a well-defined ideal family size under this definition. The following proposition describes how both breastfeeding and the gender gap in breastfeeding depend on whether the mother’s number of children is below or above her 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.*

To summarize, the negative relationship between breastfeeding and future fertility generates several distinct empirical predictions. Other theories might predict 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.<sup>5</sup> 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 the large sample of children for which mothers report breastfeeding duration. Additionally, because basic demographic information is recorded for every child born to a survey mother, we can calculate birth order and

---

<sup>5</sup>Alternative hypotheses might predict an interaction of gender and birth order. For example, the mother might have less time to breastfeed as she has more children and, on the margin, allocate more time to breastfeed sons. However, such hypotheses generally do not naturally predict the non-monotonicity or the non-linearity when ideal family size is reached (and, moreover, this particular alternative story is at odds with breastfeeding increasing with birth order.)



the sex composition of siblings for each child. The survey also includes a variety of information on desired fertility, contraception, and child health, as well as standard demographic and household characteristics.

We make several sampling restrictions. First, 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 should have been solicited, the data are missing.

Second, we exclude mothers who have eight or more children (the 95<sup>th</sup> percentile for this variable) to reduce composition bias from mothers with unusually large family size. Third, 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 a manner that does not reflect mothers' preferences regarding breastfeeding; this restriction results in a loss of about five percent of remaining observations.<sup>6</sup> Our final sample includes just over 110,000 observations.

Summary statistics for the sample used in the breastfeeding analysis are presented in Table 1. Note that the breastfeeding durations reported in the table do not account for censoring of children still being breastfed at the time of the survey. The average observed breastfeeding duration in our sample—which includes both children still being breastfed and children already weaned—is 14.8 months, whereas adjusting for censoring via a hazard estimation indicates that children in the sample will be breastfed for 22.8 months on average.

Similarly, the mean of birth order in our sample is 2.6, but many women we observe have not completed their fertility. We cannot calculate total fertility in our data, but based on Indian 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

---

<sup>6</sup>We estimate both OLS and hazard models, as described in the next section. In the hazard models, one can include children who have died and model the observed breastfeeding duration as censored at the child's age at death. In Appendix Table 1, we show that our hazard model specifications are robust to including the children who have died. Our results are also robust to including mothers with more than eight children or multiple births (or both); results available upon request.

data.<sup>7</sup>

The descriptive statistics in Table 1 compare low versus high birth-order children and also boys versus girls. High birth-order children are born into larger families, on average. Thus, not surprisingly, characteristics associated with family size such as rurality and mother’s education differ by birth order. In contrast, the family background variables are very similar for boys and girls in the sample. Accordingly, in our empirical analysis, specifications with only birth order will be sensitive to the inclusion of covariates, but specifications focusing on gender or its interaction with birth order will not be.

## 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 cease childbearing, 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  $i$ , who may or may not still be breastfeeding at the time of the survey;  $a_i$  is a vector of age-in-months fixed effects up to 36 months (the maximum value of the outcome);  $X$  is a vector of covariates and  $\varepsilon$  is an error term. We add age-in-month fixed effects to address the fact that the duration of breastfeeding is censored; if a nine-month-old child is still being nursed at the time of the survey, his duration of breastfeeding is censored at nine months. Note that the regression still estimates *observed* breastfeeding duration (mean = 14.8) and not *completed* breastfeeding duration (mean = 22.8), but by only making comparisons among children who are the same age, it does not mistakenly include the effect of age when estimating the effect of birth order on observed breastfeeding duration.

While the OLS model has the advantage that the coefficients are easy to interpret, we also estimate a proportional hazard model, which accounts for the censoring of completed breastfeeding duration due to the fact that many children in our sample are still being nursed,

---

<sup>7</sup>In the 1991 Census, mothers age 50-54 had a mean total fertility of 4.8. We assume fertility is falling at one percent a year. Disaggregated fertility data for the 2001 Census are not yet publicly available.

imposes no conditions on the baseline hazard function  $h_0(t)$ , and models independent variables as having a proportional effect on the hazard rate:

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

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: Breastfeeding duration increases monotonically with birth order.

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, column (2) includes what we will generally use as our “standard” set of covariates: 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. Column (3) is the hazard estimation analogue to column (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.

The sensitivity of the birth-order coefficient to including mother characteristics suggests that it might be useful to estimate a mother fixed-effects model, in which the identification

comes from differences across siblings; however, sample selection bias would plague this estimation. Recall that the survey only records breastfeeding duration for a look-back period of three to five years. Using the three-year look-back period as an example, a youngest child born, say, 24 months before the survey date must have a sibling no more than 12 months older than him in order to be in the sample. But, according to our hypothesis, selecting for sibling pairs in which the older child is followed quickly by a subsequent birth is equivalent to selecting for sibling pairs in which the older child was weaned early. Thus, there would be a positive correlation between birth order and breastfeeding duration in the mother-fixed-effect sample even in the absence of any such relationship in the general population.<sup>8</sup>

Thus, while the results on birth order are consistent with the hypothesis that future fertility determines breastfeeding, concerns about omitted variables remain. In addition, even compelling evidence of a causal effect of birth order on breastfeeding would not disprove alternative hypotheses such as 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 much 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.<sup>9</sup> From this measure we generate a variable that measures the current distance from the mother’s ideal family size:  $\Delta Ideal_{ij} = BirthOrder_{ij} - Ideal_j$  (where  $BirthOrder_{ij}$  is the birth order of mother  $j$ ’s  $i$ th child and  $Ideal_j$  is mother  $j$ ’s ideal family size).

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

---

<sup>8</sup>Indeed, mother fixed-effect estimations yield a positive coefficient on birth order that is far larger than those in Table 2, but the magnitude cannot be interpreted as causal given the sample selection problem.

<sup>9</sup>The survey question is, “If you could go back to the time you did not have any children and could choose exactly the number of children to have your whole life, how many would that be?”

knew she would have two girls first, then her ideal family size might be larger.

Second, to avoid cognitive dissonance, mothers might self-report their “ideal” fertility preference to match their actual fertility outcome, and thus the variable might not reflect their preferences at the time they gave birth and made breastfeeding choices. 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.,  $\Delta Ideal > 0$ ).<sup>10</sup>

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 shows a jump in the level of breastfeeding once ideal family size is reached ( $\Delta Ideal = 0$ ), as predicted by the model.<sup>11</sup>

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. Column (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 our standard set of covariates in column (5) yields a coefficient of 0.88 (an 18 percent drop).

The more exacting specification, though, is to test for a discrete increase in breastfeeding at  $\Delta Ideal = 0$  while allowing for a linear effect of  $\Delta Ideal$  that is allowed to vary on either side of  $\Delta Ideal = 0$ .<sup>12</sup> We estimate the following equation, where again the variable of interest is the indicator for  $\Delta Ideal \geq 0$ :

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

---

<sup>10</sup>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.

<sup>11</sup>To explore more rigorously the discrete jump in breastfeeding when a mother reaches her desired number of children, we create a series of dummy variables corresponding to the conditions  $\Delta Ideal \geq x$  and then run separate regressions of breastfeeding on each dummy variable (as well the standard covariates). Indeed, the regression using the variable  $\Delta Ideal \geq 0$  yields the largest  $t$ -statistic and  $R^2$ . Results available upon request.

<sup>12</sup>In the online appendix, we show that the results are similar if we omit the spline in  $\Delta Ideal$ , i.e., the regressor  $\Delta Ideal \times \mathbb{1}(\Delta Ideal \geq 0)$ .

Column (6) shows the results without our standard set of covariates included. Once mothers reach their ideal fertility, the duration they breastfeed subsequent children increases by 0.77 months. With covariates included, we find that above and beyond a flexible linear effect of  $\Delta Ideal$ , breastfeeding duration increases by 0.40 months once a mother reaches her ideal family size.

## 5.4 Discussion

There appears to be a strong positive correlation between breastfeeding and birth order, consistent with the model’s 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 0.4 months longer than older siblings.

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 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. Demand for an additional child is higher after the birth of a girl and thus mothers wean daughters sooner in the hopes

of conceiving a son (*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, 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 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)*). These variables 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 heaping of observations at multiples of six months (which likely reflects both rounding error in reporting and an actual propensity to breastfeed up to focal point) makes it difficult to see the differences across gender at those points, but elsewhere the gender differences are clearly visible. There is no apparent gender gap for the first few months of life, but then a gender gap opens up and persists through age 36 months. The gender gap is statistically significant at the one-percent level at age seven months, and remains so through age 36 months (see the online appendix).

The mean duration of completed breastfeeding is 22.33 months for girls and 23.26 months for boys.<sup>13</sup> Note that this gender gap of 0.93 months in completed breastfeeding is over twice

---

<sup>13</sup>This mean is calculated assuming that no child is breastfed beyond the top-coded value of 36 months, but there is in fact still a gender gap at 36 months. If the survival curve is exponentially extended to zero for

as large as the gender gap in observed breastfeeding reported in Table 1. This difference is due to the fact that almost all children under age six months at the time of the survey will exhibit no current gender gap, even though many of the boys in that group will eventually be breastfed longer than the girls.

The gender gap can also be quantified in terms of the percentage point gap in the likelihood of being breastfed. Over the age range of 12 to 36 months, where the gender gap is most pronounced, girls on average are 3.7 percentage points less likely to be breastfed in any given month. This effect size is consistent with, for example, 3.7 percent of girls being weaned at age 12 months who, had they been boys, would have been weaned at age 36 months or later, and all other girls experiencing no differential treatment during this period.

We next use regression analysis to estimate the gender gap in mean breastfeeding duration between boys and girls. Table 3 shows the effect of the child’s sex and of the sex composition of existing children. The first column indicates that sons receive an additional 0.37 months of breastfeeding relative to daughters. This point estimate refers to the difference in *observed* breastfeeding (as reported in Table 1) and not to the difference in expected *completed* breastfeeding (as shown in Figure 3). As OLS estimates are much easier to interpret, we will mainly focus on the differences in observed breastfeeding, though one should keep in mind that differences in completed breastfeeding are generally more relevant for child health, which we examine in Section 7.

Column (2) shows that the *Male* effect size barely moves (increasing to 0.39 months) after adding our standard set of covariates from Table 2 plus birth-order fixed effects. Column (3), the hazard-model analogue of column (2), suggests that sons have a ten percent lower probability of being weaned in any given month relative to daughters.

In the next two columns we examine whether the sex composition of siblings has an independent effect on breastfeeding even after accounting for the sex of the current child. Column (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 higher. 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

---

observations censored at 36 months, the male and female means are, respectively, 23.94 and 22.88, yielding a gender gap of 1.06 instead of 0.93 months.



“take to the breast” better than daughters.

Finally, the last column of Table 3 re-estimates the specification in column (2), allowing the male coefficient to vary by survey wave. The male-wave interactions are imprecisely estimated, but the point estimates suggest that the gender gap in breastfeeding was 0.31 months in the first wave, 0.39 months in the second wave, and 0.46 months in the most recent wave.

### 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. The prediction is 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).<sup>14</sup> 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.

We first estimate the following equation:

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

The key variables are the vector of  $Male \times BirthOrder$  dummy variables, which allow 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 in  $X$ . Sons are breastfed more than girls at every birth order, but the difference is not constant across birth order. The gap is increasing until birth order four, and then decreasing after that, in line with the model’s prediction. 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

---

<sup>14</sup>Recall the two mechanisms underlying the breastfeeding-fertility relationship. If the relationship depended entirely on women using breastfeeding as contraception, then the population-average effect would peak at the average 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.

columns of Table 4. Column (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  $i$ th child and  $Ideal_j$  is mother  $j$ 's ideal family size). Except for  $\Delta Ideal = -3$  (and recall from Figure 2 that almost all the density of the  $\Delta Ideal$  distribution falls above this point), 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. We read this evidence 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 is a discrete jump in the male breastfeeding advantage just when mothers reach their ideal family size. The effect size is 0.55 months without covariates and is relatively unchanged (increases to 0.59) when covariates are added to the regression.<sup>15</sup>

#### 6.4 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 the average degree of son preference in the state, and indeed we find that it is systematically lower in south India, where son preference is recognized to be less intense (Bhatia, 1978; Basu, 1992).

Table 5, column (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

---

<sup>15</sup>As discussed in Section 3, our breastfeeding predictions are essentially fertility-stopping predictions. In Appendix Table 2, we verify that when an indicator variable for having no younger siblings is used in place of breastfeeding duration as the dependent variable, we find the same coefficient patterns with respect to birth order, distance from ideal family size, gender and their interactions. We also show in the online appendix that we find the same patterns for mothers' self-reported desire to not have any more children.

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. Note that this result is consistent with both a stop-after-a-boy fertility preference and with parents caring more about sons' health.

We next examine how much the breastfeeding gap increases when the mother reaches her ideal family size, which is a more distinctive prediction of son-biased stopping rules. As seen in column (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 (inclusive of the child himself) minus her ideal number of sons. As seen in column (3), the duration of breastfeeding increases by 0.41 months once the mother reaches her ideal number of sons. Parallel to our earlier  $\Delta Ideal$  specifications, the specification allows for a linear effect of  $IdealSons$  that varies on either side of  $\Delta IdealSons = 0$ . This increase at  $\Delta IdealSons = 0$  is above and beyond the increase at  $\Delta Ideal = 0$ ; when a mother has reached *both* her ideal number of sons and her ideal family size, her children are breastfed 0.74 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 column (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. 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.

## 6.5 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 predictions of our model in which mothers decide how long 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, column 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 arise until mothers reach their ideal family size. This estimate is provided by the coefficient on  $Male \times \Delta Ideal \geq 0$  in column (5) of Table 4, which is 0.59. As the mean of  $\Delta Ideal \geq 0$  is 0.55, we estimate that our mechanism explains 0.32 months ( $0.59 \times 0.55$ ) or 82 percent ( $0.32/0.39$ ) of the male breastfeeding advantage. Another approach is to quantify 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 0.15, 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 6 shows the results for our main specifications when “total number of vaccinations child received” serves as the dependent variable. While there is indeed a strong male advantage, none of the other patterns found for breastfeeding hold. For example, column (1) shows vaccinations are in fact decreasing with birth order. Column (3) shows that the sex of older siblings does not affect vaccinations, and column (5) shows that the male advantage does not increase after the mother reaches her ideal family size. We find similar results, which are reported in the online appendix, using whether the child received any vaccine, received medical treatment conditional on having a cough or fever, or received

medical treatment conditional on having diarrhea as the dependent variable.<sup>16</sup>

## 6.6 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 is the average boy breastfed longer than the average girl (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 show that the gender patterns in breastfeeding depend on the mother's degree of son preference in the way that theory would predict. 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 child health in India. We first discuss the medical evidence on the health benefits of breastfeeding. We then review the patterns for child mortality across gender and birth order documented by past researchers to determine whether they appear plausibly related to the patterns we find for breastfeeding.

---

<sup>16</sup>We also assessed whether sex-selective abortion could somehow be driving our results. Because no question on the topic is included in the NFHS, we instead examined how consistent our coefficient estimates are across the 1992, 1998 and 2005 waves of our data, which cover a period when sex-selective abortion became increasingly common in India (Arnold, Kishor, and Roy, 2002; Jha et al., 2006). As shown in the online appendix, our results are quite stable across wave, suggesting that our findings are not an artifact of sex-selective abortion; if they were, one would expect muted effects in the first wave. Speculatively, the gender gap in breastfeeding, just like other types of gender discrimination, could become less pronounced when sex-selective abortion becomes more widely used because girls are then born into families with less son preference. However, our by-wave analysis shows no evidence of this hypothesized trend, at least between 1992 and 2005.

Finally, we use our NFHS sample to directly test whether child survival exhibits the same relationships with gender, birth order and ideal family size as does breastfeeding duration.

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

Medical and public health researchers have suggested several mechanisms through 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). The mother’s digestive tract inactivates pathogens, so even if she ingests contaminated food and water, her breast milk is not contaminated (Isaacs, 2005).

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 in south Asia that examine how breastfeeding beyond infancy affects mortality. Even past the age of exclusive breastfeeding, partially breastfed children consume less contaminated food and water, which lowers their risk of diarrheal disease and other water- and food-borne diseases (Prentice, 1991). 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.<sup>17</sup>

---

<sup>17</sup>HIV can be transmitted from a mother to her child through breastfeeding. If antiretrovirals are unavailable, HIV-positive mothers are encouraged to wean their child after six months, as the risk of HIV transmission could outweigh the potential benefits of prolonged breastfeeding (World Health Organization, 2010). In addition, if the food a child would consume in lieu of breast milk if he or she were weaned is more nutritious than breast milk, then weaning could be beneficial to a child. The literature suggests that breastfeeding could either increase or decrease a child’s calorie consumption and nutrient intake (Jelliffe and Jelliffe, 1978; Motarjemi et al., 1993). The positive association between prolonged breastfeeding and child survival suggests that, on average, prolonged breastfeeding is beneficial to child health in the context we study.

## **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, Behrman (1988) uses food consumption data to show that parents favor lower birth-order children in India. Our own results from Table 6 suggest that parents vaccinate lower birth-order children significantly more than later children. Finally, the mechanical positive correlation between birth order and family size may impart a health advantage to low-birth-order children. Thus, the favorable circumstances of low-birth-order 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 one year 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). In line with this fact, we find relatively little gender difference in breastfeeding for first-born children (Figure 4).

## **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 child survival.

Figure 3 showed the age range where breastfeeding differences are largest and thus any related mortality differences should be largest.<sup>18</sup> The gender gap in breastfeeding is concentrated between the ages of 12 and 36 months. As discussed in Section 6 and shown in detail in the online appendix, there is no statistically significant gender gap during the first six 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 mortality between one and six months as a placebo test.<sup>19</sup>

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, which helps shed light on the mechanism behind any health effects. The more important reason for this comparison is that it helps us to separate the effects of breastfeeding from other factors associated with future fertility. For example, higher future fertility implies a larger family size and more competition for the mother’s time and energy, which could be detrimental to child survival (Palloni and Tienda, 1986; Yamaguchi, 1989; Clark, 2000; Jensen, 2003). However, these other channels are not predicted to vary with water quality.<sup>20</sup>

## Mortality data from the NFHS

Because 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.<sup>21</sup> 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 (Byass et al., 2007).

Although we do not claim to eliminate recall bias, we try to hold it roughly constant

---

<sup>18</sup>There may also be long-term effects on the child’s health which lead to mortality at later ages; in essence we test only for contemporaneous effects.

<sup>19</sup>We use one to six months rather than zero to six months since mortality in the first month of life is typically due to genetic defects rather than health inputs; to wit, neonatal mortality for boys is higher than for girls in our sample. The online appendix shows that we also do not find the breastfeeding-related patterns when mortality between zero and six months is the outcome.

<sup>20</sup>The comparison between households with and without piped water also helps isolate the biological effect of breastfeeding from other potential effects of breastfeeding. For example, a mother who is nursing might supply less labor, leading to an income effect, but such an effect should exist in both types of households.

<sup>21</sup>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.



in all the regressions we run. Specifically, when we examine mortality rates between 1 and 6 months, we include children such that  $\text{survey date} - \text{birth date} > 6$  months and  $\text{survey date} - \text{birth date} \leq 66$  months. In this case, all observations would have been fully “at risk” for the mortality window and the recall period is five years ( $66 - 6 = 60$  months). Similarly, when we examine mortality between 12 and 36 months, we use observations satisfying  $\text{survey date} - \text{birth date} > 36$  months and  $\text{survey date} - \text{birth date} \leq 96$  months (as well as conditioning on being alive at twelve months; otherwise the outcome is not defined). This approach also ensures that we have similar sample sizes for both the 12-to-36-month mortality rate and the 1-to-6-month mortality rate that we use as a placebo test.<sup>22</sup>

The sample for the 12-to-36-month-mortality regressions has about 161,000 observations and a 12-to-36-month mortality rate of 0.0205. 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 1-to-6-month-mortality regressions is roughly the same size, with a 1-to-6-month mortality rate of 0.0153.

## 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 low-birth-order children documented in past work and in our Table 6 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-birth-order children.

The remainder of this section examines the effects on child mortality of gender and its interactions with birth order and ideal family size. Table 7 shows the results when mortality between the ages of 12 and 36 months serves as the dependent variable, with the first three columns focusing on households without piped water. Column (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 breastfed more than girls. The point estimates in the second column is suggestive that the male survival advantage, like the breastfeeding advantage estimated in the previous section, has an inverted- $u$  shape with respect to birth

---

<sup>22</sup>The results are robust to using the same sample for the two analyses, i.e., using the 12-to-36-month-mortality sample for the 1-to-6-month-mortality analysis. Note that we cannot use the 1-to-6-month-mortality sample for the 12-to-36-month mortality regressions since 12-to-36-month mortality is undefined for those who die before age 12 months.

order, although the quadratic term is not significant and the implied peak in the mortality gender gap is at birth order 6.5, which is higher than we found for breastfeeding. Column (3) is suggestive that once mothers reach their ideal number of children, the male advantage with respect to child mortality increases discontinuously.

To help distinguish breastfeeding’s effect from the effects of confounding variables, we estimate the same regressions as in cols. (1) to (3) 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. (4) through (6), the coefficients of interest are all smaller in magnitude or wrong-signed.

We formally test for the equality of the coefficients by estimating three fully-interacted models (i.e., every variable, not merely variables of interest, is interacted with *Unpipied*) that are equivalent to jointly estimating columns 1 and 4, columns 2 and 5, and columns 3 and 6. 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 specification, which tests 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 a  $p$ -value of 0.06. Thus, relative to other households, in households where breastfeeding should matter more for child health, the mortality patterns correspond more closely to the breastfeeding patterns we document earlier in the paper.

As a placebo test, we use mortality for young infants as the outcome. Since there is no gender differential in breastfeeding in the first few months of life, finding the same patterns as we do in Table 7 would suggest that something correlated with breastfeeding duration, and not breastfeeding duration itself, is driving those results. Appendix Table 3 is the exact analogue to Table 7 with the exception that mortality between ages 1 and 6 months serves as the dependent variable. The male coefficient in column (1) is statistically insignificant and an order of magnitude smaller than we found for 12-to-36-month mortality. The quadratic in birth order for households without piped water is statistically insignificant (though this was also the case for 12-to-36-month mortality). In column (3), the point estimate suggests a male survival *disadvantage* emerging in households without piped water once mothers reach their ideal family size. Comparing the households without piped water to those with piped

water, the basic male advantage in mortality differs based on the water source with a  $p$ -value of 0.06, but unlike in our 12-to-36-month mortality results, one cannot reject that the coefficients of interest are the same in households with and without piped water for the next two specifications.<sup>23</sup>

Of course, whether a household has piped water could be correlated with other factors that affect gender-specific mortality patterns. We thus also estimate regressions in which every variable is also interacted with an indicator variable for whether the household is in a rural area. The first three columns in Appendix Table 4 show the differences between households with and without piped water for 12-to-36 month mortality. The magnitude of the coefficients falls in some cases, suggesting some role for omitted variables bias. For example, the difference in the male coefficient between unpiped and piped households was -0.00465 without the rural interactions (see Table 7, column 1) and is -0.00328 with the rural interactions. However, even with the rural interactions, it remains statistically significant with a  $p$ -value of 0.03. The final three columns of Appendix Table 3 show that for mortality between age one and six months (the main placebo test), the gender differences between households with and without piped water are again not significant and in some cases wrong-signed.

In summary, where breastfeeding should play an especially important role in child health (households without piped water) and where breastfeeding gender gaps are most pronounced (between the ages of one and three years), the variation in mortality with respect to gender, birth order and  $\Delta Ideal$  is similar to the variation in breastfeeding with respect to these same variables. In settings where breastfeeding should play less of a role in child survival or where there is no breastfeeding gender gap, these mortality patterns are weaker or do not emerge at all.

## 7.4 Breastfeeding and “missing girls”

Using our results from Section 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

---

<sup>23</sup>This robustness check is somewhat conservative because, while the breastfeeding gender gap is not statistically significant before seven months, the point estimate is non-trivial as early as five months. Given the large health benefits of breastfeeding in the first six months of life, one might thus expect to see some mortality effects for this placebo group. We therefore have also run placebo tests using mortality between the ages of one and three months (where there is no evidence of a breastfeeding gender gap), as well as mortality between 37 and 60 months (where there is little evidence of either a gender gap in breastfeeding or large health benefits of breastfeeding). As shown in the online appendix, for these outcomes we also do not find any of the distinctive patterns associated with 12-to-36-month mortality. These alternative age ranges are arguably more appropriate placebo tests. However, because the mortality rate is somewhat lower for these ages, there is also less statistical power, so we use one-to-six-month mortality as our main placebo test.

much breastfeeding contributes to the gender gap in child mortality in India.

Our first calculation uses the gender gap in breastfeeding that we estimated, combined with estimates from the medical literature on the effect of breastfeeding on mortality. Based on the literature, we assume that lack of breastfeeding between the ages of 12 and 36 months increases the probability the child dies during this age interval by a factor of 2.5.<sup>24</sup> About 48.1 percent of the children aged 12 to 36 months in our sample are being breastfed, as seen in Figure 3, and about 2.05 percent die during this age interval. Thus, the implied mortality rate for breastfed children solves the equation  $0.481x + 2.5(1 - 0.481)x = 2.05$ . This calculation yields a mortality rate of 1.15 percent for breastfed children and 2.88 percent ( $1.15 * 2.5 = 2.88$ ) for non-breastfed children in this age interval. In other words, not being breastfed throughout the age range of 12 to 36 months increases the likelihood of dying by 1.73 percentage points ( $2.88 - 1.15 = 1.73$ ).

The breastfeeding survival curves (Figure 3) show that, on average, the gender gap in breastfeeding between the ages of 12 and 36 months is 3.67 percentage points. Therefore, if breastfeeding differences were the only source of mortality variation, the mortality gap would equal 0.064 percentage points ( $0.0367 * 1.73 = 0.064$ ). Given that each year about 13.2 million girls survive to age one in India, this mortality gap corresponds to about 8,400 ( $13,200,000 * 0.00064$ ) excess female deaths per year. The excess female mortality between ages 12 and 36 months in our sample is 0.75 percentage points (2.44 percent for girls versus 1.69 for boys), so breastfeeding differences explain 8.5 percent ( $0.064/0.75 = 0.085$ ) of this mortality gender gap.

Note that this calculation could underestimate or overestimate the excess female mortality due to breastfeeding. On the one hand, the calculation does not count the deaths caused by the breastfeeding gap between age 7 months and 12 months, and at these younger ages, breastfeeding likely confers an even larger survival advantage. On the other hand, the estimates from the literature on the mortality effects of breastfeeding are correlations and may overestimate the causal effect of breastfeeding on child survival.<sup>25</sup>

---

<sup>24</sup>The factor of 2.5 is the midpoint of the estimates from two studies that examine the relevant age range in developing countries. First, breastfeeding between the ages of 18 and 36 months was found to reduce mortality risk by two thirds in Bangladesh (implying a factor of 3) (Briend, Wojtyniak, and Rowland, 1988). Second, breastfeeding between ages one and two was found to reduce mortality risk in developing countries by one half (implying a factor of 2) (World Health Organization, 2000).

<sup>25</sup>For example, if the absence of breastfeeding increases mortality risk by a factor of 2.5 between age 12 to 24 months, but the literature is incorrect and there is no health benefit of breastfeeding after age 24 months, then the number of excess deaths would be 30 percent smaller: The average gender gap in breastfeeding between age 12 and 24 months is .0336 (91 percent of the average 12-to-36 month gap), and mortality between age 12 and 24 months is 1.58 percentage points (76 percent of mortality between age 12 and 36 months).

Our second calculation uses our mortality results from the previous subsection. To be conservative, we assume, first, that breastfeeding only affects survival in households without piped water. Then the relevant mortality gap is the difference-in-difference coefficient of -0.00465 from Table 7, column (1). Second, we subtract the corresponding coefficient for our placebo group, which is -0.00257 (Appendix Table 3, column 1). Differencing out the piped-unpiped gender gap in mortality for the placebo group helps address the concern that other omitted variables are correlated with whether the household has piped water; our hypothesis is that the actual effect of piped water on excess female mortality should only emerge after age 6 months. Under these two assumptions, the breastfeeding gap accounts for 0.208 percentage points ( $0.465 - 0.257 = 0.208$ ) of excess female mortality in households without piped water. As 78 percent of the sample lacks piped water, the breastfeeding gender gap appears to explain  $0.208 * 0.78 = 0.163$  percentage points (22 percent) of excess female mortality between the ages of one and three, or about  $0.00163 * 13,200,000 = 21,500$  missing girls per year.

In summary, we find that the gender gap in breastfeeding accounts for somewhere between 8,400 and 21,500 excess female deaths each year. Using the midpoint of this range—15,000 missing girls—the breastfeeding differential explains 15 percent of the excess female mortality between the ages of one and three in India, or 9 percent of excess female child mortality, defined as deaths between ages one and five.<sup>26</sup>

These estimates have important caveats that are worth reiterating. The differential gender gap in mortality in households with unclean water may not be fully due to breastfeeding. In addition, our estimate of 15,000 missing girls is the middle of a range, and the endpoints of our range themselves have wide confidence intervals because of sampling error and the assumptions that the calculations require.

## 8 Conclusion

This paper began by noting 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 model of fertility decisions that incorporates this negative covariance between breastfeeding duration and subsequent conception. The model makes a number of

---

<sup>26</sup>The gender gap in child mortality is 1.2 percentage points (Pandey et al., 1998). The midpoint of our range (15,000 deaths) is equivalent to 0.113 percentage points of excess female mortality, and  $0.113/1.2 = 9.4$  percent.

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 continue breastfeeding 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 girls: 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. Breastfeeding duration increases with birth order, and this effect is largely driven by a discrete increase once mothers reach their ideal family size. Sons are breastfed 0.9 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 not the main explanation for these results. Instead, about two thirds of the son advantage in breastfeeding is due to the value parents place on having future sons, according to our estimates. 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, appear to 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 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 six months, when no gender gap in breastfeeding exists.

We calculate that the gender gap in breastfeeding accounts for roughly nine percent of excess female child mortality (deaths between the ages of one and five years) in India. In other words, mothers breastfeeding daughters less than sons leads to about 8,000 to 21,000 “missing girls” each year. Unlike many other proposed factors causing missing girls, the channel we highlight does not require that parents value girls’ health less than boys’. Instead, excess female mortality arises as an unintended consequence of parents’ desire to have more sons rather than from an explicit decision to allocate fewer resources to daughters.

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.<sup>27</sup>

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.

---

<sup>27</sup>We know of no strategy to isolate exogenous variation in access to contraception for this setting. As a suggestive test, we constructed an imperfect measure of “access,” the average use of contraception in the respondent’s primary sampling unit (village), excluding herself. We find some evidence suggesting that access to condoms, IUDs, and the pill decreases breastfeeding, while access to sterilization increases it. The 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.
- BASU, A. M. (1992): *Culture, The Status of Women, and Demographic Behaviour*. Clarendon, Oxford, UK.
- 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.
- BHATIA, J. C. (1978): “Ideal Number and Sex Composition of Children in India,” *Journal of Family Welfare*, 24(2), 3–16.
- 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).

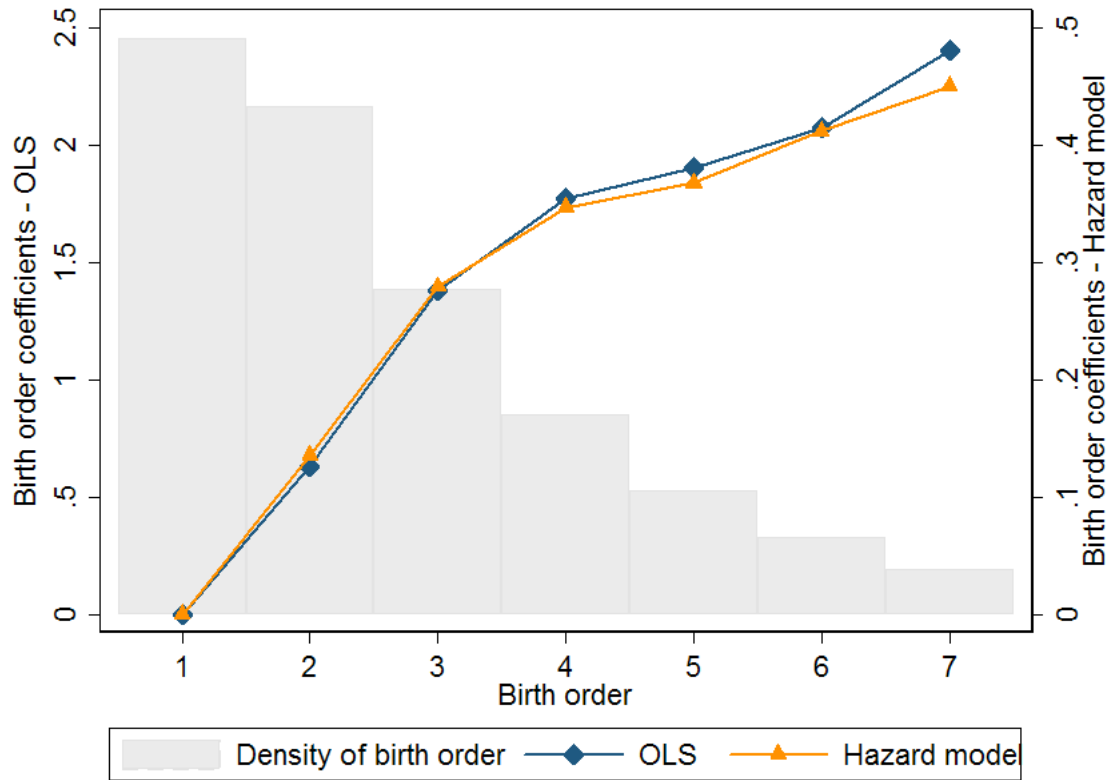


- CHEN, Y., S. YU, AND W.-X. LI (1988): "Artificial Feeding and Hospitalization in the First 18 Months of Life," *Pediatrics*, 81(1), 58–62.
- CLARK, S. (2000): "Son Preference and Sex Composition of Children: Evidence from India," *Demography*, 37(1), 95–108.
- DAS, N. (1987): "Sex Preference and Fertility 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.
- 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.
- FELDMAN, S. (2000): "Nursing through Pregnancy," *New Beginnings*, 17(4), 116–118.
- 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.
- HEMOCHANDRA, L., N. S. SINGH, AND A. A. SINGH (2010): "Factors Determining the Closed Birth Interval in Rural Manipur," *Journal of Human Ecology*, 29(3), 209–213.
- HUFFMAN, S., K. FORD, H. ALLEN, AND P. STREBLE (1987): "Nutrition and Fertility in Bangladesh: Breastfeeding and Post-Partum Amenorrhoea," *Population Studies*, 41, 447–462.
- ISAACS, C. E. (2005): "Human Milk Inactivates Pathogens Individually, Additively, and Synergistically," *Journal of Nutrition*, 40(1), 1286–1288.
- JELLIFFE, D., AND E. JELLIFFE (1978): "The Volume and Composition of Human Milk in Poorly Nourished Communities. A Review," *American Journal of Clinical Nutrition*, 31(3), 492–515.
- JENSEN, R. (2003): "Equal Treatment, Unequal Outcomes? Generating Sex Inequality Through Fertility Behavior," Mimeo, Harvard University.
- JHA, P., R. KUMAR, P. VASA, N. DHINGRA, D. THIRUCHELVAM, AND R. MOINEDDIN (2006): "Low Male-to-Female Sex Ratio of Children Born in India: National Survey of 1.1 Million Households," *Lancet*, 367(9506), 211–218.
- LEUNG, S. F. (1991): "A Stochastic Dynamic Analysis of Parental Sex Preferences and Fertility," *Quarterly Journal of Economics*, 121, 1063–1088.
- 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.
- MOTARJEMI, Y., F. KAUFSTEIN, G. MOY, AND F. QUEVEDO (1993): "Contaminated Weaning Food: A Major Risk Factor for Diarrhoea and Associated Malnutrition," *Bulletin of the World Health Organization*, 71(1), 79–92.
- MUHURI, P. K., AND J. MENKEN (1997): "Adverse Effects of Next Birth, Gender, and Family Composition on Child Survival in Rural Bangladesh," *Population Studies*, 51(3), 279–294.
- 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.
- NEMETH, R. J., AND J. M. BOWLING (1985): "Son Preference and Its Effects on Korean Lactation Practices," *Journal of Biosocial Science*, 17(4), 451–459.
- OSTER, E. (2009): "Proximate Sources of Population Sex Imbalance in India," *Demography*, 46(2), 325–339.
- PALLONI, A., AND S. MILLMAN (1986): "Effects of Inter-birth Intervals and Breastfeeding on Infant and Early Childhood Mortality," *Population Studies*, 40, 215–236.
- PALLONI, A., AND M. TIENDA (1986): "The Effects of Breastfeeding and Pace of Childbearing on Mortality at Early Ages," *Demography*, 23(1), 31–52.
- 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.
- PRENTICE, A. (1991): "Breast Feeding and the Older Infant," *Acta Paediatrica*, 80(s374), 78–88.
- RETFERFORD, R. D., M. K. CHOE, S. THAPA, AND B. B. GUBHAJU (1989): "To What Extent Does Breastfeeding Explain Birth-Interval Effects on Early Childhood Mortality?," *Demography*, 26(3), 439–450.
- RETFERFORD, R. D., AND T. K. ROY (2003): "Factors Affecting Sex-Selective Abortion in India," NFHS Bulletin No. 17.
- RINDFUSS, R. R., L. L. BUMPASS, J. A. PALMORE, AND D. W. HAN (1982): "The Transformation of Korean Child-spacing Practices," *Population Studies*, 36(1), 87–104.

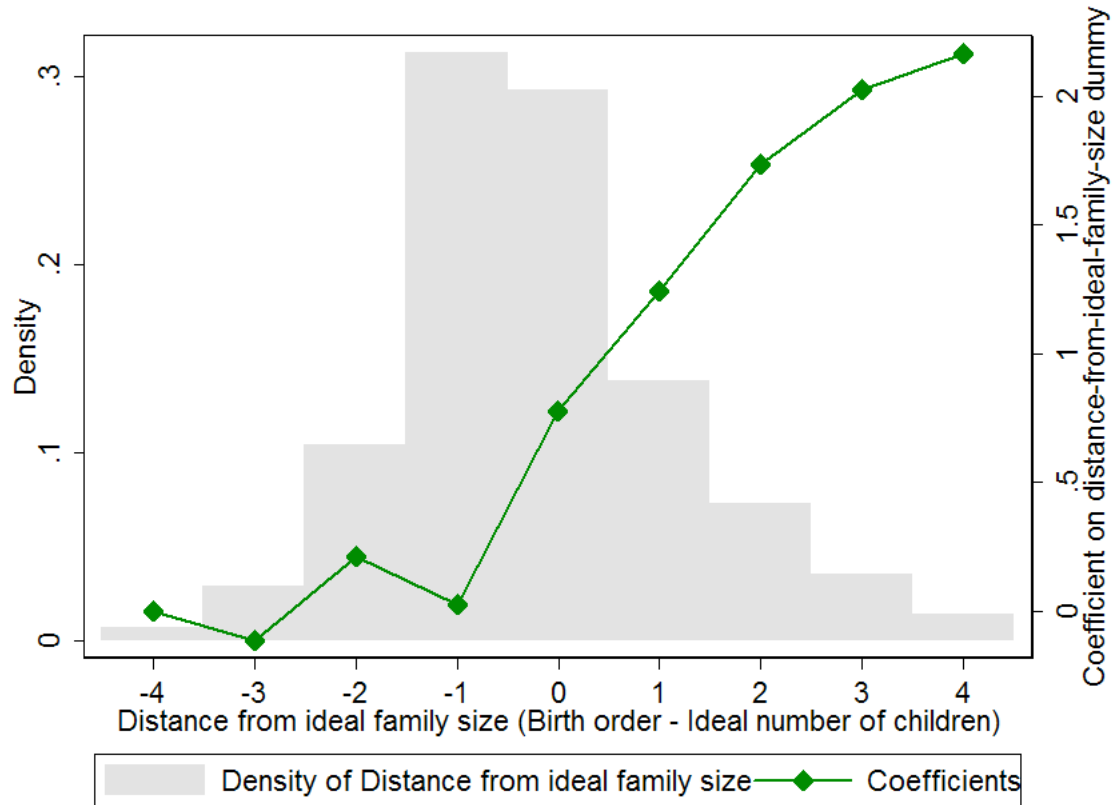
- ROSENZWEIG, M. R., AND K. I. WOLPIN (1980): "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, 48(1), 227–240.
- ROUS, J. J. (2001): "Is Breast Feeding A Substitute for Contraception in Family Planning?," *Demography*, 38(4), 497–512.
- THAPA, S. (1987): "Determinants of Fertility in Nepal: Applications of an Aggregate Model," *Journal of Biosocial Science*, 19(3), 351–365.
- TRUSSELL, J., L. G. MARTIN, R. FELDMAN, J. A. PALMORE, M. CONCEPCION, AND D. N. L. B. D. A. BAKAR (1985): "Determinants of Birth-Interval Length in the Philippines, Malaysia, and Indonesia: A Hazard-Model Analysis," *Demography*, 22(2), 145–168.
- VERD, S., J. MOLL, AND B. VILLALONGA (2008): "Miscarriage rate in pregnancy-breastfeeding overlap," *Contraception*, 78(4), 348–349.
- 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.
- (2010): "Guidelines on HIV and Infant Feeding 2010: Principles and Recommendations for Infant Feeding in the Context of HIV and a Summary of Evidence," Report.
- 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.

Figure 1: Breastfeeding duration, by birth order



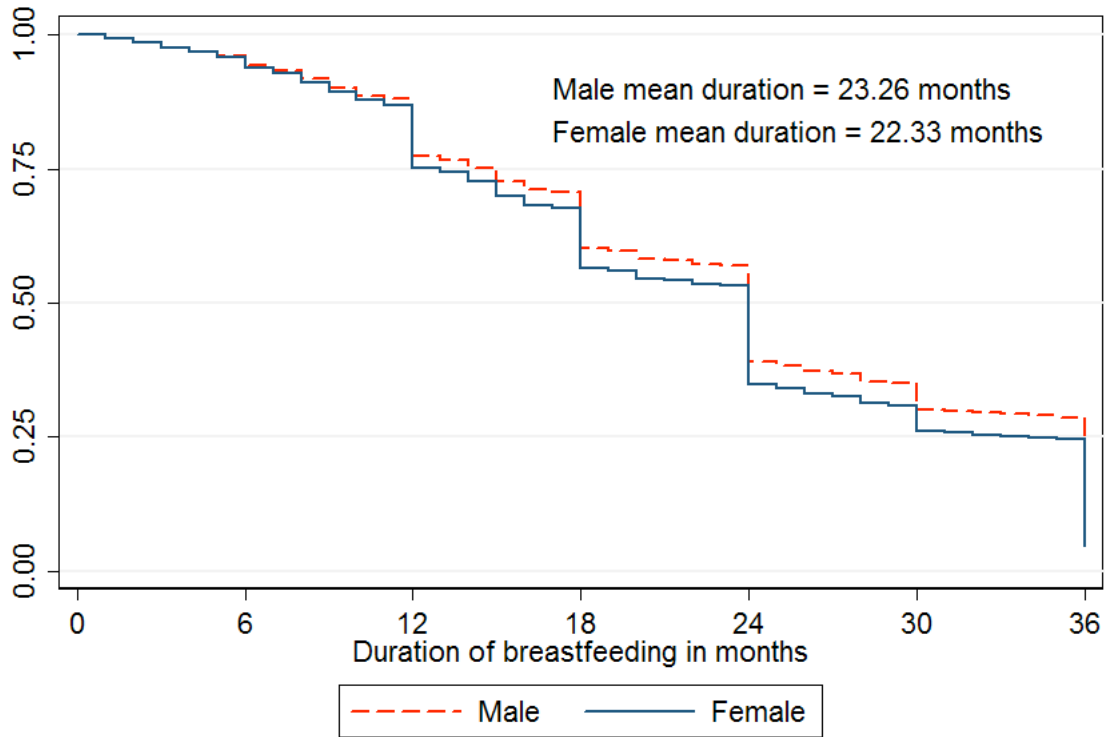
Notes: The figure plots the coefficients for birth-order dummies from a regression with breastfeeding duration in months as the dependent variable. The OLS model includes age-in-month fixed effects; no other control variables are included. The omitted category is birth order 1, for which the coefficient 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



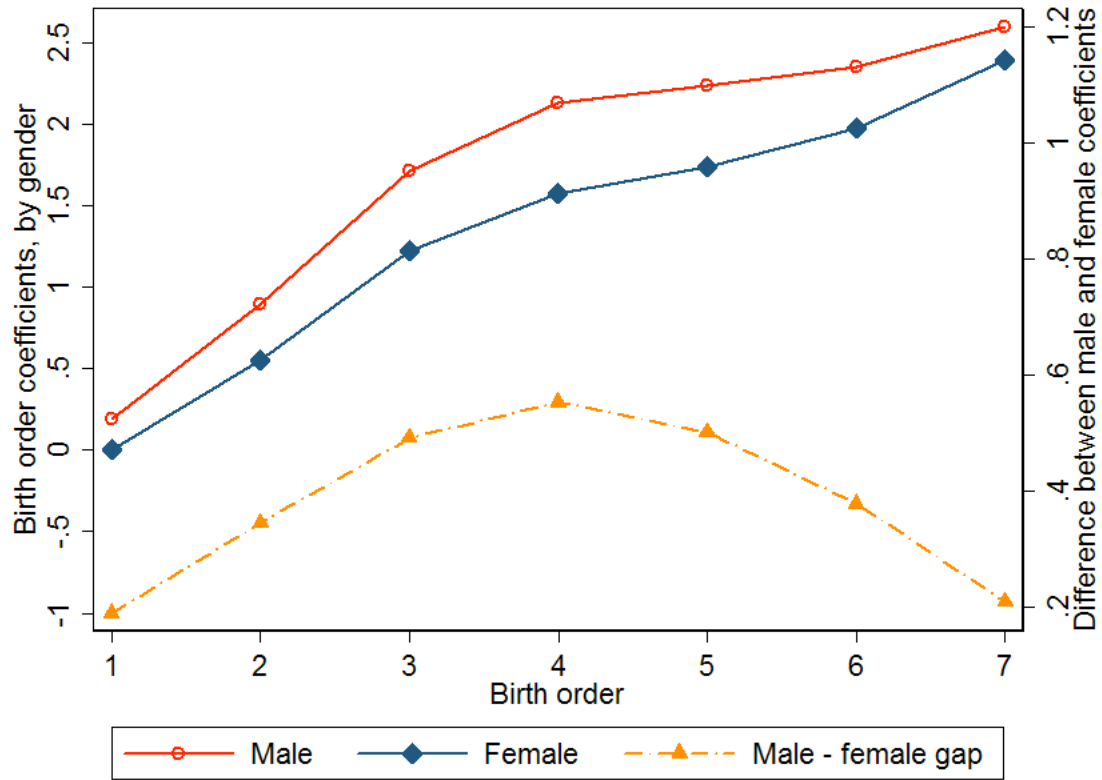
Notes: The figure plots the coefficients for “distance from ideal family size” dummies from a regression with breastfeeding duration in months as the dependent variable. Distance from ideal family size is defined as the child’s birth order minus the mother’s ideal family size. The omitted category is 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. The histogram of the “distance from ideal family size” variable is also displayed.

Figure 3: Survival function for breastfeeding, by gender



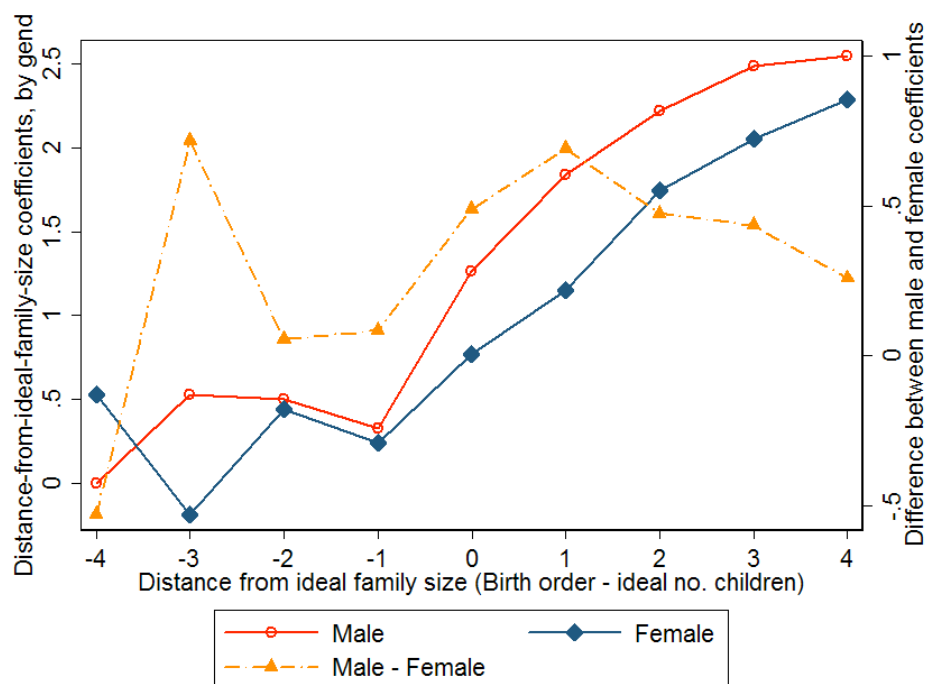
Notes: The figure plots the proportion of children, by gender, who are still being breastfed at the duration (age) given on the horizontal axis. Both the plot and the reported means account for censoring of the breastfeeding duration variable for children who are still being breastfed at the time of the survey. The means listed in the figure are calculated assuming that all observations that report having 36 months of breastfeeding (the top-coded value) have exactly 36 months of breastfeeding. If instead we exponentially extend the survival curve to zero for those observations, the male and female means are, respectively, 23.94 and 22.88, yielding a gender gap of 1.06 instead of 0.93 months. The survival curves are later used to calculate statistics used in the mortality calculation in Section 7. For example, the probability that a girl is breastfed between the ages of 12 and 36 months is the average height of the female survival curve in this age range, or 0.4597. The same probability for boys is 0.4964, yielding a gender gap of 3.67 percentage points. The average probability of being breastfed between 12 and 36 months, also needed in the mortality calculation, can be similarly calculated from the combined survival curve (not plotted), and is 0.481.

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



Notes: The solid lines plot the gender-specific coefficients for birth-order dummies from a regression with breastfeeding duration as the dependent variable, with the coefficient for birth order one for females normalized to zero. 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”



The solid lines plot the gender-specific coefficients for “distance from ideal family size” (birth order minus ideal family size) dummies from a regression with breastfeeding duration in months as the dependent variable. 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. The dashed line is the difference between the male and female coefficients.



Table 1: Summary statistics

	Birth order $\leq 2$	Birth order $> 2$	Sons	Daughters
Months of breastfeeding	14.24 [8.739]	15.54 [9.287]	14.99 [9.093]	14.56 [8.880]
Birth order	1.469 [0.499]	4.109 [1.220]	2.579 [1.571]	2.550 [1.563]
Ideal no. of children	2.404 [0.861]	3.164 [1.195]	2.687 [1.067]	2.739 [1.085]
Birth order minus ideal no. children	-0.915 [0.894]	0.882 [1.354]	-0.156 [1.402]	-0.226 [1.422]
Male	0.513 [0.500]	0.522 [0.500]	1 [0]	0 [0]
Mother has at least one son	0.631 [0.483]	0.915 [0.279]	1 [0]	0.481 [0.500]
Child has no younger sibling	0.769 [0.422]	0.833 [0.373]	0.811 [0.391]	0.779 [0.415]
Total number of vaccinations	3.972 [2.345]	3.090 [2.457]	3.689 [2.412]	3.520 [2.449]
Age of child	1.950 [1.262]	1.920 [1.252]	1.939 [1.255]	1.936 [1.261]
Age of mother	23.72 [4.228]	28.64 [4.816]	25.81 [5.097]	25.71 [5.096]
Rural	0.637 [0.481]	0.743 [0.437]	0.677 [0.467]	0.684 [0.465]
Mother's years of schooling	5.597 [5.144]	2.429 [3.767]	4.333 [4.904]	4.227 [4.852]
Observations	64,439	45,744	56,896	53,287

Notes: Data drawn from 1992, 1998 and 2005 waves of the National Family Health Survey for 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, whose mother has had no more than eight children, and who are singletons. The 1992, 1998 and 2005 waves account for 36.9 percent, 24.6 percent, and 38.5 percent of the observations, respectively. Breastfeeding duration is not 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. “Total vaccinations” is the sum of six indicator variables for whether the child received each of the polio 1, polio 2, polio 3, DPT (diphtheria/pertussis/tetanus) 2, DPT3, and measles vaccines.

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

	OLS		Hazard	OLS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Birth order	0.464*** [0.0124]	0.210*** [0.0179]	-0.0612*** [0.00421]				
$\Delta Ideal \geq 0$				1.072*** [0.0399]	0.876*** [0.0454]	0.773*** [0.0745]	0.399*** [0.0742]
$\Delta Ideal$						-0.0242 [0.0426]	0.320*** [0.0436]
$\Delta Ideal \times (\Delta Ideal \geq 0)$						0.441*** [0.0502]	-0.215*** [0.0515]
Male		0.391*** [0.0373]	-0.105*** [0.00866]		0.375*** [0.0385]		0.374*** [0.0385]
Mother's years of schooling		-0.121*** [0.00503]	0.0289*** [0.00112]		-0.136*** [0.00479]		-0.135*** [0.00489]
Rural		0.806*** [0.0478]	-0.181*** [0.0102]		0.834*** [0.0489]		0.839*** [0.0490]
Covariates	No	Yes	Yes	No	Yes	No	Yes
Observations	110183	110183	108616	104456	104456	104456	104456
R-squared	0.503	0.527		0.496	0.524	0.497	0.524

Notes: 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. The variable *Ideal* is the response of the child's mother to the question: "What is your ideal number of children?" We define  $\Delta Ideal$  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 and a linear control for mother's years of education, as well as dummy variables corresponding to the year of the survey wave and the 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 (see text for fuller explanation). 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 the coefficient estimates should have the opposite sign of those of the OLS regressions. The reason the number of observations is not constant across specifications is that (a) hazard estimations drop observations that immediately exit (i.e., duration of breastfeeding = 0) and (b) we exclude observations where  $\Delta Ideal$  is missing or not in the interval  $[-4, 4]$  in cols. 4 to 7. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table 3: Effect of gender and sibling sex composition on breastfeeding duration

	OLS		Hazard	OLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Male	0.368*** [0.0384]	0.389*** [0.0373]	-0.103*** [0.00867]	0.244*** [0.0486]	0.262*** [0.0546]	0.458*** [0.0675]
Mother has at least one son				0.280*** [0.0623]		
Male share of mother's children					0.231*** [0.0751]	
Male x First survey wave						-0.144 [0.0895]
Male x Second survey wave						-0.0654 [0.0929]
Covariates	No	Yes	Yes	Yes	Yes	Yes
Observations	110183	110183	108616	110183	110183	110183
R-squared	0.497	0.527		0.527	0.527	0.527

Notes: 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. The specifications in columns 2-6 include linear and quadratic controls for mother's age and child's year of birth and a linear control for mother's years of education, as well as dummy variables corresponding to the year of the survey wave, the state of residence, and the child's birth order. 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 (see text for fuller explanation). 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 specifications is that the hazard estimations drop observations that immediately exit (i.e., duration of breastfeeding = 0). \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table 4: Effect of gender and birth-order interactions on breastfeeding duration

	OLS		Hazard	OLS	
	(1)	(2)	(3)	(4)	(5)
Male	-0.0839 [0.135]	-0.0661 [0.131]	0.0203 [0.0301]	-0.00614 [0.134]	-0.0188 [0.130]
Male x Birth order	0.299*** [0.0944]	0.311*** [0.0923]	-0.0847*** [0.0216]		
Male x Birth order <sup>2</sup>	-0.0365*** [0.0135]	-0.0381*** [0.0132]	0.00997*** [0.00316]		
Male x ( $\Delta Ideal \geq 0$ )				0.548*** [0.150]	0.590*** [0.146]
Male x $\Delta Ideal$				-0.0827 [0.0846]	-0.102 [0.0820]
Male x $\Delta Ideal$ x ( $\Delta Ideal \geq 0$ )				0.0665 [0.0991]	0.113 [0.0962]
Covariates	No	Yes	Yes	No	Yes
Max effect of male at birth order...	4.09	4.09	4.25	N/A	N/A
Observations	110183	110183	108616	104456	104456
R-squared	0.504	0.527		0.498	0.524

Notes: Standard errors, in brackets, are adjusted for clustering by mother. See notes to previous tables for all variable definitions. 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. Cols. 1 to 3 include birth-order fixed effects and columns 4 and 5 include fixed effects for each value of  $\Delta Ideal$ . The remaining covariates and age-in-month controls are as in Tables 2 and 3. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table 5: Effect of son preference on duration of breastfeeding

	Regional variation in son pref.		Individual var. in son pref.	
	(1)	(2)	(3)	(4)
Male	-1.922** [0.842]	1.750 [1.892]	0.260*** [0.0456]	0.366* [0.202]
Male x State sex ratio	2.145*** [0.781]	-1.648 [1.760]		
Male x ( $\Delta Ideal \geq 0$ ) x Sex ratio		6.218** [2.529]		
$\Delta IdealSons \geq 0$			0.408*** [0.0890]	
$\Delta Ideal \geq 0$			0.328*** [0.0779]	
Male x ( $\Delta IdealSons = 0$ )				0.205** [0.0956]
Male x ( $\Delta IdealSons > 0$ )				-0.0603 [0.129]
Covariates	Yes	Yes	Yes	Yes
Observations	110183	104456	104333	105054
R-squared	0.527	0.525	0.524	0.525

Notes: Standard errors, in brackets, are adjusted for clustering by mother. State sex ratio comes from the 2001 Indian Census. We define *IdealSons* as the mother's answer to the question, "What is your ideal number of sons?" We define  $\Delta IdealSons$  as *Total number of sons* – *Ideal number of sons*. All regressions include the standard set of covariates and age-in-month fixed effects included in previous tables (see the notes to Table 3 for a list). In addition, column 2 includes the two-way interactions of *Male*,  $\Delta Ideal \geq 0$ , and *Sex ratio* and fixed effects for  $\Delta Ideal$ ; column 3 includes  $\Delta IdealSons$ ,  $\Delta IdealSons \times (\Delta IdealSons \geq 0)$ ,  $\Delta Ideal$ , and  $\Delta Ideal \times (\Delta Ideal \geq 0)$ ; and column 4 includes fixed effects for  $\Delta IdealSons$ . The results in columns 3 and 4 are essentially identical if we also include fixed effects for the total number of sons.

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

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

	Dep. var.: Total number of vaccinations				
	(1)	(2)	(3)	(4)	(5)
Male	0.127*** [0.0115]	0.131*** [0.0117]	0.143*** [0.0157]	0.0123 [0.0391]	0.0631 [0.0412]
Birth order	-0.146*** [0.00600]				
$\Delta Ideal \geq 0$		-0.0235 [0.0178]			
Mother has at least one son			-0.0323 [0.0202]		
Male x Birth order				0.0549* [0.0291]	
Male x Birth order <sup>2</sup>				-0.00291 [0.00432]	
Male x ( $\Delta Ideal \geq 0$ )					0.0305 [0.0455]
Covariates	Yes	Yes	Yes	Yes	Yes
Additional fixed effects	None	None	Birth order	Birth order	$\Delta Ideal$
Observations	107212	101684	107212	107212	101684
R-squared	0.406	0.403	0.406	0.406	0.406

Notes: Standard errors, in brackets, are adjusted for clustering by mother. *Total vaccinations* is the sum of six indicator variables for whether the child received each of the polio 1, polio 2, polio 3, DPT (diphtheria/pertussis/tetanus) 2, DPT3, and measles vaccines; these are the six vaccines for which data are collected in each of the three waves. See notes to previous tables for all other variable definitions. All regressions include the standard set of covariates in previous tables (see the notes to Table 3 for a list), in addition to those listed. Column 5 also includes the variables  $Male \times \Delta Ideal$  and  $Male \times \Delta Ideal \times (\Delta Ideal \geq 0)$ .

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table 7: Child mortality between 12 and 36 months

	Household lacks piped water			Household has piped water		
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.00851*** [0.000866]	0.00369 [0.00291]	-0.00522* [0.00285]	-0.00388*** [0.00103]	0.000131 [0.00376]	-0.00828* [0.00437]
Male x Birth order		-0.00619*** [0.00220]			-0.00272 [0.00316]	
Male x Birth order <sup>2</sup>		0.000476 [0.000331]			0.000313 [0.000529]	
Male x ( $\Delta Ideal \geq 0$ )			-0.00485 [0.00324]			0.00581 [0.00462]
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
$\hat{\beta}_{\text{unpiped}} - \hat{\beta}_{\text{piped}}$ , coeff(s) of interest	-0.00465	-0.00350 0.000168	-0.0106			
<i>F</i> -test of above coeff difference(s) ( <i>p</i> -value)	0.000560	0.0497	0.0600			
Observations	125857	125857	116957	35164	35164	33850
R-squared	0.00965	0.00992	0.00912	0.00749	0.00754	0.00782

Notes: Standard errors, in brackets, are adjusted for clustering by mother. All regressions are estimated by OLS and control for the standard set of covariates in previous tables (see the notes to Table 3 for a list). In columns 1, 2, 4, and 5, birth-order dummies are included. In columns 3 and 6, dummies for each value of  $\Delta Ideal$  are included, as well as controls for  $Male \times \Delta Ideal$  and  $Male \times (\Delta Ideal \geq 0) \times \Delta Ideal$ . The sample includes children born between 37 and 96 months before the survey date. The lower bound accounts for the fact that children younger than 37 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). 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 4 (where the *p*-value is for *Male*); columns 2 and 5 (where the *p*-value is for the joint test of the linear and quadratic terms); and columns 3 and 6 (where the *p*-value is for  $Male \times (\Delta Ideal \geq 0)$ ).

\* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

## Appendix: Model

### 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.

### 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  $\beta$ . Time periods are denoted by  $t$ . In each period, a woman gives birth to either one child or no children.

A mother who gives birth 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. 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).<sup>28</sup>

The decision problem is therefore:

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

---

<sup>28</sup>Our assumption that mothers choose to breastfeed after their last birth 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  $\lim_{n \rightarrow \infty} q'(n) = -\infty$ . 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  $\hat{n}$ .

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

Son preference factors into the breastfeeding decision once  $n > \hat{n}$ . Intuitively, mothers unambiguously gain from having up to  $\hat{n}$  children, after which point they weigh the net cost of having more children (which grows unboundedly) against the value of trying for more sons.<sup>29</sup>

### 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  $n^{\text{th}}$  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.

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

- (i) *the child is male; or*
- (ii) *more of his or her older siblings are male.*

*Proof.*

---

<sup>29</sup>Note that we are interested in the case where a mother wants to have at least one child, or  $\hat{n} > 0$  (equivalently,  $q(1) - q(0) > 0$ ). Lemma 1 shows that under this assumption there exists a unique  $\hat{n}$ , 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  $\hat{n}$  because of the marginal benefit of more sons  $\lambda[g(\hat{n}+1) - g(\hat{n})]$ , but every mother continues up to at least  $\hat{n}$ .

- (i) By Lemma 3, a mother will breastfeed if and only if  $u(n, s) \geq \frac{u(n+1, s) + u(n+1, s+1)}{2}$ . Dropping the terms in the utility function that do not depend on  $s$ , we have that the decision to breastfeed is increasing in  $g(s) - g(s+1)$ , which is increasing in  $s$  since  $g'' < 0$ . Holding the sex of the first  $n - 1$  children fixed,  $s$  is higher when the  $n^{\text{th}}$  child is a boy, so a son is more likely to be breastfed than a daughter. For sufficiently large  $\lambda$ , 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  $n^{\text{th}}$  child fixed,  $s$  is increasing in the number of boys among the first  $n - 1$  children, so an  $n^{\text{th}}$  child with more brothers among his or her siblings is more likely to be breastfed. For sufficiently large  $\lambda$ , 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  $\hat{n}^{\text{th}}$ . At sufficiently high birth order, breastfeeding is again the same for sons and daughters: by Lemma 2, 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  $n^{\text{th}}$  child is male, and will therefore breastfeed a son but not a daughter.

### Predictions regarding “ideal family size”

The model also predicts that breastfeeding patterns can change abruptly when  $n$  reaches  $\hat{n}$ . While a mother’s “ideal” quantity of children is ill-defined in the presence of son preference, one can think of  $\hat{n}$  as one measure of the mother’s preferred quantity: it is the quantity at which she 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  $\hat{n}$  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  $n < \hat{n}$ , by Lemma 1. Proposition 1 establishes the remainder of the proof. The mother will begin breastfeeding at some  $n \geq \hat{n}$ , the exact value depending on the gender composition of her children and the extent of her son preference.
- (ii) The proof to part (i) establishes that neither boys nor girls are breastfed for  $n < \hat{n}$ , so there is no gender gap. Proposition 2(i) establishes the possibility of a gender gap at  $n \geq \hat{n}$ . ■

The proofs for the following lemmas, which are used above, are in the online appendix.

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

**Lemma 2.** The total number of children that a mother gives birth to is bounded above; that is, there exists  $\bar{n}$  such that she never has more than  $\bar{n}$  children, regardless of sex composition.

**Lemma 3.** 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 of breastfeeding rather than childbearing).

### 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  $\phi$  (demand for children) and  $\lambda$  (son preference).

Second, the binary specification does not take into account mothers' using breastfeeding to space births as opposed to merely prevent them. The essential assumption to incorporate birth-spacing into our framework is that mothers who want to continue having children might space their births to an extent, but they will still breastfeed less than mothers who want to stop having children entirely. This assumption follows naturally from a model (such as ours) where a mother receives flow utility from her children; she will begin receiving the positive flow of utility sooner if she has her children sooner.<sup>30</sup>

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.<sup>31</sup>

---

<sup>30</sup>Stepping outside the model, fecundity declines with age, so delaying childbearing also carries the risk that the mother becomes less able to conceive.

<sup>31</sup>For example, consider a modification of the model such that a mother chooses whether to breastfeed her child both in the period the child is born and 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 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.