Explaining Third Birth Patterns in India: Causal Effect of Sibling Sex Composition

Shanjukta Nath*

University of Maryland, College Park

Abstract

Son preference is a well-established phenomenon for India. This preference gets reflected in multiple dimensions of the childbearing process such as the size of the family and birth spacing. I use the sibling sex composition of the first two children to capture its impact on the third birth interval, induced by a preference for a son. Sibling sex composition provides a credible source of exogenous variation in the Indian context for births before 1990 as gender screening became widespread only after the economic reforms in 1990. My analysis shows that on average families with two sons face a 9% lower hazard of third birth relative to families with two daughters. This hazard ratio translates into a gap of roughly one month in the average third birth interval. I also show that sibling composition affects the proportion of third births spaced below 18 months, a critical cut-off for neonatal, post-neonatal and child mortality. Inter-birth intervals of less than 18 months increase the chances of the third child's mortality by 10% in my sample. A back of the envelope calculation based on these estimates suggest that about 1,500 infant deaths every year in India can be attributed to a higher proportion of daughters among the first two births.

JEL Codes: J13, J16, O15

Keywords: Childbearing, Fertility, Gender Discrimination, Demographic Economics, India.

*Department of Economics, University of Maryland, 3114 Preinkert Dr, Tydings Hall, ,College Park, MD 20742. Contact: nath@econ.umd.edu

1 Introduction

The accumulation and growth of human capital are crucial for sustained economic development. Better child health and nutrition results in an increment in total factor productivity (Becker et al., 1990; Becker, 1994). Completed fertility and its transition play a crucial role in this process. Life-cycle fertility, in turn, is a multi-faceted process. Parents have preferences over the timing and spacing of births along with total family size. Short birth intervals usually result in a high fertility environment. Besides, the spacing of births can have ramifications for human capital outcomes of children within the household. In this spirit, birth spacing affects both the quality-quantity dimensions (Becker, 1960; Becker and Lewis, 1973) of household preferences over children. These two components are ultimately indispensable for economic development.

India is home to currently 1.3 billion people, the second most populist country in the world. Consequently, researchers have focused on learning the determinants of high fertility in this country. Factors such as infant mortality (Agrawal, 1975; Murthi et al., 1995), cultural norms favoring large families (Fawcett, 1983), and son preference resulted in above replacement level fertility over the last couple of decades.[1] The latter has emerged as decisive for parity progression as families continue childbearing till they attain the desired number of sons (Clark, 2000; Jensen, 2003; Das, 1987; Arnold et al., 1998; Devi, 2013; Chaudhuri, 2012) .[2,3]

Although there exists comprehensive research on fertility, there is limited literature on birth spacing patterns in India. However, there is a pressing need for a better understanding of birth spacing as medical studies recommend avoiding shorter birth intervals for better child health outcomes (Zhu et al., 1999; Conde-Agudelo et al., 2006). Sibling competition becomes even more intense for families with binding credit constraints. Such a situation is relevant for a significant fraction of households in India with limited household resources. Both preceding and succeeding birth intervals can affect child well-being. Children born before and after a short interval might face escalated chances of mortality. I discuss here the three primary mechanisms of heightened risk as suggested in Boerma and Bicego 1992; Conde-Agudelo et al. 2012. These mechanisms include "maternal nutritional depletion, folate depletion, transmission of infections between closely spaced siblings and sub-optimal lactation related to breastfeeding" (Conde-Agudelo et al., 2012, p95, Table 1) .[4]

Most of the earlier work on birth spacing in India is descriptive and not causal. It is hard to interpret the effects of explanatory variables in the birth spacing equation as causal due to the presence of unobserved heterogeneity across households such as the mother's biological characteristics, bargaining dynamics within the family. Also, the existence of unobservables at the level of the immediate environment such as cultural norms influencing birth spacing choices poses challenges to identification. Examples of such unobservables include the prevalence of religious practices that dictate days of abstinence on certain occasions or the use of contraceptives being dependent on group behavior. Lastly, studies have used small samples where inference might be unreliable due to low power, large confidence intervals and

[1] "Replacement level fertility is 2.1 children per woman, and this represents the average number of children a woman would need to have to reproduce herself by bearing a daughter who survives to childbearing age" "Total Fertility Rate 2018", para 1; Currently, the fertility rate in India is 2.2 children per woman, slightly above the replacement level fertility.

[2] Parity progression refers to families moving to higher order births.

[3] Evidence of such behavior exists for DLHS-2, and it is provided in section 3.4

[4] Folate depletion refers to a deficiency of folic acid which is critical for producing red blood cells and DNA (Snow, 1999).

limited external validity (Verma et al., 1990).[5]

Researchers in the past have tried to control for unobserved heterogeneity in the birth spacing mortality equation by applying mother and residential cluster fixed effects (Zenger, 1993; Hobcraft et al., 1985; Bhargava, 2003). Such fixed effects correct for time-invariant unobservables in this relationship. However, it cannot correct for time-varying unobservables at the level of households correlated with both birth intervals and child mortality. Moreover, cluster fixed effects cannot capture the impact of time-varying correlated shocks for families within a group.[6]

Apart from fixed effects, some former work used miscarriages between two live births as a source of exogenous variation in birth spacing such as Buckles and Munnich 2012 for the United States. Here, the authors use this instrument to disentangle the effect of birth spacing on child outcomes conflated with unobserved heterogeneity within and across families. The use of miscarriages as a source of exogenous variation is less appealing for a developing country like India. The average nutritional deficiencies are much higher for mothers in India compared to the developed world. Hence, biologically these mothers might face a higher risk of miscarriage, which in turn poses a threat to the IV strategy.

Given these caveats of the existing work, I make use of sibling sex composition as a source of variation which is more suited and credible for the Indian context. The suitability of this variable rests on the following factors. First, son preference is a well-established phenomenon for India. Families are expected to respond differently to the different realizations of sibling composition of the first two births. Existing research has suggested waiting time to next child delivery are longer after a son is born relative to a daughter. Hence, ex-ante I anticipate waiting time to third birth is likely to be directly related to the proportion of sons among the first two births. Second, families with three births overwhelmingly represent more than two-thirds of the total sample[7]. This aspect of data establishes the importance of the third birth in the childbearing process in India. Third, the demographic policy has actively pushed for two-births in India. But, it has failed to achieve the desired results during most of the decades following independence.

Sibling sex composition is a credible source of exogenous variation if the gender of a child is a random event. Research in medicine, anthropology, and sociology have not been able to establish meaningful biological or socio-economic links to the sex of a child (James, 1987; Gilbert and Danker, 1981; Rodgers and Doughty, 2001). This information is critical for identification in my empirical specification. However, it is not enough to ensure exogeneity of the variable as there is some evidence of sex-related abortions in India. I tackle this issue by restricting births to on or before 1990[8]. This cut-off aides identification as gender screening became widespread in India post the economic reforms in 1990.

I estimate the effect of sibling sex composition of the first two births on various aspects of third birth for India. First, does sibling composition affect the probability of third birth? I delve deeper and measure the effect of sibling composition on third birth interval patterns. Ultimately, I make use of these results

[5]For example, Verma et al. 1990 uses a sample of 73 pregnant females in a rural health center in Ludhiana, Punjab for examining the effect of gender of the previous child on succeeding birth interval.

[6]An example of a correlated shock for fertility is the time-varying growth in access and use of contraception which might be dependent on group behavior.

[7]When I restrict the DLHS-2 to mothers in the age group of 21-32 for a direct comparison with (Angrist and Evans, 1998), it shows that 67% of mothers had at least three births by 1990.

[8]Barcellos et al. 2014; Jayachandran and Pande 2017 use the same restriction to address threats to identification arising from gender screening in their empirical work; therefore, earlier work in India has noted that births before 1990 are unlikely to be manipulated by parents using the sex detection technology.

to determine the welfare implications of birth spacing on child health outcomes.

Sibling composition affects both the likelihood and spacing of third birth for Indian households. The probability of third birth is inversely related to the proportion of sons among the first two births. I find that families with two daughters have a likelihood of 0.75 for third birth relative to 0.69 and 0.67 for couples with one or two sons respectively.

Households with a higher proportion of sons are likely to wait longer for the third child relative to families with more daughters. Families with two sons face about 9% lower hazard of a third birth in every period compared to families with two daughters. Moreover, households with one son face a 5% lower hazard of third birth. Overall, the median survival time for families with two boys is a month longer than families with two girls.

The difference of a month driven by sibling sex composition is expected to have limited economic significance for child welfare. Considering I focus on short-term child health consequences, observations on the left tail of the birth interval distribution is crucial for my analysis. Particularly, birth intervals shorter than 18 months pose very high mortality risk for children (Cleland and Sathar, 1984; Hale et al., 2009; Whitworth and Stephenson, 2002). Beyond the average effect, a higher proportion of sons make parents less inclined to undertake these risky births. Parents with two sons face a 2.5% lower risk of third birth within the first 18 months of second birth relative to families with two daughters.[9]

Birth intervals shorter than 18 months raise the risk of child mortality approximately by 10%. Translating these probabilities into the number of deaths using the cumulative number of live births in India per year suggests that nearly 1,500 infant mortality deaths per year are due to the sibling sex composition effect of birth spacing outcomes.

The layout of the paper is as follows. To familiarize the readers more with the context, I follow-up the introduction with a detailed literature review on son preference and birth spacing in Section 2. Section 3 provides details on the data-sets used for empirical analysis and key variables. Section 4 discusses the identification and methodology for estimation. Next, I present the results in section 5 and robustness checks in section 6. I consider the welfare implications of birth spacing on child mortality in section 7 and I conclude the paper in section 8.

2 Literature Review

In an early article Ben-Porath and Welch 1976 proposed the mechanisms for parental preference over the sex of children. This preference can either arise due to variations in parental taste or because parents expect to derive higher net benefits from a child of a particular gender. An example of the former is parents having a liking for mixed sibling over same-sex composition in developed countries[10]. The latter can arise due to higher net gains from raising a daughter in a matrilineal society where women play the primary role in a household[11].

In the Indian context, son preference is critical for the empirical analysis. There are numerous social

[9] The corresponding coefficients using the NFHS-1 is 2.9% for families with two sons relative to families with two daughters.

[10] Angrist and Evans 1998; Conley and Glauber 2006 leverage parental preference for mixed sibling composition in the United States over same-sex composition to study its effect on fertility.

[11] Kaul 2018 suggests how there exist preference for daughters over sons in the matrilineal state of Meghalaya, India.

and economic mechanisms at work which make a son more desirable than a daughter. Cultural and religious practices help to provide a higher status to sons in the society. Such practices constitute religious rites and rituals concerning the deceased among Hindus such as "Shraddha" which can only be performed by sons (Vlassoff, 1990; Dyson and Moore, 1983; Das Gupta et al., 2003; Pande and Astone, 2007; Das Gupta, 2010). Marriage customs such as "Kanyadan" where the bride is given away to the groom by the bride's father further strengthens the rank of sons within a household (Borooah et al., 2009; Patkar, 1995) .

Besides social value, sons have meaningful economic value in the Indian community. Prior work has established the importance of sons for old age security (Nugent, 1985; Becker et al., 1988), particularly relevant for a society with near-absent social pension schemes. In addition to this, sons are a vital source of agricultural labor(Mayer, 1999; Malhotra et al., 1995). Lastly, societal customs and traditions bestow huge economic gains on sons such as dowry payments (Anderson, 2003; Srinivasan and Lee, 2004) and inheritance of land and assets (Agarwal, 1995).

Parents preference for a boy gets translated into the differential allocation of household resources. Parental discrimination is reflected in parental care time as shown in Barcellos et al. 2014. In this article, authors use a time use study for India and suggest that parents allocate less time to girls than boys. Regarding nutrition and health, such behavior gets reflected in breastfeeding duration (Jayachandran and Kuziemko, 2011; Fledderjohann et al., 2014; Nath et al., 1994), food allocation (Behrman, 1988; Pande, 2003; Borooah, 2004; Bose, 2011) and expenditure on health-care services (Pandey et al., 2002; Oster, 2009) . Previous work has similarly presented evidence of bias in educational investment within Indian households. Parents are inclined to spend more on schooling for boys than girls. Additionally, boys have a higher probability of attending relatively expensive private schools than girls. Girls are mostly sent to government schools (Azam and Kingdon, 2013; Zimmermann, 2012; Saha, 2013).

This differential allocation of resources by gender results in adverse child outcomes in several dimensions. First, infant mortality rates have been found to be consistently higher for girls (Sen, 2001; Gupta, 1987; Murthi et al., 1995; Das Gupta and Mari Bhat, 1997). Second, such bias can have implications for child health. For instance, Pal 1999 and Jayachandran and Pande 2017 find gender bias in anthropometric measures. Third, similar to health consequences, lower level of education expenditure on girls leads to low enrolment rate and a lower mean for educational attainment(Filmer and Pritchett, 1998; Gandhi Kingdon, 2002; Dercon and Singh, 2013; Kingdon, 1998).

Sibling sex composition has been utilized extensively to estimate the mother's labor supply fertility relationship in developed countries (Angrist and Evans 1998 for the United States; Daouli et al. 2009 for Greece). It has also been used to disentangle the influence of sibling size on child educational outcomes from unobservables affecting both sibling size and test scores in the economic equation of interest (Currie and Yelowitz, 2000; Goux and Maurin, 2005; Conley and Glauber, 2006). The first stage in these articles estimates the impact of sibling composition of the first two births on the probability of third birth. The first stage exists as parents have a mixed sibling preference in the high-income countries.

Sibling composition have been used as an exogenous source of variation for fertility in some developing economies (Cruces and Galiani 2007 for Argentina and Mexico; Mace and Sear 1997 for a nomadic population in Kenya; Tu 1991 for China; Coombs and Sun 1978 for Taiwan and DAddato 2006 for Morocco). In all these studies, sibling composition of the first two births affects the likelihood of third birth. Broadly, some of these studies also suggest that parents prolong the birth of the next child after a boy is born.

There is limited evidence of the use of sibling gender composition to investigate birth spacing patterns, especially for India. The closest article to my study is Nath and Land 1994 and it provided a starting point for my empirical analysis. The authors use a sample of 803 mothers from the rural areas of Karimganj district in Assam.[12] This sample also included mothers who had experienced mortality of either one of the first two births. The results suggest that a higher proportion of sons is inversely related to the hazard of third birth in every period under analysis.

However, this article suffers from a couple of limitations regarding the sample, estimation and welfare implications. Authors restrict the sample to scheduled caste population in one district in Assam among its 33 districts. It constrains the application of the predicted coefficients belonging to other socio-economic sections and regions within or outside Assam of other states in India. Besides, it works on the assumption that families care about sibling composition of the first two births irrespective of their survival status. This assumption is not convincing as parents are unlikely to care about the sex of the dead child.

My empirical analysis uses a nationally representative data-set. Consequently, the estimates apply to a much larger sample. Second, I restrict the sample to families for whom the first two children survived. Since I work with DLHS-2, which is by far the largest demographic health survey for India with complete retrospective birth histories for mothers, I have enough power for estimation even after forcing survival status restrictions. Such constraints can lead to low power in small sample studies. This distinction also separates my work from other related work in India that estimate infant mortality and birth spacing simultaneously for all births irrespective of birth order (Bhalotra and Van Soest, 2008; Whitworth and Stephenson, 2002). The advantage of using mothers with surviving children controls for unobserved biological characteristics such as various nutrient deficiencies which raise infant mortality risks for those mothers who suffered child mortality.

Regarding estimation Nath and Land (1994) applies a hazard model with fixed covariates to time to third birth. Hazard model with time-varying covariates relaxes this assumption, and I use this flexible version of the Cox model for my main estimation. Such flexible models might be more suited for the present context as there exist time-varying covariates in the model such mother's age.

Finally, Nath and Land's study does not attempt to translate their hazard ratios to differentials in the birth interval for families with different sibling composition. This inadequacy limits the ability of comparison of the magnitudes of their coefficients to related work in other developing countries. Additionally, it lacks any discussion on the economic significance of their results regarding welfare implications within the households. It does not provide any suggestion for risky births and the features of the families that undertake these births in response to sibling composition.

3 Data and Variables

I use the second round of the District Level Household and Facility Survey (DLHS) for my main estimation (IIPS, 2006). This data-set is suitable for my study as it is the largest nationally representative data-set for India with complete retrospective birth histories for mothers. The subsequent rounds of

[12] Assam is a north-eastern state in India which has a total of 33 districts

DLHS had larger sample sizes but did not incorporate the full birth histories for the respondents. For example, DLHS-3 (2007-2008) interviewed a sample of more than 600,000 women. However, it recorded the reproductive history for births within the last five years before the survey date (IIPS, 2010).

DLHS-2 interviewed a sample of 507,622 women from 593 districts in India (Table 9.1).[13] It surveyed in two phases. Phase 1 was carried out in 2002 and covered 295 districts. And, surveyors completed phase 2 in 2004. The respondents were qualified to be questioned if they were currently married and in the age group of 15-44. For the sampling design, the surveyors adopted a two-step stratified scheme. Within each district, the survey selected about 40 urban towns or rural village units. Moreover, within each urban or rural unit, around 1,000 households were selected in the final sample. Consequently, the stratification was done at two stages-first selecting urban or rural units and then randomly drawing units of households from these chosen villages or urban units.

In addition to DLHS-2, I use a smaller sample NFHS-1 for specific segments of my empirical investigation. NFHS-1 happened in 1992-1993, and it surveyed around 89,000 women in the age group of 13-49 on their entire retrospective birth records (IIPS, 2007).[14] Using this data-set renders the following advantages for estimation. First, NFHS-1 provides a suitable robustness test for my results as most of the births in this survey occurred before gender screening became widespread. Consequently, the identifying assumption of randomness of the sex of the child is expected to hold for these births. Second, this survey handles the issue of recall bias that might be conflating my main estimation. Recall bias is a problem in DLHS-2 as the study noted the birth histories in 2002-2004, unlike NFHS-1 where it was in done in 1992-1993. Third, I use it to analyze the characteristics of families and its influence on the outcome variable of interest as it has better quality data on the socio-economic aspects of respondents. For instance, NFHS has more detailed and complete data on parental education, income status, and father's characteristics.

3.1 Key variables used in the analysis

Given my source of variation, I examine patterns of third birth decisions. I begin by considering the households who decide to advance to third birth. Table 9.2 compares the fertility measures and sibling composition of DLHS-2 with Table 1 in Angrist and Evans 1998, p453. Sample weights weight the averages in Table 9.2. The two samples are comparable as they contain women in comparable age group around a similar period.[15] As expected, average fertility is higher for Indian households than American families. The mean number of children for Indian homes is 3.3 as compared to 2.5 in PUMS(1990). This distinction makes it more appropriate to study third birth spacing for India relative to the United States. Moreover, more than two-thirds of these mothers had at least three births by 1990 relative to 37.5% among similarly aged respondents in PUMS. Additionally, the mode of fertility for the sample of these mothers in DLHS-2 is 3. This modal value makes it further relevant to study third birth spacing patterns in India for the period before 1990.

The probability of having a boy in the first birth is 0.512 for the United States and 0.513 for India. For the second birth, the likelihood of a boy is 0.511 for the US and 0.509 for India.[16] These probabilities are very close to those proposed by the medical literature in the absence of any manipulation by parents.

[13] The number of districts is based on the 2001 census. Currently there are 719 districts in India as of 2018.

[14] The surveyors administered NFHS-1 in three phases in Delhi and 24 states.

[15] For Angrist and Evans 1998 the age of respondents in their sample varied from 21 to 35.

[16] For the Indian sample these probabilities are computed for mothers who had three births by 1990 to avoid gender screening.

If the sex ratio at birth is close to 105 males per 100 females, it provides evidence in support of the gender of the child being random at birth. This number corresponds to a likelihood of 0.513 for the birth of a boy. Moreover, the percentage of families with two daughters or two sons in India is very close to the proportion observed in the United States. These comparable estimates support exogeneity of sex at birth as demographers have found sex ratios post-1990s to be skewed in India.

Having confirmed the importance of third birth for families in India, I now examine third birth spacing of these households. Indian families have an average third birth interval of 29.8 months. This length is comparable to the median birth intervals in other developing economies such as the Philippines, Malaysia, Senegal, and Morocco (Smith, 1985).[17] Figure 9.1 displays the distribution of third birth interval for households who had their third birth by 1990. A few critical observations regarding the distribution are it is right skewed, a notable fraction of births fall below the crucial cut-off of 18 months, and more than three-fourths of third births occur within the first four years of the second child's birth (Figure 9.2).

Moving further, I associate third birth decisions with sibling composition of the first two births. First, I measure the differences in the likelihood of third birth by sibling composition of previous births. For this analysis, I concentrate on families for whom first two children survived. An examination of the percentage of households who proceed to third birth indicates a male bias operating this choice. Table 9.3 shows on average families with two alive daughters are 8% more likely to advance to the third birth. Besides, couples with at least one son are 5% less likely to have a third child relative to families with two daughters. Panel B in this table tests for the differences in these fractions. These differences turn out to be significant. This conclusion is critical for the discussion on sample selection in section 4.

Second, I consider birth spacing patterns for the households who had at least three births by 1990. As anticipated, parents with a higher proportion of sons among the first two births are inclined to wait longer for the third birth (Table 9.4). Table 9.4 is on the same lines as Nath and Land, 1994, p383, Table 1. The average birth interval for parents with two sons is slightly over a month longer than the mean for those with two daughters. Moreover, sibling sex composition not just moves the mean interval but also influences parents decision to space the third child within the first 18 months of previous childbirth. The percentage is 17.4% for families with two surviving daughters relative to 14.9% for families with two surviving sons.

Besides sibling composition, there are other observables of the household that has significant explanatory power for birth intervals. I classify the vector of observables into two broad categories- demographic and socio-economic variables. For the demographic variables, I control for the duration of the second birth interval, mother's birth cohort, and father's birth cohort. The socio-economic vector comprises of caste, religion, parental literacy, the standard of living, and place of residence for the household.

It is critical to account for the history of the birth spacing process (Heckman and Walker, 1990). The inter-birth interval between the first two births partly accounts for the unobserved parental behavior concerning birth spacing, coital frequency, contraceptive acceptance and usage, breastfeeding duration, and the mother's biological characteristics favoring childbirth. For instance, some related work has established the association between postpartum amenorrhea and birth spacing patterns (Popkin et al.,

[17] Median birth interval refers to the median for all births and not particularly for the third birth for these other developing countries

1993; Thapa et al., 1988; Winikoff, 1983).[18] Since I do not observe the span of post-partum amenorrhea in the data, the second birth interval gives a rough proxy for underlying characteristics influencing such factors.

Next, I incorporate mother's birth cohort in my estimation. Medical studies have shown variations in a woman's fecundity by age (van Noord-Zaadstra et al., 1991; Stovall et al., 1991). Consequently, I suspect younger mothers to face a higher risk of short birth intervals, as fertility is anticipated to decrease with age. I examine this hypothesis by studying the correlation between short birth intervals and mother's age. Figure 9.3 indicates that young mothers are prone to risky births in terms of shorter spacing. Such short spaced third births become less likely as I proceed towards higher age groups. Besides, the broad conclusion for the association between mother's age and the birth interval is similar for NFHS-1 and DLHS-2.

Amongst the socio-economic covariates, the place of residence is presumed to play a vital role in birth spacing behavior. Rural and urban mother's face varying customs and traditions associated with breastfeeding practices. Rural mothers are likely to breastfeed their infants longer than urban mothers. This duration, in turn, is closely related to postpartum amenorrhoea which is critical for inter-birth intervals (Hajian-Tilaki, 2002; Singh and Bhaduri, 1971). Moreover, young mothers in Indian villages face several taboos upon resumption of sexual activity before the child reaches a critical age (Fayehun et al., 2011). Figure 9.4 shows that shorter birth interval is more prevalent among urban than rural women in India.

Parental education and labor market participation are important behavioral determinants of child spacing. I do not observe the labor market outcome for mothers at the time of third birth. Nevertheless, educational attainment and labor market outcomes are closely associated. There are two opposing effects of education on birth spacing. First, being educated makes the mother more likely to be aware of the dire consequences of short spacing. But, mothers with higher education are more likely to closely space multiple births as they must forgo wages for staying out of the labor market. I use NFHS-1 for this section of the study. The level of education is low in the sample. Nearly 65% of mothers have no education, and another 16% have not completed primary school. However, the correlation between education and third birth is not monotonically increasing or decreasing. This non-monotonic relationship might be the manifestation of the two different mechanisms stated above at varying education levels (Figure 9.5).

Lastly, religion and caste of the household can explain child spacing behavior. For instance, existing literature has pointed out that Hindus practice abstinence on certain days of the month, something not obeyed by other religious groups. This practice might lead to marginally delayed births for Hindus than others. Besides religion, there exists some correlation between caste and a woman's bargaining power within a household (Srinivas, 1980; Rahman and Rao, 2004). Women who belong to scheduled castes/tribes face slightly lower constraints such as restrictions on working outside the household, which is often applicable to women in upper caste households.

3.2 How does son preference gets translated into fertility measures

Sibling sex composition is a valid instrument for birth spacing if there exists an underlying preference for sons. I provide some suggestive evidence of such choice by studying the fertility decisions of the

[18] A mother undergoes short-term infertility following birth due to breastfeeding which is referred to as postpartum amenorrhea.

households in my sample. The survey asks mothers on their preferred sex for the next child. Table 9.5 shows the distribution of their preferences by number and sex composition of living children. The desire for a boy is inversely related to the proportion of living sons. In a family with one child and that being a daughter, 66% desire the next child to be a son. This number steeply rises to 85% for mothers with two daughters and 91% for a family with three living daughters. I repeat the same analysis for NFHS-1(1992-1993) sample, summarized in the last two columns in Table 9.5. Mothers in this sample also show a similar desire for sons as suggested by the respondents in DLHS-2 (2002-2004). A comparison of the proportions from the two surveys conducted with a gap of ten years is that son preference has remained unchanged over the years.

Another significant indicator of son preference is differential stopping behavior (DSB). According to Clark, 2000, p95, para 3 "in DSB parents who have achieved their desired composition of children are more likely to stop child-bearing than parents who have not achieved their desired sex composition". Research on DSB has shown that girls usually have more siblings than boys and birth of a son is a major determinant of child stopping among families at different parities. Table 9.6 supports this hypothesis as a lower proportion of sons drives a higher probability of continued childbearing. For a household with two children, mothers with two sons are 14% less likely to continue childbearing relative to those with two daughters. Similarly, for a family with three children, the likelihood of continued childbearing is 87% for families with three daughters relative to 64% for those with three sons. Although, this analysis is suggestive of differential stopping behavior it should be interpreted with caution as not all mothers have completed childbearing in this sample. Table 9.7 examines the sex ratio at birth of the last child for women who have completed child-bearing. These mothers belong to the oldest age cohort at the time of the survey in 2002-2004.[19] The numbers indicate that mothers usually stop childbearing following the birth of a boy at every parity. Tables 9.6, 9.7 are on the same lines as Chaudhuri, 2012, p181,182,table 3,4.[20]

Additionally, there is much variation in son preference across India. Such variation originates from differences in cultures and traditions. This, in turn, influences the social and economic value of sons. Mothers belonging to patrilineal states of Madhya Pradesh and Rajasthan express a much higher desire for a boy child. On the contrary, mothers in a more progressive state such as Kerala show a much lower preference towards a son. Close to 48% of mothers want the sex of the next child to be a boy in Madhya Pradesh. This percentage drops to as low as 18% for Kerala. Another crucial observation is that preference for a girl is higher in the north-eastern states of India such as Meghalaya and Nagaland. These states are known for their matrilineal customs and practices. Such behavior indicates that parents desire a child of that gender which is of higher economic value (Table 9.8). Table 9.8 has been replicated from IIPS 2006,314, p65.[21]

Although sibling sex composition is a random event, it interacts with parental preferences. Mothers respond differently regarding their decision for the third birth after the revelation of the gender of the first two children. Other countries such as China and Taiwan, known for high son preference, also show a similar behavioral response to child gender (Tu, 1991; Coombs and Sun, 1978).

[19] This part of the analysis includes mothers in the age group of 40-44 at the time of the survey in 2002-2004.

[20] Chaudhuri 2012 analyzed son preference and its implications on fertility behavior for a different period in India using NFHS 2005-2006.

[21] The value of a child here refers to both differences in social and economic status created by birth of a child of a particular gender. This could be a son in Rajasthan and a daughter in Meghalaya.

4 Empirical Strategy

4.1 Identification strategy and empirical specification

Given the evidence implying a strong preference for sons in India, I propose the following model. The outcome of interest is the third birth interval B_i. I concentrate on the inter-birth interval as a measure of birth spacing that calculates the gap within the birth date of the second child and the third child in a family.[22] I start with a linear specification of log of birth interval on sibling composition of the first two births and a vector of other observable covariates. I relax the assumption of linearity in the second part of the analysis using hazard models.

$$B_i = \alpha_0 + X_i\beta + \gamma_1 d_{1i} + \gamma_2 d_{2i} + \gamma_3 d_{3i} + \epsilon_i \tag{1}$$

where

$$d_{1i} = \begin{cases} 1, & \text{if } BB_i = 1 \\ 0, & otherwise \end{cases}, d_{2i} = \begin{cases} 1, & \text{if } BG_i = 1 \\ 0, & otherwise \end{cases}, d_{3i} = \begin{cases} 1, & \text{if } GB_i = 1 \\ 0, & otherwise \end{cases}$$

The unit of observation for this regression is a household, and cross-sectional variation is the source of identification. There are four dummies for sibling sex composition of first two live births-BoyBoy(BB), BoyGirl(BG), GirlBoy(GB) and GirlGirl(GG). I include the dummies for all combinations except GG to avoid the dummy variable trap. Consequently, GG is the omitted/reference/benchmark category in all the specifications. The set of other covariates included in the specification are mother's birth cohort, father's birth cohort, mother's literacy, father's literacy, second birth interval, the standard of living index, a dummy representing rural or urban household, caste, and religion. Since these covariates have substantial explanatory power for birth interval, the inclusion of the observables improves the precision of the estimated coefficients.

The coefficients of interest are $\gamma_j, j \in \{1, 2, 3\}$. The identifying assumption depends on the independence of gender of births from the error term ϵ_i. This assumption as discussed above requires sex of a child to be a random event, which is likely to hold in the absence of gender screening. Even though the gender of a child is random, ex-post parents can manipulate the sex ratio of the total sibling size. Female infanticide (Sen, 1992) and the subsequent under-reporting of these female deaths in the retrospective health surveys is expected to pose a problem to the estimation.[23] An immediate check for such a threat to estimation is to investigate the distribution of observable characteristics by the gender of the first and second birth. Table 9.9 shows the distribution of place of residence, different religious groups and caste groups among families who had a son or daughter at the first and second birth. The reason for using these covariates is that there could be differences in the intensity of son preference by these sub-categories. For example, Hindus in India are likely to have a higher preference for sons than Christians (Clark, 2000). But, in my sample fraction of families who had a girl or a boy among the first or the second birth have a very similar distribution of religion.

Moreover, since cultural norms and traditions are likely to vary by place of residence being rural or urban, I examine differences by place of residence. Again the distribution is not statistically different at 5% level of significance. Lastly, I consider these differences by caste categories. For upper castes, I

[22] Inter-birth interval is the most commonly used measure of birth spacing used in existing literature; however, there are other medically suggested measures of spacing such as inter-pregnancy intervals.

[23] Female infanticide is the deliberate killing of new born daughters in societies with a strong son preference and high aversion towards daughters.

find a marginal difference, but it is more in favor of girl child rather than a boy. Scheduled caste and other backward caste show no differences. Scheduled Tribe is the only outlier which shows a marginally higher proportion in favor of boy child. However, the magnitude of this difference is tiny 0.7%, and scheduled tribes can be safely assumed to be highly unlikely to be exposed to any modern gender screening technology before upper, and other castes had access to the same technology. Additionally, as mentioned earlier women autonomy and bargaining power is relatively higher for women belonging to Scheduled Tribes rather than upper caste households, owing to higher seclusion of women belonging to upper castes. Hence, it is not likely that scheduled tribes were manipulating child's sex at birth even though they have lower son preference relative to other caste groups in India.

I interpret the coefficient γ_j relative to the omitted category of households with two daughters. Since parents are assumed to wait longer following the births of boys relative to girls, ex-ante the $\gamma_j s$ are expected to be positive ($\gamma_j > 0, j \in \{1, 2, 3\}$). I also expect parents to respond to third birth interval according to the proportion of sons amongst the first two children. Therefore, families with two sons are supposed to wait longer than families with one son ($\gamma_1 > \gamma_2, \gamma_1 > \gamma_2$). Lastly, if parents care relatively more about the younger child in the sibling pair, then I expect $\gamma_3 > \gamma_2$. Nevertheless, these coefficients can go in the opposite direction in the presence of eldest son preference ($\gamma_2 > \gamma_3$, Jayachandran and Pande 2017) . Else, considering the two influences are opposing each other, overall they can offset each other and $\gamma_2 = \gamma_3$.

4.2 Hazard models with fixed and time varying covariates

Although the linear model is a useful starting point for the birth interval analysis, it suffers from the following caveats. The inference for the linear model depends on the normality assumption, and the outcome variable violates this. The semi-parametric duration or survival model could be a better candidate in this situation. It does not require any distributional assumption on the duration till third birth unlike the linear specification.[24] Consequently, the existing literature on the duration of birth intervals profoundly makes use of the hazard models for their estimation (Newman, 1983; Newman and McCulloch, 1984; Trussell et al., 1985; Heckman and Walker, 1990; Nath et al., 1993). The most commonly used hazard model in the literature is the one proposed by Cox in his seminal 1972 article (Cox, 1972). The Cox model uses the order of third birth among families in the sample and estimates the duration till third birth using the ordering of birth interval.

Although the Cox model is more appropriate for time to an event analysis than a linear model, it has some limitations. The Cox model is a semi-parametric analysis; therefore, it does impose a specific functional form on the effect of explanatory variables. A fully non-parametric analysis can eliminate this functional form assumption as well. However, semi-parametric models deal with problems related to truncation or censoring more appropriately than non-parametric survival models.[25]

In duration analysis, the hazard rate is modeled instead of the probability of the time interval till the event occurs.[26] Therefore, instead of modeling the survival time which is time until households waited for third birth, I estimate the conditional probability of third birth occurring in period t as noted in

[24]One can think about the semi-parametric model as performing a binary analysis of the probability of failure at each point in time and then combining the results for all these periods without forcing any assumption on the distribution of third birth times.

[25]I discuss the sample selection and the probability of third birth at full length in the results section.

[26]The the event corresponds to the third birth in my analysis.

(Sueyoshi 1992, p28)

$$\begin{aligned}
\lambda(t) &= \lim_{\Delta \to 0} \frac{P(t \leq T \leq t + \Delta | T \geq t)}{\Delta} \\
\lambda(t, X(t), \beta, v) &= v\lambda(t, X(t), \beta) \\
&= v\lambda_0(t)\phi(X(t), \beta)
\end{aligned}$$

In this equation, $\lambda(t)$ corresponds to the instantaneous hazard rate. It depends on the baseline hazard denoted by $\lambda_0(t)$. $X(t)$ indicates the observables for households in period t which can be fixed or varying with time. Lastly, β is the vector of parameters to be estimated, and v is the error term. I use the same set of covariates as in the linear model. In addition to the third birth interval, sibling composition can also modify the probability of third birth itself.[27] Consequently, limiting the sample to families with at least three births by 1990 can result in sample selection. This constraint can incorporate more of those families who had fewer sons in first two births. I tackle this problem by expanding my sample to households that had at least two births by 1990. Additionally, I censor those observations that did not have the third birth by 1990. The censored model as noted in Cleves et al. 2008, p31

$$L\{\beta|(t_1, x_1), ..., (t_n, x_n)\} = \prod_{i=1}^{k} S(t_i|x_i, \beta)\lambda(t_i|x_i, \beta) \prod_{k+1}^{n} S(t_i|x_i, \beta)$$

In this equation, $S(t) = 1 - F(t)$ (Cleves et al. 2008, p7), where $F(t)$ is the cumulative distribution function. The survivor function $S(t)$ measures the probability that family i did not have the third birth by period t. For a censored household, I know that the family did not have a third birth by 1990. Accordingly, I replace $S * \lambda$ by just the survivor function for these households in the sample.

The underlying assumption needed for using the aforementioned right censored models is that the censoring of households is random and is not related to reasons for third birth. It will hold in my case if 1990 is an arbitrary year and it is not a determinant of the timing of third birth. This assumption is reasonable for this examination of third births. However, one shortcoming of the censored model is that it forces the parameters to be identical for the non-censored and the censored observations. It can be made more flexible by estimating two coefficient vectors for the two type of households and testing for the equality of these coefficients. This test will produce additional validation for the empirical specification.

5 Findings

5.1 Male bias and the probability of third birth

In this section, I replicate the first stage results of Angrist and Evans for the Indian context. Parental male bias can affect family decisions at two margins, the total number of children and the spacing between the children. I analyze the impact of sibling sex composition on the probability of third birth. I find that couples with a higher fraction of sons among the first two births are less likely to proceed to third birth. A simple comparison of means along with a logit analysis for the probability of third birth as shown below supports this hypothesis.

[27]Section 5.1 shows that households with a higher proportion of sons are less inclined to progress to third birth.

I report the estimates of the effect of male bias on the probability of third birth in Table 9.3. These estimates are like those reported in panel B of Table 3 in (Angrist and Evans, 1998). The second and the third column summarizes the fraction of third birth by gender composition of the first two births for mothers in DLHS-2 sample who experienced two or more births by 1990. Parents with a higher proportion of sons have a lower likelihood of third birth unadjusted for any covariates. The test of the difference in fractions for these sub-groups have been shown in panel B. The differences in the fractions for couples who opt for third birth is statistically different for groups with different proportion of sons among the first two children.

The critical difference between these results and those observed in (Angrist and Evans, 1998) emanates from underlying differences in parental preferences. Research has shown that parents prefer mixed sibling composition in the United States, families with same-sex composition are more likely to have a third child. On the contrary parental preference for sons gets reflected in this preliminary analysis for India where parents with a higher proportion of sons are less likely to go for a third birth. The other point of divergence is that the proportion of families who have third birth is much higher in India than the sample in Angrist and Evans. The proportion of families that have more than two children ranges between 0.37 to 0.43 for the American households as compared to 0.66 to 0.75 for the Indian sample.

Although the difference between couples with two sons and two daughters is high in the DLHS-2 sample, some of these differences could be the result of the differences in other observable and unobservable variables. Therefore, I estimate the relationship between these variables controlling for other covariates. The regression that I estimate is given as follows:

$$y_i = \begin{cases} 1, & \text{if } y_i^* = X_i\beta + \kappa_1 BB_i + \kappa_2 BG_i + \kappa_3 GB_i + u_i \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

The binary outcome variable y_i is whether mother i had the third birth by the year 1990. The dummy BB_i, BG_i explain the latent variable y_i^* which considers the sibling sex composition of the first two births.[28] Table 9.10 shows that sibling sex composition of the first two births affects the probability of third birth. In both the specifications, with and without additional observable covariates, the probability of having the third birth decreases with an increase in the proportion of sons among the first two births. The omitted category is families who have two daughters. The marginal effects of third birth by sibling sex composition show that families with two daughters have an 8% lower likelihood of third birth relative to families with two sons (Table 9.10).

5.2 Parametric model

I first begin by modeling the third birth interval as a linear process. Taking cues from the medical literature, I restrict the birth interval to start from 9 months. In an OLS regression such as given below ϵ_i follows a normal distribution which has the entire real line as its domain. On the contrary, birth intervals are always positive. In order to address this concern, I use a log-linear regression model where the outcome variable has a positive domain.

Table 9.11 shows the coefficients of interest γ_1 and γ_2 conditional on a set of covariates. In the model,

[28] In the logit regression the dummy BB takes a value 1 if first two children are sons and 0 if first two children are daughters; BG takes a value 1 if the couples had one daughter and one son and 0 if first two births were daughters; hence, the omitted category is two daughters.

Table 9.11 I generate dummies for the sibling composition of first two births on the same lines as (Nath and Land, 1994). The reference category is families with two surviving daughters. Since this is a log-linear model of the third birth interval, the interpretation of the coefficients is in percentages. Families with two surviving sons are likely to have birth intervals longer by 3.0% relative to the reference category of families with two surviving daughters. Besides, I observe a slightly longer third birth interval for families with one son, particularly for families with a male second child. These estimates support my hypothesis that parental investment is directly related to the proportion of sons among the first two births.

Although the log-linear regression is suggestive, the estimates should be interpreted with caution. For any inference, the errors in the above log-linear regression to have a normal distribution. I perform the skewness-kurtosis diagnostic test for testing the normality assumption on the residuals generated from the above two models (D'agostino et al., 1990; D'Agostino and Belanger, 1992; Royston, 1992). Table 10.1 shows that I can reject the skewness-kurtosis tests of normality for the residuals from the two models. This result is in line with the distribution of the log of third birth interval plotted in figure which provides visual evidence of a violation of the normality assumption. The hazard model makes a more appropriate substitution for the normality assumption used in the least squares estimation.

5.3 Non-parametric tests for survival analysis

5.3.1 Wilcoxon test

Wilcoxon test is a non-parametric analysis which examines the difference in average survival time to third birth for different subgroups (Breslow, 1970; Gehan, 1965). I create these groups by characteristics of families observed at baseline period. For my model, this test generates a contingency table consisting of the number of families with third birth and hence the proportion at risk of third birth for every period. It then computes and compares the expected survival times for each of the groups for every period. The Wilcoxon test statistic is a weighted average of the expected difference in survival times to the third birth. There are multiple variations of this rank test, and the critical difference lies in the weighting scheme used for the test statistic. Wilcoxon statistic is more suited to my data-set as the weights decrease as one moves closer to the end point in the duration analysis when fewer and fewer families are at risk of third birth. I prefer this test over another commonly used test statistic, the log-rank test as there all the summary tables are equally weighted. Table 9.12 suggests differences in survival times by socio-economic and demographic characteristics.[29]

[29] The following two equations refer to equation 8.3 and 8.4 in Cleves et al. 2008, p124. "For every time period the Wilcoxon test statistic combines the following two equations

$$u' = \sum_{j=1}^{k} W(t_j)(d_{1j} - E_{1j}, ..., d_{rj} - E_{rj})$$

$$V_{il} = \sum_{j=1}^{k} \frac{W^2(t_j) n_{ij} d_j (n_j - d_j)}{n_j(n_j - 1)} (\delta_{il} - \frac{n_{ij}}{n_j})$$

. Here, d_{ij} and E_{ij} refer to number who had third birth and expected third birth for individuals belonging to subgroup i in period t_j respectively. Lastly, δ_{il} is 1 if i and l are same and 0 otherwise".

5.4 Hazard model

5.4.1 Baseline model with fixed covariates

I start duration analysis with the basic Cox model with fixed covariates.[30] The hazard rate simplifies to the following equation as shown in Sueyoshi 1992, p28

$$\begin{aligned}\lambda(t, X, \beta, v) &= v\lambda(t, X, \beta)\\ &= v\lambda_0(t)\phi(X, \beta)\end{aligned}$$

This model measures the covariates at the baseline, and these variables are assumed to remain fixed throughout the wait time to third birth. A popular function ϕ accepted in the current literature on duration models is the exponential function as this makes sure that the hazard rate is non-negative. Consequently, I estimate the following specification

$$\lambda_i(t) = \lambda_0(t)e^{\alpha_0 + X\beta + \delta_1 d_{1i} + \delta_2 d_{2i} + \delta_3 d_{3i}}$$

Since the covariates are time-invariant, the ratio of hazards for two individuals j(BB) and k(BG) at a period $t \in \{0, ..., T\}$ as suggested in Cleves et al. 2008, p22 is given by[31]

$$\frac{\lambda_j(t)}{\lambda_k(t)} = \frac{\lambda_0 e^{\beta_0 + X_j\beta + \delta_1}}{\lambda_0 e^{\beta_0 + X_k\beta + \delta_2}}$$

I present the estimated coefficients of this empirical model in Table 9.13. Column 2 gives the coefficient estimates, and column 4 shows the exponentiated coefficients for ease of interpretation. For the primary explanatory variable of interest, a hazard ratio of less than one corresponds to a lower risk of having the third birth than the reference category of households with two daughters. As anticipated, $e^{\delta_j} < 1, j \in \{1, 2, 3\}$. Consequently, families with one or two sons face 5 to 9% lower instantaneous hazard respectively than the reference category of households.

Cox model allows flexibility to the baseline hazard as no ex-ante assumption on the shape of λ_0 is imposed. Using the coefficients of this model, I plot the instantaneous hazard rate in figure 9.6. The instantaneous hazard rate has been estimated by first calculating the period by period hazard contribution. Once, the hazard contribution for every period has been estimated; I use the Gaussian kernel to smooth the curve. It shows that mothers who had two boys face a lower hazard of third birth in every period in $\{0, .., T\}$ relative to families with two alive daughters. Besides, I observe that the baseline hazard function is an increasing function of time and the gradient of the curve is not constant. It is plausible that mothers in the sample do prefer some spacing between the second and third child, but they do not want to delay the third birth excessively. Therefore, the gradient of the curve becomes steeper as one moves to the right on the time axis.

In order to compare these results with related work, I translate the hazard ratios into median survival time. Median survival time in duration analysis refers to the median time till the parents choose to wait for the third birth. In line with economic theory, the median survival time increases with the number of sons in the first two births. Families with two boys wait one month longer than families with two daughters. Given, the average third birth interval for Indian mothers is 30 months, a difference of one month translates into a 3% shorter interval for households with two daughters. Although there is not

[30] This version of the Cox model is also commonly known as the proportional hazards model.

[31] The abbreviations in the parentheses correspond to the sibling composition of the first two births

precisely comparable evidence of sibling composition on third birth, a few studies have examined the impact of gender of previous birth on the subsequent birth interval. Birth intervals were reported to be 2.3% shorter for births preceded by girls in Bhalotra and Van Soest 2008 for a sample of mothers from Uttar Pradesh, India.

5.4.2 Main model with fixed and time varying covariates

In this section, I extend the preliminary examination by relaxing the assumptions of the basic Cox model. Instead of the fixed covariate model, I allow some covariates to change over time. Mother's birth cohort entered as a fixed variable earlier. However, mother's fecundity varies over time, and hence it is appropriate to incorporate it as a time-varying covariate in the hazard function. I use a variant of the link test to check the proper functional form for the mother's birth cohort-fixed or time-varying. I fit the baseline model with an additional interaction of mother's birth cohort with time as follows

$$\lambda_i(t) = \lambda_0(t)e^{\alpha_0+X\beta+\delta_1 d_{1i}+\delta_2 d_{2i}+\delta_3 d_{3i}+\beta_1(mother'sage*t)} \tag{3}$$

After fitting this model, I test whether $\beta_1 = 0$. Since the test rejects the null of $\beta_1 = 0$, I include mother's age as a time-varying covariate. The coefficients for sibling sex composition remain unchanged (Table 9.14). I apply the estimated coefficients to predict the survival curve for families with two daughters and two sons. Having more sons among the first two births prolongs the survival time for those families (Figure 9.7).

5.4.3 Estimation with additional covariates

I use the NFHS 1 survey to estimate the third birth interval with additional covariates. First, I include a variable for household economic status. The ideal measure for the financial condition is the one that measures income at the time of third birth. However, these household surveys do not contain such detailed income data. Consequently, it is more relevant to include the economic status variable for NFHS-1 than DLHS-2 as the data on economic status was collected in 1992-1993 for NFHS-1 as opposed to 2002-2003 for DLHS-2.

Second, DLHS-2 does not collect data on husband's additional characteristics as part of the eligible woman survey. Surveyors collect other data on husbands as part of the husband's questionnaire, but they administer this questionnaire on a much smaller sample than the number of mothers. However, NFHS includes detailed data on father's characteristics as part of the woman's survey. Therefore, I control for father's birth cohort in addition to all the covariates used earlier.

Lastly, I have better quality information on educational attainment in NFHS-1. In the earlier specification, I use a crude measure of education that captures whether or not the parents are capable of reading and writing. Here, the survey question captures the education attainment as completion of various schooling levels. Mainly, I divide schooling into six categories-no education, incomplete primary school, complete primary school, incomplete secondary, complete secondary and higher.

Table 9.15 summarizes the results with the additional set of covariates. Father's age is inversely related to the hazard of third birth. Households with older men are less prone to third birth relative to families with younger men. Economic status of the house is computed using a measure of household possession of durables as the survey did not have a direct measure of financial condition.[32] The better the

[32] The economic status index is a simple average of the possession of a group of indicator variables-household has electricity, radio, television, refrigerator, bicycle, motorcycle and car.

economic condition, the higher is the relative hazard of third birth. The economic intuition behind this observation could be that women belonging to higher income households face more seclusion regarding practices such as purdah relative to women in low-income families. Finally, father's education has a limited impact on the third birth interval, in line with related work. Meanwhile, mothers with secondary or higher education face a lower hazard of third birth relative to the reference category of mothers with no schooling.

5.5 Sibling composition and risky births

Sibling sex composition changes the average third birth interval by about a month. This difference in spacing between families with two daughters relative to families with two sons is significant in the statistical sense but is of limited economic importance for child outcomes. Nevertheless, for 17% of these births, the interval was shorter than 18 months. As stated earlier, this cut-off of 18 months is a critical determinant of child mortality. The proportion of boys amongst the first two births lowers the unconditional probability of these risky births (Table 9.9). Besides, a percentile distribution of third birth intervals by sibling sex composition reveals that families with two daughters have a consistently lower interval for every given percentile below 50 than households with two sons (Figure 9.8). Given this observation, I estimate the following linear probability model

$$z_i = \begin{cases} 1, & \text{if } z_i^* = X_i\beta_z + \zeta_1 BB_i + \zeta_2 BG_i + \zeta_3 GB_i + \eta_i \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

In the equation above, the outcome variable is a binary which takes the value 1 if the third birth interval is less than 18 months and 0 otherwise. The explanatory variables of interest are those corresponding to the sibling composition dummies. The identification of the $\zeta_j's$ rests the assumption that sibling composition is independent of η_i. Ex-ante given parents underlying son preference I expect $\zeta_j < 0, j \in \{1, 2, 3\}$. These negative coefficients indicate that the higher the proportion of sons among the first two births, lower is the likelihood of riskier births. Besides, assuming that parents care and attention on first two children increases with the proportion of sons, $|\zeta_1| > |\zeta_2|, |\zeta_1| > |\zeta_3|$. Lastly, depending on parents concern for the older or younger son in the sibling pair, either $|\zeta_2| \geq |\zeta_3|$ or $|\zeta_2| \leq |\zeta_3|$.

Table 9.16 shows the results of this estimation. The coefficients on sibling sex composition are in line with the expectation ($\zeta_j's < 0$). Households with two sons have a 2.4% lower likelihood of third birth within the first 18 months after second birth relative to the reference category of couples with two daughters. The corresponding probability for families with one son is slightly more than 1%. Finally, I find $|\zeta_3| > |\zeta_2|$, which provides some evidence that parents might care somewhat more about the younger child in the sibling pair. This variation might be due to the reason that the second child faces additional risk to illness due to a shorter subsequent birth interval.

6 Robustness Checks

6.1 Recall bias

The primary dataset for my analysis is DLHS-2. Surveyors interviewed the mothers between 2002 and 2004. However, the identifying assumption for sibling composition holds for the restricted sample of births before 1990. Although, the data on retrospective birth history is mostly complete and does not show irregularities in reporting by the gender of the child. Nonetheless, given the time gap between the date of interview and birth dates I want to eliminate any possibility of recall bias by comparing the results from another data-set. NFHS-1 is unlikely to suffer from such bias as the survey happened in 1992-1993. I re-estimate equation (1) using NFHS-1 and the results are given in Table 9.15. Sibling sex composition has a very similar effect on third birth interval as noted in Table 9.15.

Furthermore, the probability that the third birth interval is less than 18 months changes by sibling composition. Figure 9.9 displays the coefficients from the estimation of equation 4 using NFHS-1. Having two sons among the first two births lowers the probability of third birth interval being shorter than 18 months by 2.9% relative to those couples who had two daughters. Besides, even one son among first two births lowers this likelihood considerably.

6.2 Sensitivity analysis

I use the sensitivity analysis in Altonji et al. 2005 to assess my empirical model. I do this analysis for equation 4.[33] The authors build their test on the assumption that selection on unobservables determines selection on observables. I state the assumption Altonji et al. 2005, p175 below

$$\frac{E(\epsilon|BB=1)-E(\epsilon|BB=0)}{Var(\epsilon)}=\frac{E(X'\gamma|BB=1)-E(X'\gamma|BB=0)}{Var(X'\gamma)}$$

First, I begin with the preliminary examination of how much does the coefficient $\zeta_j s$ in equation 4 change once I introduce the set of controls. Figure **??** shows that there is hardly any change in the estimated coefficients once I add the set of controls in the regression specification. This test is also an indirect test for the exogeneity of the primary explanatory variable of interest in the birth spacing equation.

Next, I use the assumption mentioned above to compute the extent of selection bias required to explain the sibling composition effect. According to the authors, if this ratio is well above 1.5, it is an indication of the presence of a real sibling sex composition effect, and it is not entirely driven by selection in unobservables. Table 9.17 provides the implied ratio for equation 4. The implied ratio for sibling sex composition is 21.0, mainly because the observables cannot predict sibling sex composition. This test provides additional evidence that the results are not driven by selection in observables.

6.3 Regional variation in male bias

Son preference in India has its roots in cultural norms and traditions which vary considerably across the country. There are several states such as Madhya Pradesh, Bihar and Rajasthan where patriarchal ways are more widespread than other parts of the country. Evidence of such ways gets reflected in various sociological and demographic indicators. For instance, nearly 48% of the mothers want the next child to be a boy, 10 percentage points above the all-India average of 38%. On the contrary, southern India has shown notable progress in female education, labor market participation and other indicators of female autonomy. Especially, Kerala is one of the most successful states when it comes to these indicators. Given the apparent contrast in the underlying cultures in Madhya Pradesh and Kerala, I speculate that

[33] Figure **??** are the coefficients from the regression of probability of risky birth on sibling composition of the first two births using NFHS-1.

sibling sex composition is expected to have a more significant impact in Madhya Pradesh than Kerala. I estimate the following two equations to test the conjecture-

$$z_{si} = \begin{cases} 1, & \text{if } z^*_{si} = X_{si}\beta_{sz} + \zeta_{s1}BB_{si} + \zeta_{s2}BG_{si} + \zeta_{s3}GB_{si} + \eta_{si} \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where the variables are the same as defined in equation 4 except that $s \in \{M, K\}$ depending on samples belonging to Madhya Pradesh or Kerala. Ex-ante I expect $|\zeta_{Mj}| > |\zeta_{Kj}|, j \in \{\ 1, 2, 3\}$. Also, given that the outcome is the probability of having the third birth within first 18 months of second birth, all $\zeta_{Mj}s$ are expected to be negative. For Kerala, I expect the $\zeta_{Kj}s$ to be close to zero.

Figure 9.10 shows the estimates of equation 5. The confidence intervals for the point estimates for Madhya Pradesh and Kerala are much larger than the All-India estimate. These large confidence intervals are due to small sample size in state-wise regressions. However, some broad conclusions can be drawn from these results even though the coefficients are not precisely estimated. First, coefficients for Kerala turn out to be either marginally negative or positive for families with either one son or daughter. These signs provide support to my hypothesis that son preference is remarkably low in Kerala. Moreover, I observe the $\zeta_{Mj}s$ to be highly negative for Madhya Pradesh than the all-India average. For instance, couples who had two sons had a 6.2% lower likelihood of risky birth than families with two daughters. These results are suggestive, but one should not read too much into this as the regressions have low power.

7 Discussion

7.1 Birth spacing and child mortality

Research in medicine, economics, and sociology has highlighted several mechanisms that drive this relationship. First, if a mother undergoes two births with a spacing shorter than 18 months, the subsequent child is prone to suffer from "intrauterine growth retardation" (Warkany et al., 1961; Chiswick, 1985; Vandenbosche and Kirchner, 1998) . This aberration occurs due to maternal depletion as mothers do not get enough time to recuperate from the previous pregnancy before the onset of the next one. Previous work has documented that maternal depletion can heighten the danger of low birth weights and infant mortality.

The other channel through which narrow spacing of births can affect child health outcomes within a household is sibling rivalry. A pair of closely spaced siblings often ends up competing for limited household resources such as parental care time, breastfeeding duration, nutrition and expenditure on health care services (Alam, 1995a). The final mechanism which directs the mortality-spacing connection is the higher possibility of transmission of contagious diseases to the younger child of the pair in the presence of a closely spaced older sibling. Aaby et al. 1984; Aaby 1988 suggest that one of the chief ways of transmission of measles is through sibling communication within the same household and not nutrient deficiency among children. Further, Alam 1995b suggest that the presence of older siblings in similar age categories raises the chance of diarrhea for younger one in the sibling pair.

Having ascertained the underlying mechanisms that amplify the risk of mortality following short birth intervals, I present the distribution of various measures of mortality in DLHS-2 for the third birth and its association with the preceding birth interval. Among the 78,000 third-order births recorded in DLHS-2 on or before 1990, 11.5% of those children have been reported to be dead at the time of the survey in 2002-2004. More than 90% of these reported deaths occurred when children were below the age of five. For NFHS-1, a similar fraction (11%) of live births were reported to be dead by the time of survey in 1992-1993.

An analysis of the under-five age deaths among children shows that there is a high incidence of neonatal deaths. Neonatal mortality refers to death within the first 28 days since the day of birth. I use monthly data to calculate mortality by age category and so any incidence of death before the completion of the first month has been classified as a neonatal death.[34] A comparison of incidence of mortality per 1,000 live births between NFHS-1 and DLHS-2 suggests similar estimates for neonatal deaths. However, the incidence of post-neonatal mortality and hence, infant mortality is higher in the NFHS sample compared to DLHS-2.[35,36]

Studying the distribution of third births by the third birth interval, approximately 17-18% occurred within one and half years of the second birth. For this count, I drop observations with a birth interval of lower than nine months as that is expected to incorporate pre-mature births.[37] I eliminate pre-mature births because there can be critical biological deficiencies in those mothers which results in pre-mature births. Figure 9.11 presents the correlation between spacing and mortality for the third birth. The incidence of death is systematically higher for births spaced below 18 months for all three mortality measures. Moreover, the sharp drop in gradient around the 18-month cut-off is evident in both DLHS-2 and NFHS-1. The decline in slope is highest for mortality before the child finishes the first month.

Next, I measure the odds ratio for third births by the previous birth interval.[38] I do this examination for children belonging to households for whom the first two children survived. I require this restriction for two reasons. First, it is to maintain the same sample for the sibling composition and mortality analysis. Second, this constraint separates mothers that might have a higher plausibility of child death due to unobservable biological characteristics. If the second child died due to mother's health deficiencies, then the third birth interval is automatically shorter. Consequently, the mother's unobservable deficiency is driving both the third child mortality and third birth length. Figure 9.12 provides evidence of the variation in third birth intervals by the survival status of the second child.

Table 9.19 summarizes the odds ratio of several measures of mortality by the length of the third birth interval. The omitted category for the birth spacing is interval longer than 18 months. I also control for observable characteristics such as mother's birth cohort, mother's and father's literacy, place of residence and standard of living. The odds ratio shows that the risk of mortality amongst children is more than twice if the previous interval is shorter than 18 months. This ratio is comparable to similar risk ratio measures suggested in related work. Boerma and Bicego 1992 find that birth intervals shorter than 18 months more than doubles the relative risks of neonatal mortality for most of the sub-Saharan and north African countries in their sample. These include Burundi, Ghana, Mali, Senegal, Togo,

[34] Bhalotra and Van Soest 2008 use a similar approximation in their analysis as there is heaping at one month in their sample for Uttar Pradesh, India.

[35] Post-neonatal mortality is death between 1 and 12 months of birth.

[36] Infant mortality is total likelihood of dying within the first year of birth; it is the sum of neo-natal and post-neo natal mortality.

[37] Restricting the birth interval at nine months is a conventional cut-off used in the infant mortality related literature.

[38] The odds ratio of higher than 1 suggests that shorter birth intervals increase the risk of mortality; aforementioned is suggestive evidence based on correlations and does not necessarily imply a causal relationship between short birth intervals and infant deaths.

Uganda, Egypt, Morocco and Tunisia.

Linking the odds ratio and the results from quantile treatment effects, I present a crude approximation of the number of missing children. The probability of the incidence of infant mortality is 8% higher for families with a third birth interval of shorter than 18 months. I use the infant mortality in 2016 as the baseline for my estimates. IMR was approximately 34 per 1,000 live births in 2016. Moreover, the number of live births in India per year is 49481 per day. This number implies a total of 18,060,565 births per year. I use the following formula for my estimate of additional deaths per year due to a shorter birth interval.

$$
\begin{aligned}
Additionaldeathsperyear &= IMR * (Increaseinprobabilityfor < 18) * (Livebirthseveryyear) \\
Additionaldeathsperyear &= \frac{34}{1000} * \frac{8}{100} * 18,060,565 \\
&= 61,405
\end{aligned}
$$

As, suggested in the estimates of sibling composition on the birth interval, families with two sons are 2.5% less likely to have risky births. This difference ends in 1,535 fewer infant deaths in India in a year.

8 Conclusion

In India, son preference has its roots in cultural beliefs and traditions. Parents sex preference gets translated into differential gender allocation in various household resources. Consequently, sibling sex composition affects different components of third birth decisions. The likelihood of third birth is directly related to the number of daughters among the first two births. According to my evaluation, couples with two girls have a probability of 0.76 of third birth relative to 0.69 and 0.66 respectively for families with one or two sons.

Since more than two-thirds of mothers in the sample had at least three births, I delve further and examine the influence of sibling composition on the third birth interval. Overall, the median survival time is about a month longer for families with two sons compared to households with two daughters. This result is robust to linear and non-linear specifications of the empirical model. Besides, the magnitude of the difference is comparable to related work on birth spacing in Uttar Pradesh where birth intervals preceded by a daughter are about 2.3% longer than those preceded by a son.

A variation of one month in the birth interval is not expected to have a considerable influence on child health outcomes. However, sibling composition does affect the left tail of the birth spacing distribution. Having two sons reduces the likelihood of third birth within 18 months of prior birth by 2.5 to 3% relative to families with two daughters. This number, when combined with the infant mortality spacing relationship, suggests that about 1,500 infant deaths in India can be attributed to a higher proportion of daughters among the previous births.

Although I present useful insights into the third birth decision for families in India, I want to point towards some limitations. First, the identification strategy works in the absence of gender screening. Sharp changes in economic growth and development have taken place in India post-1990s. But I cannot study birth spacing, and health outcomes post the 1990s as it coincided with the widespread use of ultrasound technology and it poses a threat to identification. Next, I concentrate on the short-term implications of spacing on child outcomes. I do not provide links to long-term effects such as child test scores and later labor market returns. There is some previous work that establishes the long-term implications of birth spacing[39] and that could be an interesting topic to explore for future work on India.

This analysis reveals that the proportion of sons among the first two births does affect different dimensions of third birth decisions for parents in India. Therefore, sibling composition is a credible candidate for an instrument for birth spacing in the birth interval-child outcome relationship. This instrument provides a quasi-experimental set-up to disentangle the causal effect of birth spacing on child mortality and various other anthropometric measures. Besides, it is also possible to identify the characteristics of the compliers in my sample. These complier characteristics could be very informative for any targeted policy intervention for improving child health outcomes in India.

[39] For example, Buckles and Munnich 2012 estimate the impact of birth spacing on test scores for siblings within a household; authors find a significant impact on older siblings and muted effect on younger ones using NLSY-79 data-set for the United States.

9 Tables and Figures

Table 9.1: Summary Statistics of Eligible Women Sample in DLHS-2

Variable	Place of residence		
	Rural	Urban	Total
	Mean (or proportion)	Mean (or proportion)	Mean (or proportion)
Fertility behavior of households			
Children ever born	2.83	2.51	2.73
Surviving children	2.52	2.33	2.46
Surviving sons	1.32	1.22	1.29
Surviving daughters	1.20	1.11	1.17
Proportion of sons	0.54	0.54	0.54
Socio-economic characteristics of households			
Age of mother	29.47	30.80	29.89
Religion			
Hindu	0.78	0.75	0.77
Muslim	0.09	0.16	0.11
Christian	0.07	0.05	0.06
Sikh	0.03	0.02	0.03
Others	0.03	0.03	0.03
Caste of household head			
Scheduled Caste	0.18	0.14	0.17
Scheduled Tribe	0.19	0.08	0.15
Other Backward Class	0.38	0.38	0.38
Other	0.24	0.39	0.29
Do not know	0.01	0.01	0.01
Mother can read and write			
Yes	0.43	0.72	0.53
No	0.57	0.28	0.48
Father can read and write			
Yes	0.69	0.86	0.74
No	0.31	0.14	0.26
Household standard of living index			
Low	0.57	0.14	0.43
Medium	0.31	0.33	0.32
High	0.12	0.53	0.25
Sample size	343,342	160,994	504,336

.

Table 9.2: Comparison of fertility measures in India with the sample of American households in Angrist and Evans (1998)

Sample	Means and percent
A.American household sample used in Angrist and Evans (1998)	
	Women aged 21-35 in PUMS 1990
Children ever born	2.50
At least 3 children	0.375
Boy first	0.512
Boy second	0.511
Two boys	0.264
Two girls	0.241
B.Indian household samples from DLHS-II(2002-2004)	
	Women aged 21-32 in 1990 in DLHS II
Children born (alive)	3.35
At least 3 children	0.670
Boy first	0.513
Boy second	0.509
Two boys	0.267
Two girls	0.246

Table 9.3: Likelihood of third birth by the sex composition of first two births for families with two or more children by 1990

	DLHS-2 (75,863)
Gender composition of the first two births	Proportion that had third birth
(a)Both sons	0.665
(b)Son Daughter	0.691
(c)Daughter Son	0.691
(d)Both daughters	0.747
	Difference in proportions
(a)-(d)	-0.082***
	[0.004]
(b)-(d)	-0.056***
	[0.004]
(c)-(d)	-0.056***
	[0.004]

.

Table 9.4: Third birth intervals by sex composition of first two births

	Third birth interval in months				
Sex composition of first two births	Mean	<18	19-24	25-36	>36
(a)Both daughters alive	29.18	0.174	0.249	0.362	0.215
(b)Daughter and son alive	29.87	0.162	0.244	0.364	0.229
(c)Both sons alive	30.39	0.149	0.239	0.367	0.244
	Difference in means/proportions				
(a)-(b)	-0.68***	0.012***	0.005	-0.001	-0.015***
	[0.145]	[0.004]	[0.004]	[0.005]	[0.004]
(a)-(c)	-1.20***	0.024***	0.010*	-0.005	-0.028***
	[0.164]	[0.004]	[0.005]	[0.005]	[0.005]
(b)-(c)	-0.52***	0.012***	0.005	-0.003	-0.014***
	[0.145]	[0.004]	[0.004]	[0.005]	[0.004]

Table 9.5: Desired sex for additional child by number and sex composition of living children

	Desired sex preference for next child			
	DLHS-2		NFHS-1	
	Son	Daughter	Son	Daughter
One child	38.7	14.2	45.6	18.1
One daughter	65.9	1.2	76.6	0.9
One son	12.8	26.6	16.3	34.4
Two children	50.2	12.4	60.4	16.0
Two daughters	85.3	0.9	93.4	0.4
One son, one daughter	43.2	2.2	59.4	2.4
Two sons	7.0	54.2	5.9	70.9
Three children	55.6	9.3	69.5	14.4
Three daughters	90.8	0.7	94.8	0.6
One son, two daughters	65.2	1.3	83.6	1.0
Two sons, one daughter	18.5	12.9	32.2	28.4
Three sons	6.3	59.1	6.0	82.2

Notes:The percentages do not add up to 100 as there are two more categories-does not matter and up to God in the survey questionnaires..

Table 9.6: Percentage of Women who Continued Child Bearing at Different Parities

Parity/No. of sons	% who stopped childbearing	% who continued childbearing	Total
Parity one	4.6	95.4	
No son	4.1	95.9	
One son	5.1	94.9	
No. of cases	7,001	144,334	151,335
Parity two	19.9	80.1	
No son	11.0	89.0	
One son	21.2	78.8	
Two sons	25.0	75.0	
No. of cases	28,782	115,548	144,330
Parity three	32.1	67.9	
No son	13.5	86.5	
One son	28.0	72.0	
Two sons	40.5	59.5	
Three sons	36.1	63.9	
No. of cases	37,145	78,402	115,547

Notes:Analysis here includes women who belonged in age group 35-44 at the time of survey in 2002-2004 .

Table 9.7: Sex Ratio at Birth, According to Whether Women Stopped Childbearing at a Given Parity

Parity	Sex ratio by last born child	
	Stopped childbearing	Continued childbearing
One	154	112
Two	161	102
Three	157	96

Notes:The sample for this analysis includes women in the age group 40-44 who had completed childbearing.

Table 9.8: Variation in sex preference across states in India

State	Desire for next child (sex) Boy	Girl	No. of cases
Andhra Pradesh	23.4	13.3	4245
Arunachal Pradesh	42.2	16.0	3348
Assam	24.2	11.2	3101
Bihar	56.3	6.3	11595
Chhatisgarh	40.9	9.6	3804
Delhi	18.7	6.3	870
Goa	28.0	11.7	368
Gujarat	40.7	8.3	4217
Haryana	37.9	5.8	3421
Himachal Pradesh	34.8	8.1	1181
Jammu & Kashmir	25.4	3.9	2185
Jharkhand	50.4	8.5	4925
Karnataka	29.1	11.0	5404
Kerala	18.1	14.3	1904
Madhya Pradesh	47.6	8.9	9284
Maharashtra	35.2	8.8	5674
Manipur	38.5	18.0	2523
Meghalaya	6.2	10.2	2397
Mizoram	28.1	19.8	2084
Nagaland	13.1	8.2	1628
Orissa	46.0	8.1	5803
Punjab	39.5	3.9	2025
Rajasthan	57.5	7.7	8812
Sikkim	24.6	11.3	459
Tamil Nadu	20.3	11.3	5465
Tripura	37.0	18.9	519
Uttar Pradesh	36.1	6.3	18530
Uttaranchal	35.5	5.0	2140
West Bengal	34.5	13.1	1777
Union Territory			
A & N islands	12.4	9.0	323
Chandigarh	24.1	7.1	112
Dadra & Nagar Haveli	29.9	14.5	234
Daman & Diu	25.1	9.2	347
Lakshadweep	10.4	13.0	470
Pondicherry	22.1	18.5	788
Total	38.3	9.1	121962

Table 9.9: Characteristics of families and gender of first two births

	Proportion First child		Diff	Proportion Second child		Diff
	Boyfirst	Girlfirst		Boysecond	Girlsecond	
Religion						
Hindu	0.775	0.778	-0.003	0.775	0.777	-0.002
			[0.004]			[0.004]
Muslim	0.116	0.112	0.004	0.112	0.116	0.004
			[0.003]			[0.003]
Christian	0.062	0.062	-0.000	0.062	0.062	-0.000
			[0.002]			[0.002]
Sikh	0.027	0.031	-0.004***	0.028	0.030	-0.002
			[0.002]			[0.002]
Others	0.019	0.018	-0.001	0.018	0.019	0.000
			[0.001]			[0.001]
Place of residence						
Rural	0.690	0.686	-0.005	0.692	0.685	-0.006*
			[0.004]			[0.003]
Caste						
Scheduled Caste	0.184	0.186	-0.002	0.187	0.184	0.003
			[0.003]			[0.003]
Scheduled Tribe	0.157	0.150	0.007***	0.157	0.150	0.007***
			[0.003]			[0.003]
Other Backward Caste	0.382	0.382	0.000	0.382	0.382	-0.000
			[0.004]			[0.004]
Upper Caste	0.266	0.272	-0.006*	0.265	0.274	-0.009***
			[0.003]			[0.003]

Table 9.10: Logit estimates of impact of sibling composition of first two births on the probability of third birth

Sex composition of first two births	Married women	
	(1)	(2)
Daughter Son	-0.056***	-0.055***
	[0.004]	[0.004]
Son Daughter	-0.056***	-0.057***
	[0.004]	[0.004]
Both Sons	-0.083***	-0.087***
	[0.004]	[0.004]
With other covariates	no	yes
Observations	119629	119629

Notes:a.Standard errors are in square brackets and $*p < 0.1, **p < 0.05, ***p < 0.01$.

b.The omitted category is two daughters in the above specifications.

c.The other covarites included in column (2) are mother's age at the time of survey, mother's age at first birth, whether mother can read and write, caste and religion. These covariates are in line with the covariates uses in Angrist and Evans (1998).

d.This analysis uses the DLHS-2 sample used in Table **??**.

Table 9.11: Log linear regression of third birth interval on sex composition of first two births

Variable	(1)
Sex composition of first two births if both children are alive[b]	
Daughter and Son	0.019***
	[0.005]
Son and Daughter	0.013***
	[0.005]
Both Sons	0.030***
	[0.005]
With other covariates	yes
Observations	50,588

Notes:a.Standard errors are in square brackets and $*p < 0.1, **p < 0.05, ***p < 0.01$.

b.The omitted category for specification (1) is when both daughters survived.

c.The other covariates in (1) are mother's age cohort, length of the second birth interval, standard of living, mother's and father's literacy, religion, caste and rural dummy.

d.Both specification include district fixed effects and the errors are clustered at the district level.

e.The sample for the estimation includes women who at least had three births by 1990, which is before the introduction of ultrasound technology..

Table 9.12: Comparison of differences in survival times of sub-groups using non-parametric tests

Groups	df	χ^2
	Wilcoxon test	
Sex composition of first two births	5	786.7***
Second birth interval	3	2985.02***
Place of residence	1	14.73***
Mother's literacy	1	1.80
Father's literacy	1	12.94***
Standard of living index	2	44.63***
Caste	4	2.46
Religion	4	113.06***
Mother's birth cohort	2	5465.38***

Table 9.13: Estimates from Cox model on time to third birth with fixed covariates

	Married women		
Variable	β	se	e^{β}
Sex composition of first two births[a]			
Daughter and Son	-0.041**	[0.018]	0.960**
Son and Daughter	-0.052***	[0.019]	0.949***
Both Sons	-0.090***	[0.018]	0.914***
With other covariates	yes	yes	yes
Observations	55,043		55,043

Notes:a.Standard errors are in square brackets and $*p < 0.1, **p < 0.05, ***p < 0.01$.

b.The omitted categories for the categorial variables are ^{a}*two surviving daughters,* ^{b}*Mother can read and write,* ^{c}*Father can read and write,* ^{d}*Scheduled caste,* ^{e}*Hindu and* ^{f}*Place of residence is rural.*

c.The model is weighted and the standard errors are robust adjusted for heteroscedasticity..

Table 9.14: Estimates from Cox model on time to third birth with fixed and time-varying covariates

	Married women		
Variable	β	se	e^{β}
Sex composition of first two births[a]			
Daughter and Son	-0.040**	[0.018]	0.961**
Son and Daughter	-0.052**	[0.019]	0.949**
Both Sons	-0.090***	[0.019]	0.914***
Age of mother	-0.063***	[0.004]	0.939***
Mother literate[b]	-0.015	[0.016]	0.986
Father Literate[c]	-0.003	[0.015]	0.997
Caste[d]			
Scheduled Tribe	0.072***	[0.026]	1.075***
Other Backward Class	0.065***	[0.019]	1.067***
Other	0.047**	[0.021]	1.048**
Do not know	0.044	[0.062]	1.045
Religion[e]			
Muslim	0.021	[0.020]	1.021
Christian	0.074**	[0.040]	1.077**
Sikh	0.179***	[0.048]	1.197***
Others	-0.06	[0.059]	0.942
Second birth interval	-0.000	[0.000]	0.999
Place of residence[f]	0.037**	[0.016]	1.038**
Time varying covariates			
Age of mother	-0.000*	[0.000]	1.000*
Observations	55,043		55,043

Notes:a.Standard errors are in square brackets and $*p < 0.1, **p < 0.05, ***p < 0.01.$

b.The omitted categories for the categorial variables are [a]*two surviving daughters,* [b]*Mother can read and write,* [c]*Father can read and write,* [d]*Scheduled caste,* [e]*Hindu and* [f]*Place of residence is rural.*

c.The model is weighted and the standard errors are robust adjusted for heteroscedasticity..

Table 9.15: Estimates from Cox model on time to third birth with fixed and time-varying covariates

	Married women		
Variable	β	se	e^{β}
Sex composition of first two births[a]			
Daughter Son	-0.073***	[0.016]	0.930***
Son Daughter	-0.068***	[0.016]	0.934***
Both Sons	-0.100***	[0.016]	0.905***
Religion[b]			
Muslim	0.049***	[0.018]	1.050***
Christian	0.115***	[0.022]	1.122***
Sikh	0.080**	[0.031]	1.083**
Other	0.009	[0.040]	1.009
Place of residence[c]			
Rural	-0.003	[0.013]	0.997
Mother's education[d]			
Incomplete primary	0.012	[0.016]	1.012
Complete primary	0.054	[0.034]	1.055
Incomplete secondary	-0.061***	[0.019]	0.941***
Complete secondary	-0.506	[0.334]	0.603
Higher	-0.243***	[0.048]	0.784***
Father's education[e]			
Incomplete primary	0.025	[0.016]	1.025
Complete primary	0.007	[0.027]	1.007
Incomplete secondary	0.018	[0.017]	1.019
Complete secondary	-0.026	[0.032]	0.974
Higher	-0.085***	[0.027]	0.918***
Mother's birth cohort[f]			
20-24	-0.082	[0.134]	0.921
25-29	-0.222*	[0.133]	0.801*
30-34	-0.329**	[0.133]	0.720**
35-39	-0.401***	[0.134]	0.670***
40-44	-0.435***	[0.135]	0.647***
45-49	-0.485***	[0.136]	0.616***
Father's birth cohort	-0.002**	[0.001]	0.998**
Economic Status of household	0.059**	[0.023]	1.060**
Observations	32879		32879

.

Table 9.16: Estimates of the likelihood of time to third birth being less than or equal to 18 months

Variable	Linear Probability Model (1)	(2)	Logit Model (3)	(4)
Sex Composition of first two births^{a}				
Daughter and Son	-0.0134***	-0.0127***	-0.0134***	-0.0127***
	[0.0045]	[0.0045]	[0.0045]	[0.0044]
Son and Daughter	-0.0103**	-0.0102**	-0.0103**	-0.0101**
	[0.0045]	[0.0045]	[0.0045]	[0.0045]
Both Sons	-0.0245***	-0.0229***	-0.0245***	-0.0228***
	[0.0043]	[0.0043]	[0.0043]	[0.0043]
With other covariates	no	yes	no	yes
Observations	55,492	55,492	55,492	55,492

Notes:a.Standard errors are in square brackets and $*p < 0.1, **p < 0.05, ***p < 0.01$.

b.The other covariates in the model are place of residence rural or urban, mother's literacy, father's literacy, second birth interval, caste, religion and mother's age by birth cohort.

c.The omitted categories for the categorial variables are ^{a}*two surviving daughters,* ^{b}*Mother can read and write,* ^{c}*Father can read and write,* ^{d}*Scheduled caste,* ^{e}*Hindu and* ^{f}*Place of residence is rural..*

Table 9.17: Amount of selection on unobservables relative to selection on observables required to attribute the entire sibling sex composition effect to selection bias

Outcome	$[E(X'\gamma \mid BB=1) - E(X'\gamma \mid BB=0)] \div \hat{Var}(X'\gamma)$ (1)	$\hat{Var}(\hat{\epsilon})$ (2)	$E(\epsilon \mid BB=1) - E(\epsilon \mid BB=0)$ (3)	$\frac{Cov(\epsilon, \tilde{BB})}{Var(\tilde{BB})}$ (4)	$\hat{\alpha}$ (5)	Implied Ratio^{a} (6)
	$\hat{\alpha}$ estimated from the OLS regression with restricted set of controls^{b}					
Third birth interval shorter than 18 months	$\frac{-0.001}{0.02*0.02}$	1.0	-2.5	$\frac{-0.001*0.497}{0.367}$	-0.0297	21.93

Notes:a.The implied ratio in column 6 is the ratio of selection on unobservables to observables under the hypothesis that there is no sibling sex composition effect.

b.The analysis is based on the restricted set of controls^{a}*-sibling sex composition of first two children, religion, caste, mother's literacy, father's literacy, mother's age at the time of third birth and place of residence.*

d.The sensitivity analysis is done on OLS regressions and not on the duration models used in section V.

e.The sensitivity analysis outlined in Altonji et al. (2005) works for binary choice explanatory variable. I restrict the sensitivity analysis to the subsample with two surviving daughters or two surviving sons to keep in line with the above methodology. Also, the variation between these two groups are most consistent across all specifications, hence, the restriction to these two groups.

Table 9.18: Third child mortality rate and third birth spacing in DLHS-2

Variable	DLHS-2	NFHS-1
Number of deaths by age category (months)		
<1	3,484	2,136
1-12	2,482	1,937
$\leq$ 12	5,966	4,073
$\leq$ 60	8,050	5,477
Third birth interval (%)(months)		
$\leq$ 18	18.3	16.7
$\leq$ 24	36.3	37.5
$\geq$ 36	22.3	30.7
Incidence of mortality (per 1,000 live births)		
Neonatal	44.6	44.1
Post neonatal	31.8	40.0
Infant	76.4	84.1
Under-five	103.1	113.1
Number of third live births	78,088	48,411

Table 9.19: Odds ratio of measures of mortality by third birth interval (DLHS-2)

Variable	Total mortality	Neo-natal mortality	Infant mortality	Under-five mortality
Variable	(1)	(2)	(3)	(4)
Third birth interval				
Less than 18 months[a]	2.348***	2.834***	2.655***	2.470***
With other covariates	yes	yes	yes	yes
Observations	78,088	78,114	78,114	78,114

Notes:a.The omitted category for the third birth interval is the length of interval being greater than 18 months.

Figure 9.1: Third birth interval distribution in DLHS-2

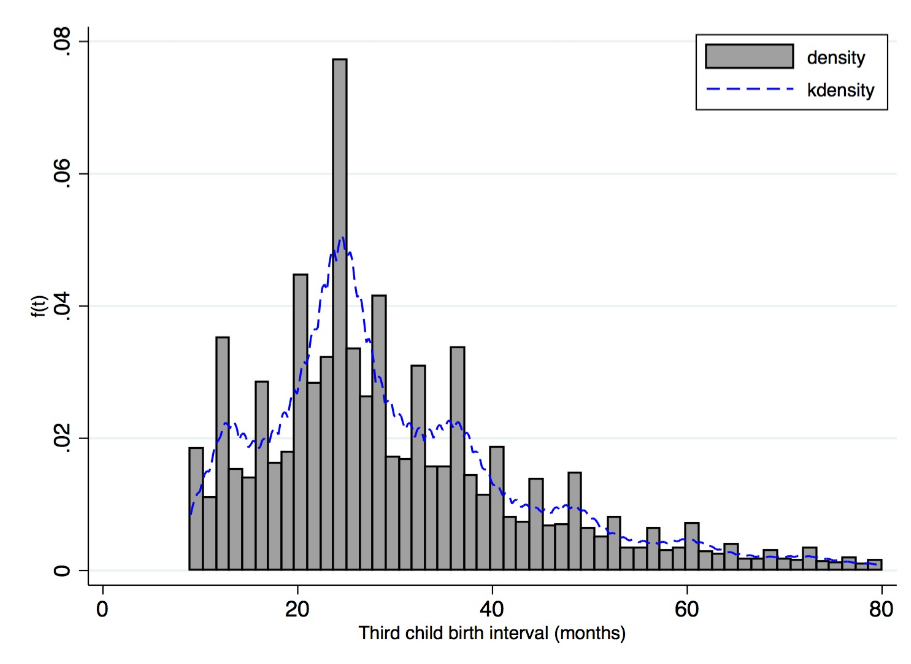


Figure 9.2: Kernel density of third birth interval in India for births prior to 1990

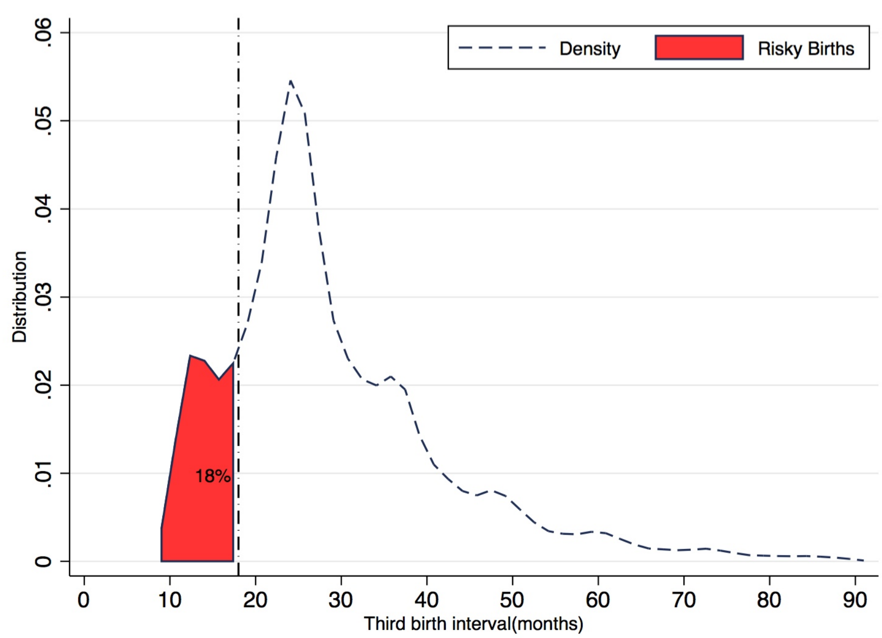


Figure 9.3: Likelihood of shorter birth interval and mother's age

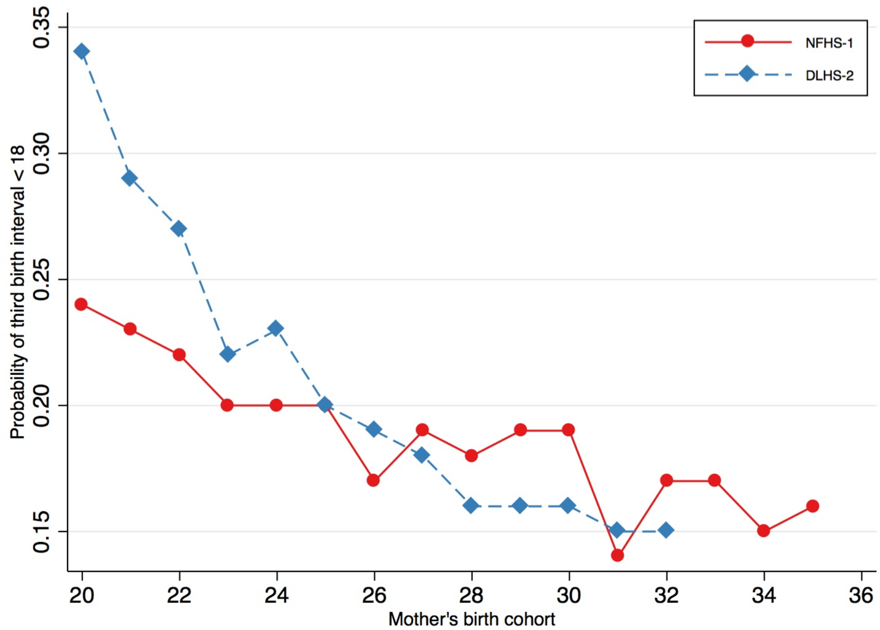


Figure 9.4: Urban and rural differentials in birth spacing

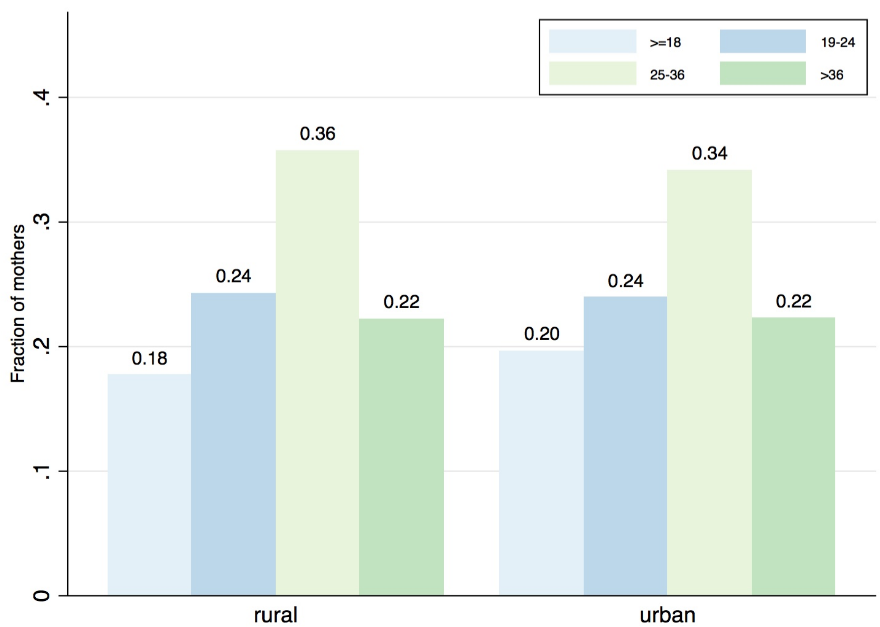


Figure 9.5: Third birth interval and mother's educational attainment

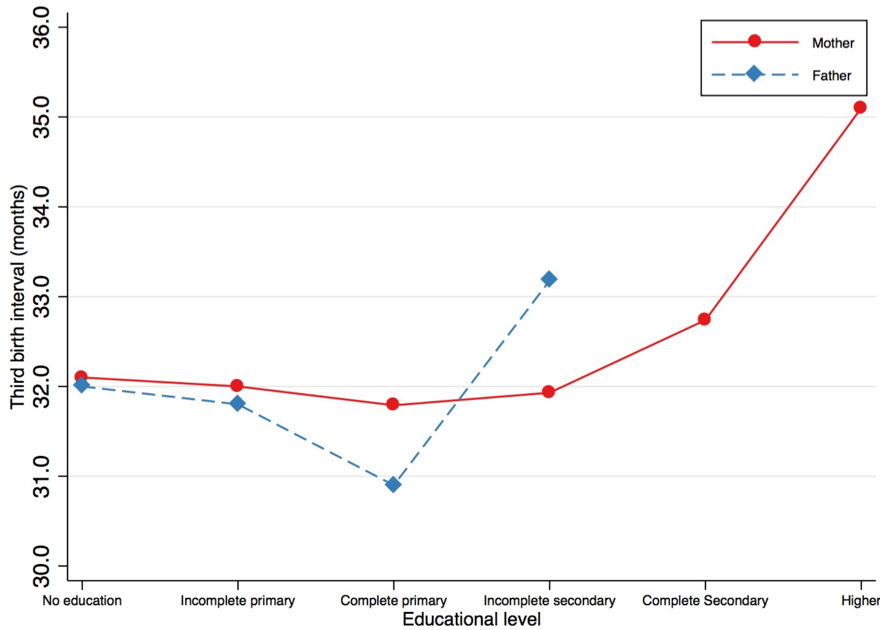
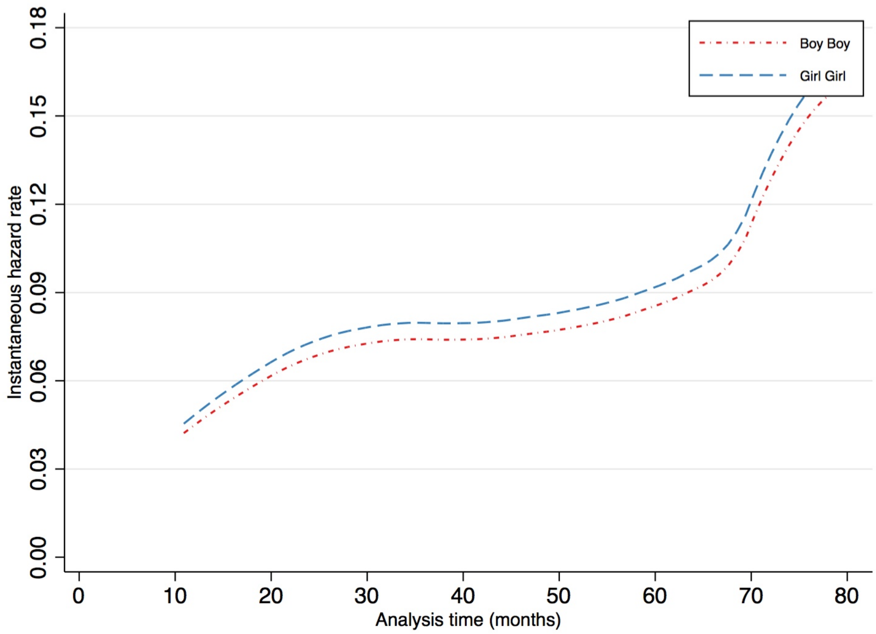


Figure 9.6: Difference in hazard rate by sibling sex composition of first two births estimated from the Cox model with fixed covariates



Notes:a.

Figure 9.7: Difference in survival time by sibling sex composition of first two births estimated from the extended Cox model with time varying covariates

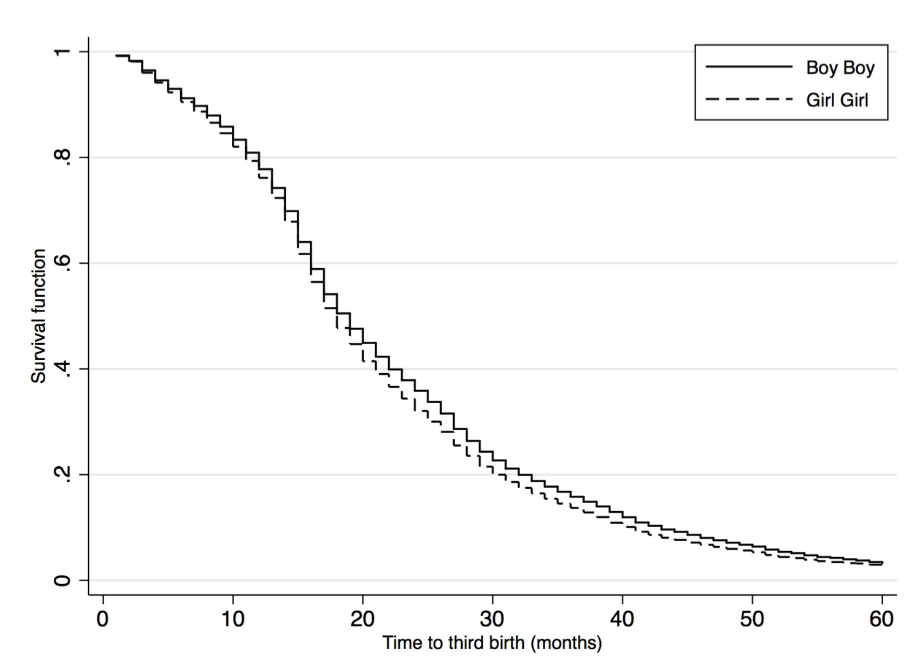


Figure 9.8: Percentile distribution of third birth interval by sibling composition

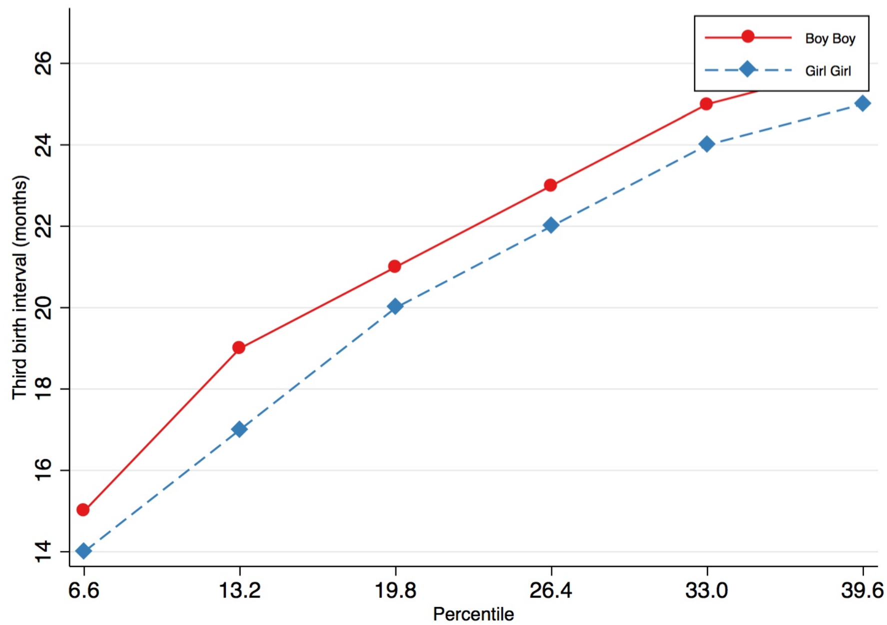


Figure 9.9: Probability of birth interval shorter than 18 months using NFHS-1

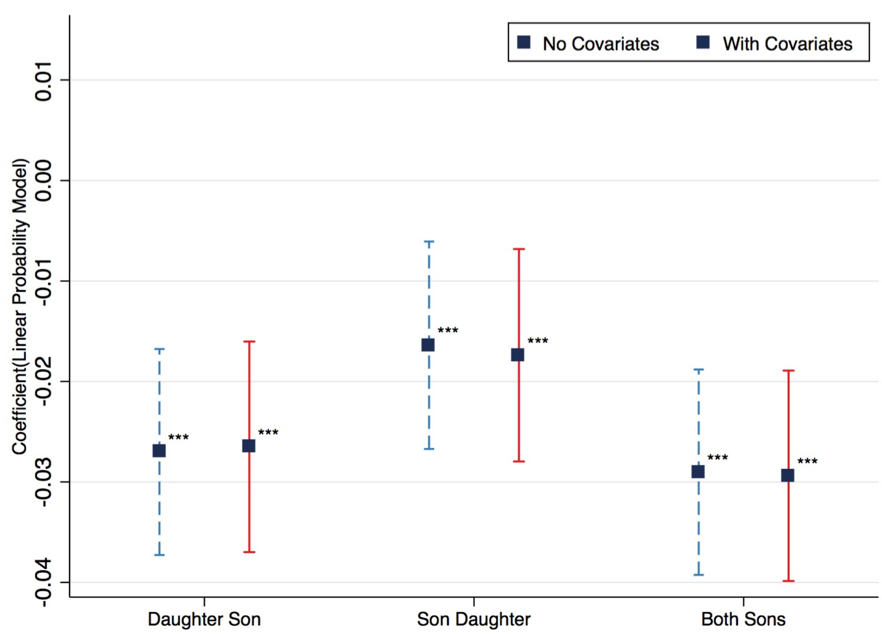


Figure 9.10: Variation in probability of birth interval shorter than 18 months by states

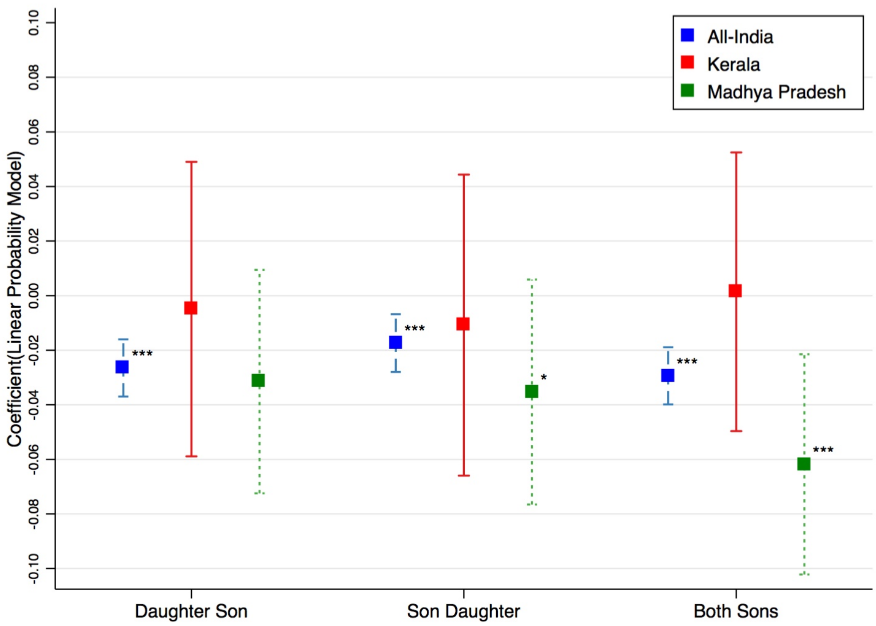


Figure 9.11: Correlation between third birth interval and mortality for third birth

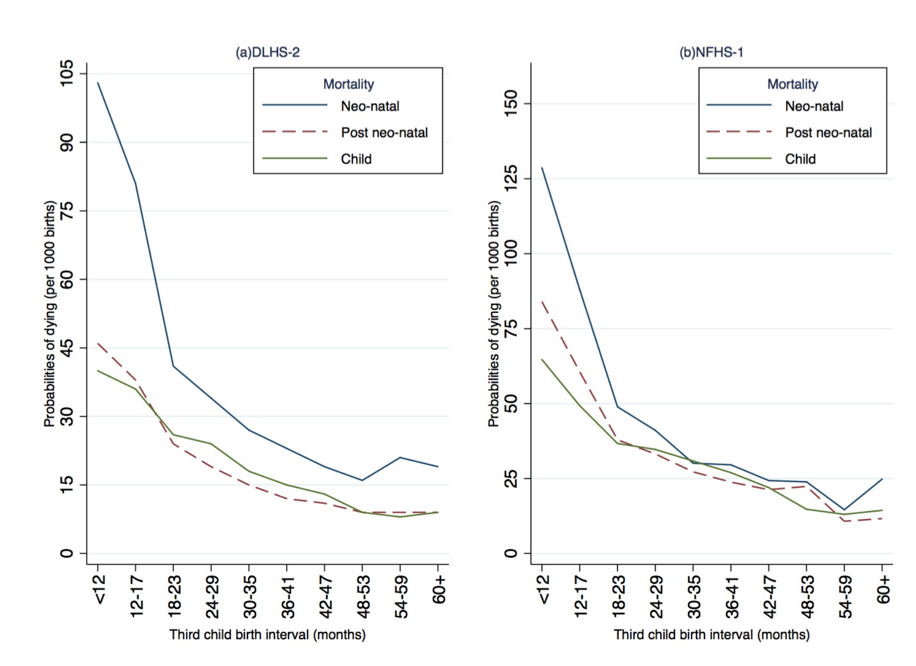
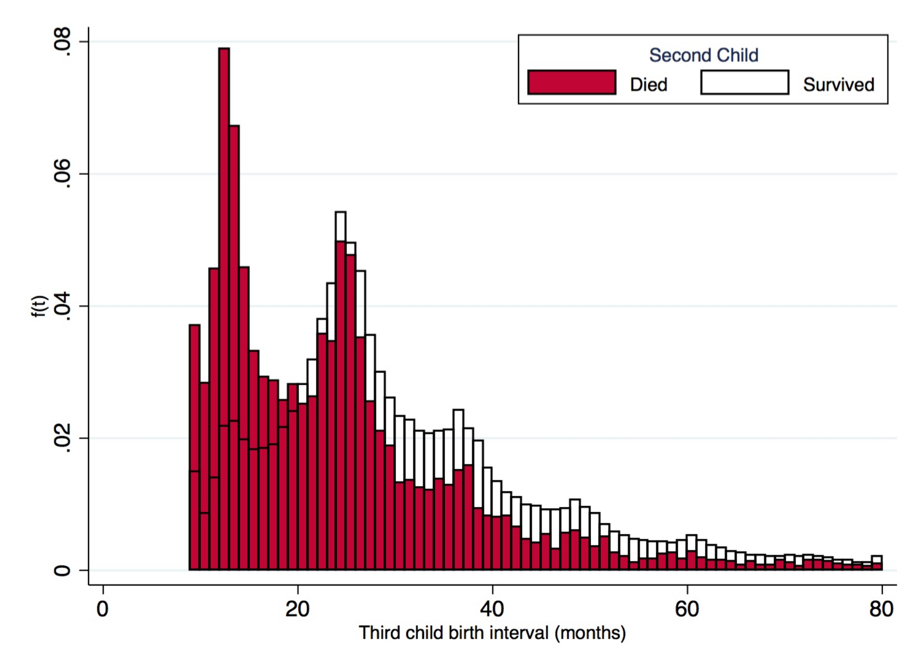


Figure 9.12: Distribution of third birth interval by the survival status of the second birth



References

Aaby, P. (1988). Malnutrition and overcrowding/intensive exposure in severe measles infection: review of community studies. *Clinical Infectious Diseases*, 10(2):478–491.

Aaby, P., Bukh, J., Lisse, I. M., Smits, A. J., Gomes, J., Fernandes, M. A., Indi, F., and Soares, M. (1984). Determinants of measles mortality in a rural area of guinea-bissau: crowding, age, and malnutrition. *Journal of tropical pediatrics*, 30(3):164–168.

Agarwal, B. (1995). Gender and legal rights in agricultural land in india. *Economic and Political Weekly*, pages A39–A56.

Agrawal, R. (1975). Influence of infant and child mortality on fertility in india.

Alam, N. (1995a). Birth spacing and infant and early childhood mortality in a high fertility area of bangladesh: age-dependent and interactive effects. *Journal of Biosocial Science*, 27(4):393–404.

Alam, N. (1995b). Predictors of diarrhoea in young bangladeshi children. *Journal of tropical pediatrics*, 41(5):278–280.

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy*, 113(1):151–184.

Anderson, S. (2003). Why dowry payments declined with modernization in europe but are rising in india. *Journal of Political Economy*, 111(2):269–310.

Angrist, J. D. and Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, 88(3):450–477.

Arnold, F., Choe, M. K., and Roy, T. K. (1998). Son preference, the family-building process and child mortality in india. *Population studies*, 52(3):301–315.

Azam, M. and Kingdon, G. G. (2013). Are girls the fairer sex in india? revisiting intra-household allocation of education expenditure. *World Development*, 42:143–164.

Barcellos, S. H., Carvalho, L. S., and Lleras-Muney, A. (2014). Child gender and parental investments in india: Are boys and girls treated differently? *American Economic Journal: Applied Economics*, 6(1):157–89.

Becker, G. S. (1960). An economic analysis of fertility, demographic and economic change in developed countries: a conference of the universities. *National Bureau Commitee for Economic Research*, 209.

Becker, G. S. (1994). Human capital revisited. In *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)*, pages 15–28. The university of Chicago press.

Becker, G. S. et al. (1988). Family economics and macro behavior. *American economic review*, 78(1):1–13.

Becker, G. S. and Lewis, H. G. (1973). On the interaction between the quantity and quality of children. *Journal of political Economy*, 81(2, Part 2):S279–S288.

Becker, G. S., Murphy, K. M., and Tamura, R. (1990). Human capital, fertility, and economic growth. *Journal of political economy*, 98(5, Part 2):S12–S37.

Behrman, J. R. (1988). Intrahousehold allocation of nutrients in rural india: Are boys favored? do parents exhibit inequality aversion? *Oxford Economic Papers*, 40(1):32–54.

Ben-Porath, Y. and Welch, F. (1976). Do sex preferences really matter? *The Quarterly Journal of Economics*, 90(2):285–307.

Bhalotra, S. and Van Soest, A. (2008). Birth-spacing, fertility and neonatal mortality in india: Dynamics, frailty, and fecundity. *Journal of Econometrics*, 143(2):274–290.

Bhargava, A. (2003). Family planning, gender differences and infant mortality: evidence from uttar pradesh, india. *Journal of Econometrics*, 112(1):225–240.

Boerma, J. T. and Bicego, G. T. (1992). Preceding birth intervals and child survival: searching for pathways of influence. *Studies in family planning*, 23(4):243–256.

Borooah, V., Do, Q.-T., Iyer, S., and Joshi, S. (2009). *Missing women and India's religious demography*. The World Bank.

Borooah, V. K. (2004). Gender bias among children in india in their diet and immunisation against disease. *Social science & medicine*, 58(9):1719–1731.

Bose, S. (2011). The effect of women's status and community on the gender differential in children's nutrition in india. *Journal of biosocial science*, 43(5):513–533.

Breslow, N. (1970). A generalized kruskal-wallis test for comparing k samples subject to unequal patterns of censorship. *Biometrika*, 57(3):579–594.

Buckles, K. S. and Munnich, E. L. (2012). Birth spacing and sibling outcomes. *Journal of Human Resources*, 47(3):613–642.

Chaudhuri, S. (2012). The desire for sons and excess fertility: a household-level analysis of parity progression in india. *International perspectives on sexual and reproductive health*, pages 178–186.

Chiswick, M. L. (1985). Intrauterine growth retardation. *British medical journal (Clinical research ed.)*, 291(6499):845.

Clark, S. (2000). Son preference and sex composition of children: Evidence from india. *Demography*, 37(1):95–108.

Cleland, J. G. and Sathar, Z. A. (1984). The effect of birth spacing on childhood mortality in pakistan. *Population studies*, 38(3):401–418.

Cleves, M., Gould, W., Gould, W. W., Gutierrez, R., and Marchenko, Y. (2008). *An introduction to survival analysis using Stata*. Stata press.

Conde-Agudelo, A., Rosas-Bermudez, A., Castaño, F., and Norton, M. H. (2012). Effects of birth spacing on maternal, perinatal, infant, and child health: a systematic review of causal mechanisms. *Studies in family planning*, 43(2):93–114.

Conde-Agudelo, A., Rosas-Bermúdez, A., and Kafury-Goeta, A. C. (2006). Birth spacing and risk of adverse perinatal outcomes: a meta-analysis. *Jama*, 295(15):1809–1823.

Conley, D. and Glauber, R. (2006). Parental educational investment and childrens academic risk estimates of the impact of sibship size and birth order from exogenous variation in fertility. *Journal of human resources*, 41(4):722–737.

Coombs, L. C. and Sun, T.-H. (1978). Family composition preferences in a developing culture: The case of taiwan, 1973. *Population Studies*, 32(1):43–64.

Cox, D. R. (1972). Models and life-tables regression. *JR Stat. Soc. Ser. B*, 34:187–220.

Cruces, G. and Galiani, S. (2007). Fertility and female labor supply in latin america: New causal evidence. *Labour Economics*, 14(3):565–573.

Currie, J. and Yelowitz, A. (2000). Are public housing projects good for kids? *Journal of public economics*, 75(1):99–124.

D'agostino, R. B., Belanger, A., and D'Agostino Jr, R. B. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4):316–321.

D'Agostino, R. B. and Belanger, A. J. (1992). Comment on tests of normality. *Stata Technical Bulletin*, 1(3).

Daouli, J., Demoussis, M., and Giannakopoulos, N. (2009). Sibling-sex composition and its effects on fertility and labor supply of greek mothers. *Economics Letters*, 102(3):189–191.

Das, N. (1987). Sex preference and fertility behavior: A study of recent indian data. *Demography*, 24(4):517–530.

Das Gupta, M. (2010). Family systems, political systems and asia's missing girls the construction of son preference and its unravelling. *Asian Population Studies*, 6(2):123–152.

Das Gupta, M. and Mari Bhat, P. (1997). Fertility decline and increased manifestation of sex bias in india. *Population studies*, 51(3):307–315.

Das Gupta, M., Zhenghua, J., Bohua, L., Zhenming, X., Chung, W., and Hwa-Ok, B. (2003). Why is son preference so persistent in east and south asia? a cross-country study of china, india and the republic of korea. *The Journal of Development Studies*, 40(2):153–187.

Dercon, S. and Singh, A. (2013). From nutrition to aspirations and self-efficacy: gender bias over time among children in four countries. *World Development*, 45:31–50.

Devi, P. K. (2013). Impact of son preference on 3rd birth transition in manipur: A logistic regression analysis. *OSR Journal Of Humanities And Social Science (IOSR-JHSS)*, 15(3):1–4.

Dyson, T. and Moore, M. (1983). On kinship structure, female autonomy, and demographic behavior in india. *Population and development review*, pages 35–60.

DAddato, A. V. (2006). Progression to third birth in morocco in the context of fertility transition. *Demographic Research*, 15:517–536.

Fawcett, J. T. (1983). Perceptions of the value of children: Satisfactions and costs. *Determinants of fertility in developing countries*, 1:347–369.

Fayehun, O., Omololu, O., and Isiugo-Abanihe, U. (2011). Sex of preceding child and birth spacing among nigerian ethnic groups. *African journal of reproductive health*, 15(2).

Filmer, D. and Pritchett, L. (1998). *Educational enrollment and attainment in India: Household wealth, gender, village, and state effects*. World Bank Washington, DC.

Fledderjohann, J., Agrawal, S., Vellakkal, S., Basu, S., Campbell, O., Doyle, P., Ebrahim, S., and Stuckler, D. (2014). Do girls have a nutritional disadvantage compared with boys? statistical models of breastfeeding and food consumption inequalities among indian siblings. *PloS one*, 9(9):e107172.

Gandhi Kingdon, G. (2002). The gender gap in educational attainment in india: How much can be explained? *Journal of Development Studies*, 39(2):25–53.

Gehan, E. A. (1965). A generalized wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52(1-2):203–224.

Gilbert, K. and Danker, H. (1981). Investigations on the changes of sex ratio in germany from 1826 up to 1978. *Acta anthropogenetica*, 5(2):89–110.

Goux, D. and Maurin, E. (2005). The effect of overcrowded housing on children's performance at school. *Journal of public Economics*, 89(5-6):797–819.

Gupta, M. D. (1987). Selective discrimination against female children in rural punjab, india. *Population and development review*, pages 77–100.

Hajian-Tilaki, K. O. (2002). Factors affecting the pattern of postpartum amenorrhea. *Annals of Saudi medicine*, 22(5-6):404–407.

Hale, L., DaVanzo, J., Razzaque, A., and Rahman, M. (2009). Which factors explain the decline in infant and child mortality in matlab, bangladesh? *Journal of Population Research*, 26(1):3–20.

Heckman, J. J. and Walker, J. R. (1990). The third birth in sweden. *Journal of Population Economics*, 3(4):235–275.

Hobcraft, J. N., McDonald, J. W., and Rutstein, S. O. (1985). Demographic determinants of infant and early child mortality: a comparative analysis. *Population studies*, 39(3):363–385.

IIPS, I. I. f. P. S. (2006). District level household survey (dlhs-2), 2002–04.

IIPS, I. I. f. P. S. (2007). *India National Family Health Survey (NFHS-3), 2005-06*, volume 1. International Institute for Population Sciences.

IIPS, I. I. f. P. S. (2010). District level household and facility survey (dlhs-3), 2007–08: India.

James, W. H. (1987). The human sex ratio. part 1: A review of the literature. *Human biology*, pages 721–752.

Jayachandran, S. and Kuziemko, I. (2011). Why do mothers breastfeed girls less than boys? evidence and implications for child health in india. *The Quarterly journal of economics*, 126(3):1485–1538.

Jayachandran, S. and Pande, R. (2017). Why are indian children so short? the role of birth order and son preference. *American Economic Review*, 107(9):2600–2629.

Jensen, R. T. (2003). *Equal Treatment, Unequal Outcomes?: Generating Sex Inequality Through Fertility Behaviour*. Citeseer.

Kaul, T. (2018). Intra-household allocation of educational expenses: Gender discrimination and investing in the future. *World Development*, 104:336–343.

Kingdon, G. G. (1998). Does the labour market explain lower female schooling in india? *The Journal of Development Studies*, 35(1):39–65.

Mace, R. and Sear, R. (1997). Birth interval and the sex of children in a traditional african population: an evolutionary analysis. *Journal of biosocial science*, 29(4):499–507.

Malhotra, A., Vanneman, R., and Kishor, S. (1995). Fertility, dimensions of patriarchy, and development in india. *Population and development review*, pages 281–305.

Mayer, P. (1999). India's falling sex ratios. *Population and development review*, 25(2):323–343.

Murthi, M., Guio, A.-C., and Dreze, J. (1995). Mortality, fertility, and gender bias in india: A district-level analysis. *Population and development review*, pages 745–782.

Nath, D. C. and Land, K. C. (1994). Sex preference and third birth intervals in a traditional indian society. *Journal of Biosocial Science*, 26(3):377–388.

Nath, D. C., Land, K. C., and Singh, K. K. (1994). Birth spacing, breastfeeding, and early child mortality in a traditional indian society: a hazards model analysis. *Social Biology*, 41(3-4):168–180.

Nath, D. C., Singh, K. K., Land, K. C., and Talukdar, P. K. (1993). Age of marriage and length of the first birth interval in a traditional indian society: Life table and hazards model analysis. *Human Biology*, pages 783–797.

Newman, J. L. (1983). Economic analyses of the spacing of births. *The American Economic Review*, 73(2):33–37.

Newman, J. L. and McCulloch, C. E. (1984). A hazard rate approach to the timing of births. *Econometrica: journal of the Econometric Society*, pages 939–961.

Nugent, J. B. (1985). The old-age security motive for fertility. *Population and development review*, pages 75–97.

Oster, E. (2009). Does increased access increase equality? gender and child health investments in india. *Journal of development Economics*, 89(1):62–76.

Pal, S. (1999). An analysis of childhood malnutrition in rural india: role of gender, income and other household characteristics. *World Development*, 27(7):1151–1171.

Pande, R. P. (2003). Selective gender differences in childhood nutrition and immunization in rural india: the role of siblings. *Demography*, 40(3):395–418.

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

Pandey, A., Sengupta, P. G., Mondal, S. K., Gupta, D. N., Manna, B., Ghosh, S., Sur, D., and Bhattacharya, S. (2002). Gender differences in healthcare-seeking during common illnesses in a rural community of west bengal, india. *Journal of Health, Population and Nutrition*, pages 306–311.

Patkar, A. (1995). Socio-economic status and female literacy in india. *International Journal of Educational Development*, 15(4):401–409.

Popkin, B. M., Guilkey, D. K., Akin, J. S., Adair, L. S., Udry, J. R., and Flieger, W. (1993). Nutrition, lactation, and birth spacing in filipino women. *Demography*, 30(3):333–352.

Rahman, L. and Rao, V. (2004). The determinants of gender equity in india: examining dyson and moore's thesis with new data. *Population and Development Review*, 30(2):239–268.

Rodgers, J. L. and Doughty, D. (2001). Does having boys or girls run in the family? *chance*, 14(4):8–13.

Royston, P. (1992). Tests for departure from normality. *Stata Technical Bulletin*, 1(2).

Saha, A. (2013). An assessment of gender discrimination in household expenditure on education in india. *Oxford Development Studies*, 41(2):220–238.

Sen, A. (1992). Missing women. *BMJ: British Medical Journal*, 304(6827):587.

Sen, A. (2001). The many faces of gender inequality. *New republic*, pages 35–39.

Singh, S. and Bhaduri, T. (1971). On the pattern of post-partum amenorrhea. In *Proceedings of All-India Seminar on Demography and Statistics, Banaras Hindu University*, page 284.

Smith, D. P. (1985). Breastfeeding, contraception, and birth intervals in developing countries. *Studies in Family Planning*, 16(3):154–163.

Snow, C. F. (1999). Laboratory diagnosis of vitamin b12 and folate deficiency: a guide for the primary care physician. *Archives of internal medicine*, 159(12):1289–1298.

Srinivas, M. N. (1980). *The remembered village*. Number 26. Univ of California Press.

Srinivasan, P. and Lee, G. R. (2004). The dowry system in northern india: Women's attitudes and social change. *Journal of Marriage and Family*, 66(5):1108–1117.

Stovall, D. W., Toma, S. K., Hammond, M. G., and Talbert, L. M. (1991). The effect of age on female fecundity. *Obstetrics and gynecology*, 77(1):33–36.

Sueyoshi, G. T. (1992). Semiparametric proportional hazards estimation of competing risks models with time-varying covariates. *Journal of econometrics*, 51(1-2):25–58.

Thapa, S., Short, R. V., and Potts, M. (1988). Breast feeding birth spacing and their effects on child survival. *Nature*, 335(6192):679–82.

Total Fertility Rate, W. (2018). Total fertility rate. In *Retrieved from http://www.searo.who.int/entity/health*$_situation_trends/data/chi/TFR/en/$.

Trussell, J., Martin, L. G., Feldman, R., Palmore, J. A., Concepcion, M., DatoAbu, D. N. L. B., et al. (1985). Determinants of birth-interval length in the philippines, malaysia, and indonesia: A hazard-model analysis. *Demography*, 22(2):145–168.

Tu, P. (1991). Birth spacing patterns and correlates in shaanxi, china. *Studies in family planning*, 22(4):255–263.

van Noord-Zaadstra, B. M., Looman, C. W., Alsbach, H., Habbema, J. D., te Velde, E. R., and Karbaat, J. (1991). Delaying childbearing: effect of age on fecundity and outcome of pregnancy. *Bmj*, 302(6789):1361–1365.

Vandenbosche, R. C. and Kirchner, J. (1998). Intrauterine growth retardation. *American family physician*, 58(6):1384–90.

Verma, J., Sachar, R., Prakash, V., Sehgal, R., and Aggarwal, S. (1990). Effect of sex of preceding child on birth interval. *Indian journal of maternal and child health: official publication of Indian Maternal and Child Health Association*, 1(2):50.

Vlassoff, C. (1990). The value of sons in an indian village: How widows see it. *Population studies*, 44(1):5–20.

Warkany, J., Monroe, B. B., and Sutherland, B. S. (1961). Intrauterine growth retardation. *American journal of diseases of children*, 102(2):249–279.

Whitworth, A. and Stephenson, R. (2002). Birth spacing, sibling rivalry and child mortality in india. *Social Science & Medicine*, 55(12):2107–2119.

Winikoff, B. (1983). The effects of birth spacing on child and maternal health. *Studies in family planning*, pages 231–245.

Zenger, E. (1993). Siblings neonatal mortality risks and birth spacing in bangladesh. *Demography*, 30(3):477–488.

Zhu, B.-P., Rolfs, R. T., Nangle, B. E., and Horan, J. M. (1999). Effect of the interval between pregnancies on perinatal outcomes. *New England journal of medicine*, 340(8):589–594.

Zimmermann, L. (2012). Reconsidering gender bias in intrahousehold allocation in india. *Journal of Development Studies*, 48(1):151–163.

10 Appendix

10.1 Additional Figures and Tables

10.1.1 Test on Residuals of log-linear model

Table 10.1: Tests of normality on the residuals of the log-linear model

Variable	Observations	Pr(Skewness)	Pr(Kurtosis)
Residuals from model(1)	60,390	0.000	0.000
Residuals from model(2)	49,737	0.000	0.000

Notes:a. .

Figure 10.1: Evidence suggesting violation of normality assumption by the third birth interval from DLHS-2