

# How Missing Men Lead to Missing Women: Revising Natural Sex Ratios at Birth for the Fragile Male

Joshua Wilde \*

December 7, 2020

**Abstract:** Estimates of missing women rely critically on estimates of the natural sex-ratio at birth (SRB). Current estimates ignore in utero male fragility, where male conceptions are disproportionately terminated in the presence of maternal stress. Using natality data from the United States, we document large correlations between a wide range of predictors of spontaneous terminations and SRBs, such as education, poverty, age, parity, birth interval, and even month, day, and time of birth. We show that controlling for maternal stress overturns many commonly held beliefs about natural SBRs. By correcting existing age, parity, and interval estimates by employing woman fixed effects, we show that there should be more "missing men" at birth than currently observed in many developing nations, implying that globally the number of missing women is underestimated by about 30%, and that 20% of the increased SRBs over the past 50 years are due naturally to the demographic transition.

**Keywords:** Missing Women, Sex Ratio, Demographic Transition, Sex Selection, Abortion, Fertility.

**JEL codes:** J11, J13, J16, O12, O15.

---

\*Joshua Wilde: Corresponding Author. Max Planck Institute for Demographic Research, Konrad-Zuse-Straße 1, 18057 Rostock, Germany. E-mail: joshua.kent.wilde@gmail.com.

# 1 Introduction

Historically, sex ratios at birth (SRBs) have varied in a narrow range around 105 female births per 100 male births, with only a few known variations among racial groups (Chao et al 2019). However, over recent decades, SRBs have risen in a number of Asian countries and in eastern Europe due to increasing demand for sex selective abortion, as well as easier access to technologies which can make sex selection possible. This growing imbalance between male and female births led Amartya Sen in 1990 to pioneer the idea of “missing women” – the number of women which should be observed in the absence of sex-preferential behavior, but who no longer exist. In particular, estimates of missing women at birth rely critically on estimates of the natural sex-ratio at birth.

In this paper, we challenge the long-held consensus belief that sex ratios at birth are a biological constant at 105 males for every 100 females, with some small variation among racial groups. A growing literature has demonstrated that male conceptions are more likely to be spontaneously terminated before birth in the presence of maternal stress (Bruckner and Catalano, 2006 & 2009, Nobles and Hamoudi 2014, Wilde et al 2017, Wilde et al 2019, Almond et al XXXX). Building upon this theoretical idea, we document that in spite of the consensus that SRBs are a biological constant, that they fluctuate significantly with indicators which predict maternal stress – even in settings where sex selective abortion is thought to be minimal, and where SRBs have not changed over time. Using completed birth history data for millions of women in the US and sub-Saharan Africa – two societies in which sex selective behavior is believed to be very low – we show that the probability a woman bears male rather than a female is highly correlated with indicators of maternal stress, such as socioeconomic status, education, poverty, mothers age at birth, parity, and even the month, day, and time of birth.

In order to distinguish correlation from causality, we introduce a woman-fixed effect into our analysis – unused previously in the sex ratio at birth literature – to control for the fact that healthier women should be more fertile and have both more children and more males. Controlling for women fixed effects reveals significant female bias in US and African births for higher parities, shorter birth intervals, and higher ages, demonstrating that women with higher levels of maternal depletion are more likely to have females.

Using this information, we provide updated estimates of missing women in India, controlling for the fact that women in India face higher levels of maternal stress due to shorter birth intervals, higher fertility rates, and a higher incidence of poverty. Using completed birth history data from India, we perform a counterfactual analysis of what the natural SRB in India would be if the effects of maternal stress on the probability of a female birth estimated using US data were applied to India. Our findings indicate that Indian sex ratios at birth should naturally be lower due to the higher incidence of “missing men” – males who would have been carried to term, but were spontaneously terminated due to high levels of maternal stress. As a result, missing women in India are being underestimated, because the comparison sex ratio at birth

is too high. We estimate that by adjusting for demographics alone – ignoring the effects of higher levels of Indian poverty – missing women in India have been underestimated by 30%. However, we find that 20% of the increase in SRBs in India over the last 40 years would have occurred naturally, due to India’s recent demographic shift in transitioning from a high to a low fertility society.

Beyond missing women, our findings explain or overturn multiple commonly held beliefs regarding SRBs. For example, differences in SRBs across racial groups – which persist even in societies where sex selective abortion is rare – have commonly been believed to be due to innate genetic differences. We show that SRBs in the US correlate almost perfectly with socioeconomic status, implying that part of what was supposed to be genetic is more likely caused more by socioeconomic factors. We demonstrate that approximately 25% of the differences across racial groups in the US can be explained by variation in demographic and socioeconomic characteristics of mothers rather than race.

It also calls into question the theoretical interpretation of several seminal papers in the economics literature. For example, we show that research documenting evidence for a demand for sons in the US likely suffers from reverse causality. For example, the negative correlation found by Dahl and Moretti (2006) between bearing a female and socioeconomic outcomes is likely not due to women being penalized for having girls, but rather because women with worse outcomes are more likely to bear girls in the first place. Nobles and Hamoudi (2014) point this out specifically for Dahl and Moretti’s divorce result, and this paper extends their idea to a larger number of maternal stress variables using large, population-based datasets over a variety of countries.

This paper is organized as follows. Section 2 reviews the literature on SRBs, discusses the consensus that SRBs are a biological constant of around 1.05 males per female, and demonstrates there is considerable evidence that this consensus is erroneous using US natality data. Section 3 describes our empirical methodology, including the importance of introducing women fixed effects in order to control for a spurious correlation between the probability of a male birth and the probability of an additional birth. Section 4 presents our results for US, which we use as a baseline estimate of the natural SRB. Section 5 extends this analysis to the developing world, considering the cases of India – where SRBs are thought to be highly influenced by sex selective abortion – and Africa, where there is believed to be minimal sex-selective behavior. Section 6 describes how the findings of this paper affect the existing literature, including providing updated estimates of missing women in India, as well as describing other papers our findings explain or overturn.

## 2 Background

### 2.A Literature Review

**To be completed, ran out of time before deadline.** Mention work by Seema Jaychandaran, Debraj Ray RES, Bongaarts 2015 PDR, Chao et al 2019 PNAS, everything Bruckner and Catalano have ever done, Amar Hamoudi and Jenna Nobles super undercited Demography paper, Dahl and Moretti RES, Guilmoto 2009, 2013, and others, William James reviews. 2 Wilde temperature and SRB papers. Doug Almond 2008 PNAS, 2013 PDR.

### 2.B Simple Correlates of Sex Ratios at Birth in the US

That SRBs are an invariant biological constant at approximately 105 males per 100 females is generally taken as given, and variation from this constant is generally attributed to sex selective abortion, with some smaller variation attributed to innate racial differences. For example, the fact that birth masculinity increases with parity or household wealth in India is explained by increasing demand for males at higher parities or higher access to sex selective technologies by the wealthy. While these effects are surely a large driver of skewed sex-ratios at birth in India, they are less salient an explanation for countries with lower levels of son preference, such as the US, Europe, or sub-Saharan Africa.

To make the case that SRBs vary significantly with other factors beyond race and sex selection, this section uses approximately 4 million birth observations in the 2016 US Natality files to present simple correlations between a large set of socioeconomic and maternal health indicators with sex ratios at birth. It should be noted that these are correlations, and therefore are not well identified. This analysis is simply to document new and considerable evidence – even in a cross-section – which suggests that SRBs are not a simple biological constant.

In Figure 2 panels A and B, I show the relationship between sex ratios at birth and maternal income and education. In both cases, higher SES women are more likely to have male births, and the gradient is both highly statistically significant and significant in magnitude. In the case of maternal education, mothers who did not graduate from high school have a sex ratio at birth of slightly less than 1.04, while women with bachelor's degrees or higher have sex ratios at birth of over 1.05 – a difference a third as large as the estimated effect of sex selection in India. The differences with income are even more stark: sex ratios at birth from women below the poverty level have sex ratios at birth of 1.012, while those with household incomes higher than 5 times the poverty threshold have a SRB of 1.10 – a difference almost twice the size of estimated sex selection in India.

That SRBs vary with race is well known – however, as shown in Figure 3, these differences are almost perfectly predicted by the relative socioeconomic status of each group. This suggests that income may at least partially explain SRB differences between these groups. Figure 4 also makes this point: averages in sex ratios at birth across countries – usually thought to be solely a

difference in racial composition – are highly predicted by income per capita in those countries. This is true in Panel A, a cross section of countries excluding 12 countries which have seen increases in sex ratios at birth after the 1970s, as well as Panel B, the set of all countries before 1970.<sup>1</sup>

Large difference in SRBs even show up in month of birth data, as predicted either by socioeconomic status or maternal stress. Figures 5, 6, and 7 show that there is seasonal variation in sex ratios at birth, but that the seasonal variation in SRBs can be explained by seasonal variation in maternal education by month of birth (Buckles and Hungerman). In addition, temperature at the time of conception is a known driver of spontaneous abortion, and the seasonality of SRBs is even more strongly predicted by temperature seasonality 9 months before birth than seasonality in maternal education.

Similar correlations are found when analyzing SRBs on different days of the week, and even in the time of day of birth, as shown in Figures 7 - 9. Figure 7 shows that births which occur on the weekend are much more likely to be male, consistent with the hypothesis that births at higher risk for miscarriage are scheduled for inductions or C-sections during the weekdays, and such births are more likely to be female. In addition, Figure 8 shows that births which occur during times for scheduled C-sections and inductions are significantly more female than births which occur after hours.<sup>2</sup> Figure 9 shows that these differences are not driven by different racial groups selecting into C-sections or inductions at different times of the day – the pattern arises for all racial groups, although there are different level effects for each racial group, consistent with the hypothesis of innate racial SRB differences.

Similarly, Figures 10 and 11 show that births are more likely to be female if the mother is older, has had more births previously, or has had twins or a birth interval of just 12-17 months – all indicators of maternal stress or depletion. In results not shown, we also document that smoking before pregnancy predicts birth femininity, as well as premature birth, being underweight before pregnancy, and birth plurality, where triplets are more likely to be female than twins, which are subsequently more female than singleton births.

While all of these SES and maternal depletion indicators likely vary with each other, taken together there are a large set of indicators of maternal stress, depletion, or poor health which are strongly correlated with lower sex ratios at birth. Therefore, there is at least superficial evidence that the current consensus on this topic is not well empirically grounded. Testing whether this is actually the case is the goal of sections 3 and 4 of this paper.

---

<sup>1</sup>The fact that this pattern occurs before 1970 is significant since this is a time period in which sex-selective technologies such as ultrasounds were less available. For example, it was not until the mid to late 1970s that SRB began rising in India, China, and other countries for which there is evidence of sex-selection.

<sup>2</sup>C-sections and inductions are usually performed between 7 am and 5 pm, with the majority happening towards the beginning of this range.

### 3 Theory and Methodology

In this section we derive theoretical implications of the fragile male hypothesis for sex ratios at birth, and demonstrate that the fragile male hypothesis solves puzzles regarding the gradient of biological sex with age and parity.

#### 3.A Conceptual Framework

To model how the probability of a birth being female may change with maternal stress, begin by defining states of the world which determine whether a birth will be male or female, beginning at conception. Once conceived, a conceptus is either male or female, and will either be born or terminated before birth. Let  $P(B)$  and  $P(T)$  be the unconditional probability of birth and termination of a given conceptus. Similarly let  $P(M)$  and  $P(F)$  be the unconditional probability that a conceptus is male or female. Note that each of these probability pairs forms a set of mutually exclusive events, leaving four possible outcomes: a conceptus being male and successfully being born, with probability  $P(M \cap B)$ ; being male but being terminated before birth, with probability  $P(M \cap T)$ ; being female and born, with probability  $P(F \cap B)$ ; and being female and terminated, with probability  $P(F \cap T)$ .

Given a large enough number of births and barring the effects of selection into childbirth, the fraction of female births observed in data will be equal to the conditional probability of being female conditional on being born for the representative woman. By the law of conditional probability, this can be expressed as:

$$P_w(F|B) = \frac{P_w(F \cap B)}{P_w(B)} = \frac{P_w(F \cap B)}{P_w(M \cap B) + P_w(F \cap B)}. \quad (1)$$

Now assume that there is some form of stress or poor health to the mother denoted by  $S$ , which may affect the probability of termination. It can be shown that the change in the probability of a birth being female can be expressed as:

$$\frac{\partial P_w(F|B)}{\partial S} = \frac{\frac{\partial P_w(F \cap B)}{\partial S} \cdot \frac{P_w(M|B)}{P_w(F|B)} - \frac{\partial P_w(M \cap B)}{\partial S}}{P_w(B)}. \quad (2)$$

Notice that the denominator of this expression will always be positive, and therefore whether maternal stress changes the sex ratio at birth – and in which direction – will depend solely on the sign of the numerator. Also notice the the expression  $\frac{P_w(M|B)}{P_w(F|B)}$  is simply the sex ratio at birth, commonly thought to be 105 males per 100 females, or 1.05, which is very close to 1. Therefore, so long as sex ratios at birth are not too skewed, the direction of the change in the probability of a female birth will depend on whether stress decreases the probability of birth more for male or female conceptuses, such that if  $\frac{\partial P_w(M \cap T)}{\partial S} > \frac{\partial P_w(F \cap T)}{\partial S}$ , then  $\frac{\partial P_w(F|B)}{\partial S} > 0$ .

That maternal stress will disproportionately affect male terminations is precisely the asser-

tion of the fragile male research outlined above. If this is the case, we can better estimate the “natural” sex ratio at birth by incorporating information on potential maternal stressors. Assume there is some baseline probability of a female birth in the absence of stressors. In this case, there will be some natural rate of gender specific miscarriage, which will increase in the presence of stressors. Formally, let the probability that live birth  $i$  from woman  $w$  be:

$$P_{i,w}(F|B) = \alpha + \beta S_{i,w} + \epsilon_{i,w} \quad (3)$$

where  $\alpha$  is the “natural” fraction of female births, and  $S$  is a variable or variables which capture stress during the critical windows of gestation. In this case  $\beta$  could be a broad set of coefficients which predict the effect of different stressors on gender-biased terminations, which according to the theory should be negative.

In an optimal research design, one would hope for random shocks to  $S_{i,w}$  in order to properly identify  $\beta$ . There are such studies which look at economic shocks (cite), temperature (cite), and even the September 11th attacks (cite) as random stressors, which studies show that nine months after such random stress shocks the sex ratios at birth become female biased. However, each of these stressors are very specific to the individual shock, and not suited for revisiting the overall consensus on natural sex ratios at birth. As a result, in this paper we focus on long-term variables which can be easily gleaned from large population datasets which might permanently affect a woman’s ability to bring a pregnancy to term, but which also are affected by broad patterns of demographic change.

In seeking variables which affect miscarriage which are not based on temporary shocks and can also be extrapolated from large datasets, several sets of variables may be considered. First, individuals with lower socioeconomic status have miscarriages at higher rates, which may be because of poor health or higher levels of maternal stress. Variables which are correlated with both socioeconomic status and higher miscarriage include education, certain racial groups, and income. Second, women which have given birth to more children traditionally have higher miscarriage rates, potentially due to higher levels of maternal depletion. Maternal age and short birth intervals are also associated with miscarriage rates.

### 3.B Methodology and the Case for Woman Fixed Effects

Studies which estimate the effects of parity, age, and other indicators on sex ratios at birth generally do not include mother fixed effects, and simply look at estimates from birth registry data in the cross section which do not have information on other births by the same woman. However, if maternal stress causes miscarriage, and age and parity are a form of maternal stress or depletion, then this introduces a statistical problem. Specifically, there may be some women who are generally more healthy in all phases of life due to – for example – a better genetic endowment, higher social class, or more privileged upbringing. In this case, these women will have lower rates of miscarriages, and disproportionately less male miscarriages, meaning they

will be more likely to have both more births and more male births. If this effect were strong enough, we may even see in the cross-section that higher parities would be more male, even if the probability of having a female for a given woman rose with higher parities, simply because women who are able to have higher parity births are likely to be more healthy throughout their lives. Through the same logic, older women who successfully have a birth are more likely to be in better health, meaning they are also more likely to birth a male than the general population.

Therefore, while the probability of a given birth to a woman into childbirth  $P_w(F|B)$  is as given in equation (2), the sex ratio at birth in the population by parity, age, and other demographic characteristics must also take into account the selection effects by maternal health of which women are able to have births at those ages or parities. Formally, the observed sex ratio at birth given a specific value of stress will be:

$$P_{i,w}(F|S, B = 1) = \alpha + \beta S_{i,w} + E(\epsilon_{i,w}|S, B = 1) \quad (4)$$

Optimally one would independently model both the selection equation into childbirth and the probability of a female birth conditional on childbirth independently, such as through a Heckman selection model. However, this is statistically impossible in this case, since stress is both the cause of selection into birth *and* the cause of the change in the probability female conditional on birth. Since these selection models require at least one variable to determine selection into miscarriage without effecting the sex ratio at birth, it is impossible to independently identify either effect.

Our strategy in this paper therefore is a 2nd best solution. Specifically, we control for as much of the selection as possible to refine the estimates of the “correct” sex ratio at birth by parity and age by introducing a woman fixed effect into our main regression specification. This innovation, while simple, has been omitted from the literature up to this point. This will reduce the problem that women with higher levels of general health are both more likely to have higher parity births and more males, since we control for time-invariant woman-specific covariates related to stress, such as a lower genetic health endowment, lower social class, or a less privileged upbringing. Finally, we choose three indicators of maternal stress to non-parametrically estimate our  $S$  equation: parity, maternal age, and birth interval with the previous birth, such that:

$$S_{j,w} = F(A, O, I) = \psi_w + \sum_{a \neq r_a} \omega_a A + \sum_{o \neq r_o} \sigma_o O + \sum_{i \neq r_i} \pi_i I + \eta_{j,o,w} \quad (5)$$

where  $A$ ,  $O$ , and  $I$  are indicator variables which take a value of 1 if an individual was born in a certain parity, age of mother, or birth interval bin. The three summations denote that each indicator of maternal stress is estimated non-parametrically through a series of age by year, parity by parity, and interval by year fixed effects to flexibly allow for non-linearities. A reference bin is omitted for each variable group to make the estimates on the fixed effects relative to the omitted bin. For mother’s age, the omitted reference age is 25 years, while for



parity and interval the omitted groups are first-born children and a birth interval of 3 years respectively. Plugging this back into equation (4) yields our main linear probability model:

$$P_{j,w}(F|S, B = 1) = \alpha_w + \sum_{a \neq r_a} \zeta_a A_a + \sum_{o \neq r_o} \theta_o O_o + \sum_{i \neq r_i} \phi_i I_i + E(\mu_{j,w}|A, O, I, B = 1) \quad (6)$$

Note that selection is still an issue in this specification – however, there is less selection than before. This improvement on the previous literature should allow us to make better predictions regarding sex ratios in the absence of behavioral sex selection, which predictions will be used to reestimate the prevalence of missing women in the developing world. In the spirit of higher predictability, as a robustness check we also estimate a fully interacted model to capture interaction effects between age, parity, and birth interval. We do not use this as our main model because of power concerns due to the high number of bins, triple interactions, and woman fixed effects:

$$\begin{aligned} P_{i,w}(F|S, B = 1) = & \kappa_w + \sum_{a \neq r_a} \sum_{o \neq r_o} \sum_{i \neq r_i} \gamma_{a,o,i} A \cdot O \cdot I + \\ & \sum_{a \neq r_a} \sum_{i \neq r_i} \gamma_{a,-,i} A \cdot I + \sum_{a \neq r_a} \sum_{o \neq r_o} \gamma_{a,o,-} A \cdot O + \sum_{o \neq r_o} \sum_{i \neq r_i} \gamma_{-,o,i} O \cdot I + \\ & \sum_{a \neq r_a} \gamma_{a,-,-} A + \sum_{o \neq r_o} \gamma_{-,o,-} O + \sum_{i \neq r_i} \gamma_{-,-,i} I + E(\nu_{i,w}|A, O, I, B = 1) \quad (7) \end{aligned}$$

## 4 Estimating Baseline Sex Ratios at Birth

Changes in sex ratios at birth due to maternal stress are very small. In addition, in order to employ woman fixed effects one must have data on multiple births for a single woman, such that one can tease out individual effects of age, parity, and interval. Therefore, extremely large numbers of observations are needed in order to have enough statistical power to estimate such changes. In addition, in order to estimate baseline sex ratios at birth with which to compare sex ratios at birth in the developing world, one must have a setting where sex-selective abortion does not exist.

To our knowledge, no such setting exists in which there are millions of accurate, completed birth histories in a location and time with no sex selective abortion. For example, pooling DHS data gives a large number of completed birth histories, but these are from the developing world where sex preference may play a role. Scandinavian birth registry data are also available, but these countries generally have lower populations, and therefore have fewer numbers of observations. The correlational estimates above come from the US Natality files, and provide a setting where sex selective behavior is likely to be low and has cumulatively 150 million observations over 50 years, but lacks information on other births by the same woman, making woman-specific time invariant characteristics a possible confounder for age and parity effects.

For this paper, we use two sources of data to serve as our baseline. First, we use the 1990

US Census because the United States is a large, developed country, with minimal sex selective practices, but crucially since it is the last US Census to include a question on the number of children ever born to a woman. By determining if the number of a woman's children which we observe in the household is equal to the number of children the woman reports to have ever birthed, then with high probability the current household constitutes all the children ever born to this woman. Therefore, we can use the information on these children to determine the birth interval, parity, and age of the mother of each child.

Second, we use data on completed birth histories for women in sub-Saharan Africa using the Demographic and Health Surveys (DHS) to serve as an alternative, developing world alternative to our US Census data. The benefit of this setting is that it is perhaps more similar to India due to its developing country status, but yet is still a region which a minimal levels of sex-selective behavior. Since our Indian data is also from the DHS, we will postpone introducing this data until section 5 of this article.

#### **4.A 1990 US Census Data**

Data from the 1990 US Census were obtained from IPUMS USA. It contains information on each individuals the household, including their relationship to the household head, education, gender, age, race, and household income as a fraction of the poverty level. We restrict the dataset to only include households where the number of children reported as ever born to the householder (if female) or the householder's spouse (if male) is equal to the number of children observed in the household. We then assume that these children are the biological children of the head female in the household, and thereby assign each child a parity rank, birth interval (defined as the number of years since the preceeding birth), and age of mother at birth. We also delete individuals which report birth intervals outside the range of 0-20 years, children over the age of 25, and individuals whose age of mother at birth was outside the ages 15-45 in order to eliminate possible outliers or miscoded observations. This methodology yields a final dataset of 2,377,415 births across 1,209,517 women.

#### **4.B Cross-Sectional vs. Fixed Effects Results**

To demonstrate how the inclusion of mother fixed effects fundamentally changes the baseline estimates of the effect of demographic indicators on sex ratios at birth, we estimate equation (6) both with and without fixed effects and present the results graphically in Figures 12 - 14. In these figures, we plot the point estimates for each bin for indicator variable group of interest, both with and without fixed effects. For readability we do not show overlapping confidence intervals for each line. The value of each estimate represents how different the sex ratio at birth for that maternal age, parity, or birth interval is from respective reference group. For the maternal age coefficients, the reference group is 25 years old; for parity, it is first births; and for birth intervals, it is three years. Standard errors are clustered at the woman level.

Figure 12 shows the effect of parity on the probability a given birth is female. In the cross section in Figure 10, we see that sex ratios at birth tend to become more female with higher parities. However, studies which have controlled for maternal age generally find that higher order births are male biased, which is what we find in Figure 12 without fixed effects. Once we include mother fixed effects, the female biased birth order gradient returns – second births are approximately 1.5 percentage points more likely to be female, which effect persists up until parity 5, after which it becomes statistically insignificant due to exploding standard errors from a small sample size. However, the difference between the fixed effects and non-fixed parity-specific estimators remain statistically different from each other until parity 8. The without fixed effects estimators are not statistically different from zero at any parity.

Our results on parity are consistent with the fragile male hypothesis. Inasmuch as healthier women are more likely to be fecund enough to have higher birth orders of children, and also more likely to birth males, there should be a correlation between higher order births and higher sex ratios at birth. This is exactly what we find in Figure 12, where the difference between the lines rises sharply between parities 1 and 2, and increases (insignificantly) with each additional parity.

Figure 13 shows the effect on maternal age. The effect of each single year of age is estimated relative to a 25 year old mother. Interestingly, this evidence is seemingly inconsistent with the fragile male hypothesis, since the fixed effects results show an increasing birth masculinity by maternal age. This is in contrast to what we see in both the cross section and the results without fixed effects, where birth masculinity decreases with age. However, there is one very large caveat to this analysis: while the *permanent* effects of socioeconomic status are controlled for in this regression, the *transient* effects are not. Inasmuch as women have access to more resources and higher incomes as they age, we are confounding the effect of income and age. If this is the case, then it makes sense according to the fragile male theory that the effect is downward sloping – it merely implies the effect of income is stronger than the effect of age.

Figure 14 show the effect of birth interval, but only for the fixed effects model. We do not show the results without fixed effects because they are essentially identical to the fixed effects results. We find that births which have a birth interval between 0-12 months from the previous birth (including multiple births) are significantly more female than births with at least 12 months spacing. This is consistent with the fragile male hypothesis, and with the cross sectional correlation from Figure 10 Panel B.

## 5 Sex Ratios at Birth in the Developing World

We repeat the analysis from the 1990 US Census, but for countries in the developing world. For now, we focus on two regions of the developing world: sub-Saharan Africa, and India. However, in the long run we intend to do this for many countries or country groups. We chose sub-Saharan Africa because it provides a developing country analog to the US, in that

sub-Saharan Africa is a location which also presumably has no demand for sex selection. We chose India because it is a large, important country for which sex selective abortion is an issue: almost 40% of the world's total estimated missing women are from India. In addition, India is the country for which we have the most observations, with approximately 3 million births. Finally, the data span a large timespan – 1970-2016 – which allows us to show the effect of changes in demographic structure over time.

## **5.A Data**

Our data for sub-Saharan Africa and India come from the Demographic and Health Surveys (DHS). These data contain information on households, including complete birth histories for women ages 15-44 in the sample. For India, this includes data on over 1 million women and 3 million births over the four sample waves used in this study: 1992-93, 1998-99, 2005-06, and 2015-16. While these data are not panel, we are able to fix sex ratios at birth by year and maternal characteristics through the retrospective birth histories. For sub-Saharan Africa, we pool together DHS data from 100 sub-Saharan African surveys over 36 countries, consisting of a dataset with birth histories for over 1 million women and 3.5 million births between 1975-2018.

## **5.B Cross-Sectional vs. Fixed Effects Results**

Unlike the US data, changes in SRBs in India cannot be assumed to be biological effects only, since sex selective abortion is prevalent in India, particularly among higher order births. Therefore, we expect the difference between the coefficients on the fixed effects model to be a mix of biological effects and behavioral effects, which when compared to the US results, we will be able to back out behavioral sex selection in India. Similarly, sub-Saharan Africa is also a region with minimal sex selective tendencies, and therefore provides another comparison group for India.

Figures 15 - 17 repeat the analysis of Figures 12 - 14, but for India. To highlight the importance of our fixed effects model, we plot both the with and without fixed effects coefficient as before. Figure 15 shows the effect of parity on sex ratios at birth. Here the differences between the with and without fixed effects models is enormous: after controlling for age and birth interval, there is no effect of parity on sex ratios in the model without fixed effects. However, once we include maternal fixed effects it reveals large and increasing sex selection by parity. Note that in the US fixed effects model, the probability female was increasing by 1.5 percentage points for parities higher than 1. In the case of India, the probability female falls by approximately 2.5 percentage points for each additional parity. For context, this amounts to a sex ratio at birth for parity 7 of approximately 150 boys per 100 girls.

Such extreme sex ratios are not necessarily surprising for high parity births in the context of India. Jaitley (2018) shows that among children who are the last child born, the sex ratio at

birth varies from between 185 males per 100 females for children of parity one to 145 males per 100 females for parity five. In an even more extreme example, Manchanda et al. (2011) found that in a large Delhi hospital known for maternal care, SRBs varied greatly by the sex composition of children already in the household. While the overall SRB was 124:100, if the parents already had a daughter the SRB was 136:100, and was an astounding 562:100 for parents with two daughters.

Figure 16 shows the effect of maternal age on SRBs, for both the with and without fixed effects models. We find little effect of age on SRBs after age 25, and little difference between the models with and without fixed effects. However, large and significant differences rise for young women under the age of 20. In the fixed effects model, very young women have significantly lower probabilities of birthing a female, which is both statistically different from 0 and from the without fixed effects model coefficient.

Finally, Figure 17 shows the effect of birth intervals on the probability of a female birth. Unlike the US data, we find no effect of short birth spacing on birth femininity in either model, and the coefficients from both models are essentially identical.

Figure 18 repeats the analysis of Figures 12 and 15 for parity, but for the sub-Saharan African sample. We see very steep, and very sharp monotonic decreases in the sex ratios at birth as women have more children. The increase gradient with respect to parity in sub-Saharan Africa is both consistent with the fragile male hypothesis, but also with the theory that higher parities are more stressful in Africa than in the developed world due to higher levels of poverty.

Figure 19 shows the effect of education on SRBs in the cross section in sub-Saharan Africa. As before, we see a large and sharp increase in SRBs with socio-economic status, also consistent with the fragile male hypothesis.

## **6 Implications for Existing Literature**

### **6.A Refining the Estimates of Missing Women at Birth**

My results show that sex ratios at birth should be lower for women with higher birth orders, shorter birth intervals, lower socioeconomic status, and poorer health. Comparing across countries, we see that individuals in poorer areas generally have lower SRBs and higher levels of fertility. If the fragile male hypothesis is correct, the demographic structure of these nations, in addition to their poorer levels of health and lower levels of income, could be driving this correlation.

This implies that poor countries with high fertility rates should have higher rates of “missing men” – male conceptuses who would have been born, but were not because they were spontaneously aborted due to poor maternal conditions. This implies that the estimates of the “natural” sex ratio may in fact be too high, since we should see more missing men in these countries than the “natural” sex ratios suggest. As a result, the calculations of missing women

may be relying on baseline sex ratios which are already too male biased to begin with, leading to an underestimation of missing women.

To demonstrate this point, we return to the example of India. In Ray (2010), Bongaarts et al (2015) and Chao et al (2019), they take as the natural SRB the sex ratio at birth observed in India before 1970, or 105.9 males per 100 females. However, comparing this to the sex ratio at birth among Indians in the United States, yields a sex ratio at birth of 104.1, and restricting to women with less than a high school education (perhaps the most representative of the majority of Indian women in India), sex ratios at birth are merely 102.1. According to the fragile male hypothesis, the lower sex ratios are a result of poor maternal health among uneducated women in the US, who are arguably provide a better guide to the natural sex ratio among Indian women in India who may on average have even lower standards of living and worse maternal health. In addition, fertility is much higher in India than in the US, which would depress natural SRBs even more.

This suggests a series of comparisons. First, assume that the estimates from the 1990 US Census are the “true” biological effects of parity, age, and birth interval in the absence of sex selective abortion. It may be informative to import these “true” effects of age, parity, and interval to the Indian DHS observations, to come up with a predicted probability of female birth for each Indian birth in our sample. Essentially, we are asking the counterfactual of what the probability female of each Indian birth should be, if the demographic structure of India was the same as the US. We do this, and then aggregate by year of birth to find a “predicted” Indian SRB based on the “true” biological effects from the US without sex selection for the years 1970-2016.

We show these results in Figure 20. The black line shows the actual sex ratio at birth in India according to Chao et al 2019. The green line shows the counterfactual currently used in the literature to estimate missing women at birth, and the area between the green line and the black line are a representation of the current estimate of missing women at birth. The blue line shows our counterfactual estimates for India. In other words, it shows what the sex ratio at birth would be if the effects of age, parity, and interval (but not socioeconomic status) on SRBs were identical to the United States. As we can see, counterfactual SRBs in India in 1970 are much lower than there predicted baseline – beginning at 104.1 in 1970 and rising to approximately 105.5 by 2017. The reason for the increase is that during this period India underwent a demographic transition – fertility (and thereby parity) fell, birth intervals extended, and women gave births at older ages. Since these women biologically were less stressed, they should give birth to relatively more males at baseline after this demographic shift occurs.

This has two implications for the estimates of missing women. First, since demographically adjusted natural sex ratios at birth imply that there should be more “missing men” in India, the correct counterfactual against which to estimate missing women is even lower than observed. Since missing women is based on the area between the actual and the natural sex ratios at birth, a lower counterfactual natural sex ratio implies more missing women – because it now correctly

takes into account the fact that there should be more missing men.

The second implication is regarding the rate of change of sex ratios at birth. As India goes through a demographic transition, the baseline counterfactual sex ratio at birth should actually rise, not remain constant. As a result, while the *overall* number of missing women has been underestimated, the *rate of change* over time has been overestimated. Some of the increase in the sex ratio over time should happen naturally, as fertility falls, age of mother at birth rises, and birth intervals lengthen, all implying less maternal stress and a higher natural rate of male births.

This is not to say that adjusting natural sex ratios at birth for demographic change explains away a majority of the increase in sex ratios at birth. That sex selective abortion, infanticide, and other sex-selective behaviors are severely skewing sex ratios at birth in India is undeniable given the preponderance of evidence for it. However, by our estimates about 20% of the change in the sex ratios at birth observed from 1970 to 2015 has been due to natural demographic change. This is represented in Figure 20 by the red line, which is simply the actual sex ratios at birth (black line) minus the estimated demographically-induced change in the natural sex ratio at birth (blue line). Therefore, the residual red line can be interpreted as the effect of induced abortion only on sex ratios at birth. As can be seen on the figure, changes in sex selective abortion behavior represented by the red line form the vast majority of the increase in the change in SRBs since 1970.

## **6.B Sex Ratios at Birth and the 1.05 Biological Constant**

The previous literature has assumed 1.05 is a biological constant, conditional perhaps on some small racial variation. My paper shows it is not that simple – there may be a biological constant sex ratio, but it must be conditional on maternal stress.

### **6.B.1 Differences by Race**

The current consensus is that race is the only factor by which sex ratios consistently vary, and that sex ratios by racial group are fixed. However, after controlling for socioeconomic factors and indicators of maternal stress, I show that 25% of the racial differences in sex ratios go away. Therefore, while differences in natural SRBs across racial groups play a large and important role in natural SRBs, it is less than otherwise believed.

### **6.B.2 Maternal Stress**

I have shown that SRBs strongly vary with socioeconomic status and maternal health, contrary to the current consensus.

## 6.C Sex Preference in the US and the Demand for Sons

Dahl and Moretti (2006) find that women in the US who have girls are worse off in a number of SES categories. Since their theoretical framework assumes that having a girl random and independent of SES, the gender of a birth is a good natural experiment to determine whether husbands prefer boys. This paper turns their argument on its head: rather than a female birth causing bad outcomes for women, this paper suggests that women with bad outcomes are more likely to have girls. As a result, the results from Dahl and Moretti (2006) are likely a combination of the mechanism they suggest, as well as the reverse causality of outcomes on birth femininity.

We are not the first person to suggest this reverse causality. Amar Hamoudi and Jenna Nobles (2014) noticed that if divorce was a source of maternal stress, this could lead divorce to cause the birth of girls, rather than the birth of girls to cause divorce. This article simply builds upon this theoretical idea, but in a large dataset with many indicators of maternal stress, rather than focusing on divorce alone.

## 7 Conclusion

Estimates of the number of missing women rely critically on correct estimates of the biological sex-ratio at birth. Generally, sex ratios at birth (SRBs) in the absence of sex-selective induced abortion are believed to be a biological constant of around 105 males per 100 females, with small racial variations. However, this literature is based on small and underpowered samples, and ignore possible in utero male fragility, where male conceptions are disproportionately terminated in the presence of maternal stress or poor health. Using 150 million birth observations over 50 years in the United States – a country with presumably low levels of sex-selective abortion – we document sizeable and highly significant correlations between proximate determinants of spontaneous terminations and sex ratios at birth, including poverty, age, parity, educational group, previous birth interval, and smoking, even after controlling for racial differences. We show that controlling for maternal stress explains – or even overturns – many commonly held beliefs about natural sex ratios at birth, such as the downward cross-country relationship between fertility rates and sex ratios at birth, the presumed US demand for sons, and 25% of the disparity in the assumed biological racial differences in natural SRBs. Since the fragile male hypothesis predicts that healthier women are more likely to give birth to more children, more males, and be able to conceive in older ages, we correct the age and parity specific estimates in the literature by employing a woman fixed effect model to control for time-invariant woman-specific differences in health and SES. Using the resulting selection adjusted estimates, we show that there should be more "missing men" at birth than currently observed in many developing nations, leading to the number of missing women being underestimated by about 30%. Finally, we estimate that 20% of the increase in sex ratios at birth over the past



50 years are naturally due to the demographic transition, and would have occurred even in the absence of sex-selective induced abortion.

## References

- [1] Seema Jaychandaran AEJ:AP
- [2] Debraj Ray RES
- [3] Buckles and Hungerman
- [4] Bongaarts and Guilmoto 2015 PDR
- [5] Chao et al 2019 PNAS
- [6] Everything Bruckner and Catalano have ever done
- [7] Lots of stuff by Doug Almond, 2008 PNAS, 2013 PDR
- [8] Amar Hamoudi Jenna Nobels Demography
- [9] Dahl and Moretti RES
- [10] Guilmoto 2009, 2013, 2015 review article and others
- [11] William James, lots of reviews
- [12] Wilde, Apouey, and Jung 2017 EER
- [13] Wilde, McLaren, Chen and Lohman, Working paper
- [14] Trivers, Robert L., and Dan E. Willard. 1973. "Natural Selection of Parental Ability to Vary the Sex Ratio of Offspring." *Science*, 179(4068): 90–92.
- [15] Sen, Amartya. 1990. "More than 100 Million Women are Missing." *New York Review of Books*, 37(20): 61–66.
- [16] Sen, Amartya. 1992. "Missing Women." *British Medical Journal*, 304: 587–88.
- [17] Abrevaya, Jason. 2009. "Are There Missing Girls in the United States? Evidence from Birth Data" *American Economic Journal: Applied Economics*, 1(2): 1–34
- [18] IPUMS USA data
- [19] DHS data
- [20] US Natality file data
- [21] Jaitley (2018) *Economic Survey 2018: Gender and Son Meta-Preference: Is Development Itself and Antidote?*
- [22] Manchanda, S., Saikia, B., Gupta, N., Chowdhary, S., and Puliyeel, J. M. (2011). Sex ratio at birth in India, its relation to birth order, sex of previous children and use of indigenous medicine. *PLoS One*, 6(6), e20097.
- [23] Sanders, N. J. and Stoecker, C. (2015). Where have all the young men gone? using sex ratios to measure fetal death rates. *Journal of health economics*

- [24] Halla, M. and Zweimüller, M. (2013). Parental Response to Early Human Capital Shocks: Evidence from the Chernobyl Accident. Discussion Paper 7968, Institute for the Study of Labor (IZA).
- [25] Breakfast skipping, extreme commutes, and the sex composition at birth B Mazumder, Z Seeskin - *Biodemography and social biology*, 2015 - Taylor & Francis
- [26] <https://onlinelibrary.wiley.com/doi/full/10.1002/ajhb.23074>

Figure 1: Figure on sex ratios at birth and fertility rates in the cross section

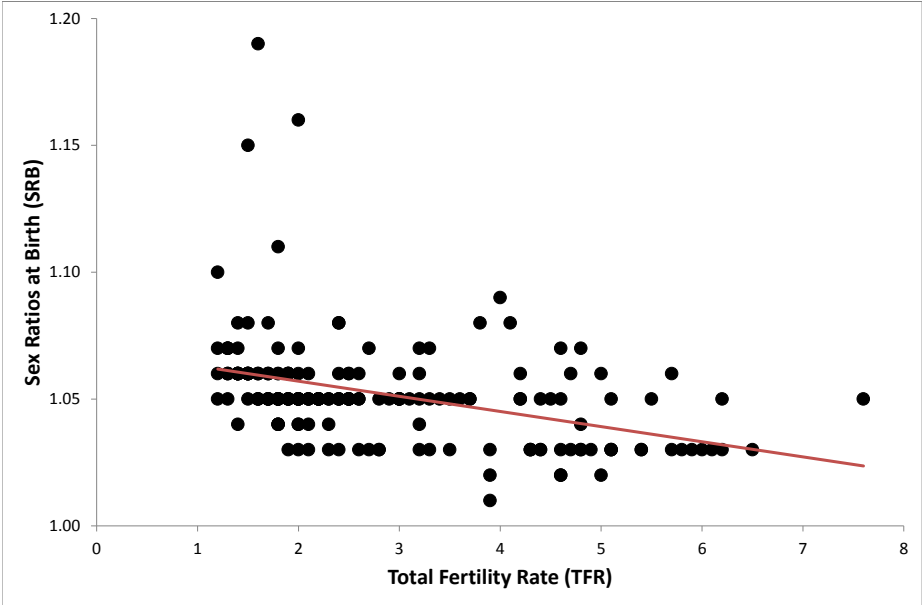
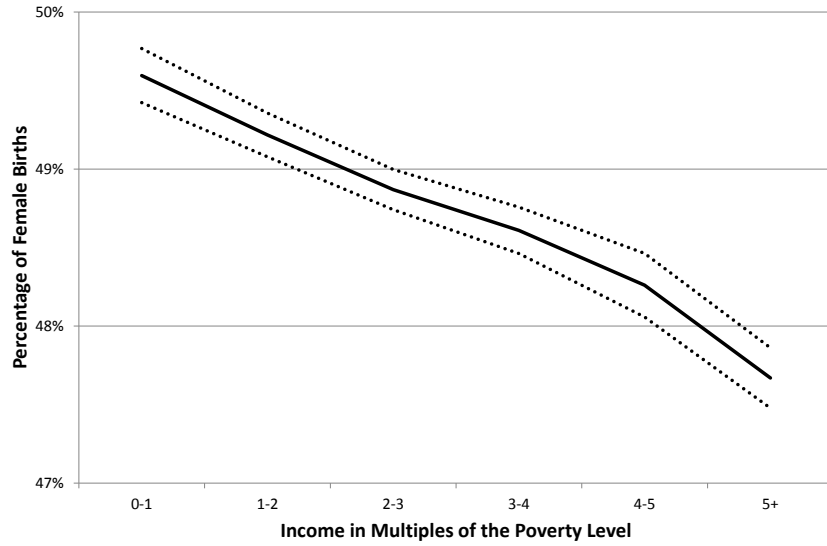


Figure 2: Cross-Sectional Relationship Between Mother’s Socioeconomic Status at Time of Survey and Sex-Ratios at Birth



Panel A: Income

Panel B: Education

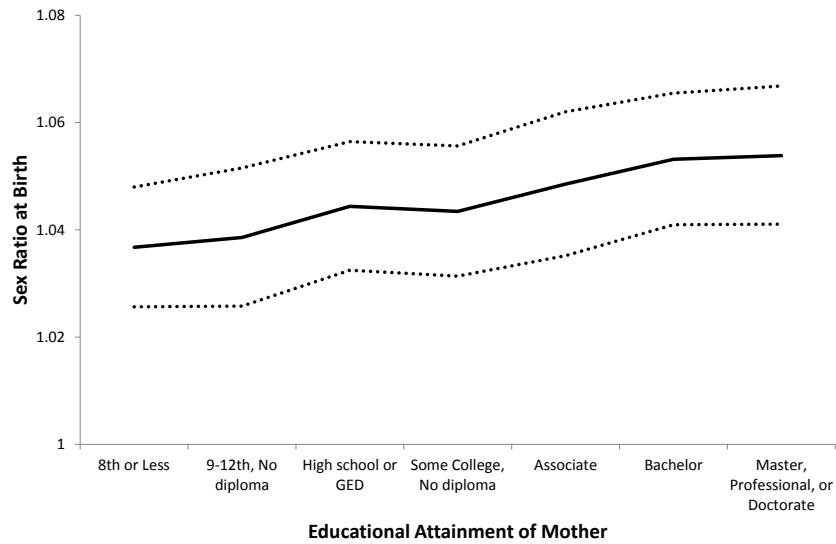


Figure 3: Cross-Sectional Relationship Between Race, Income, and Sex-Ratios at Birth

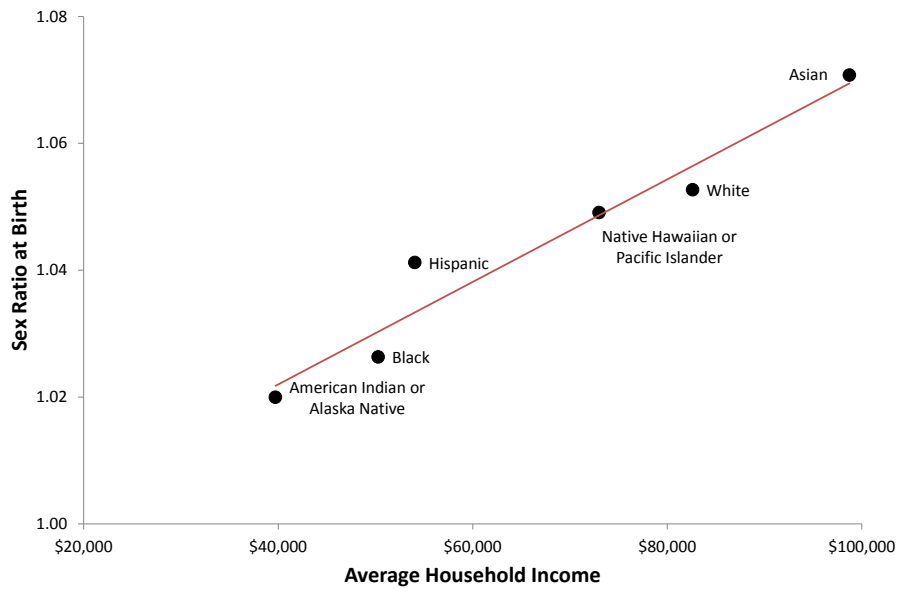


Figure 4: Cross-Sectional Relationship Between National Sex Ratios at Birth and Income

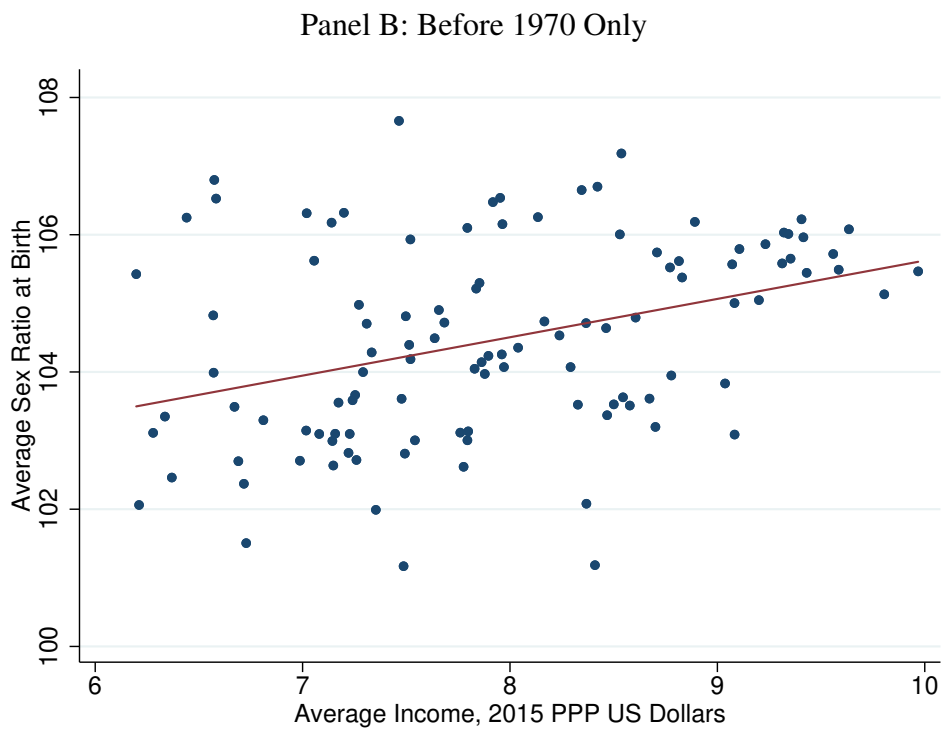
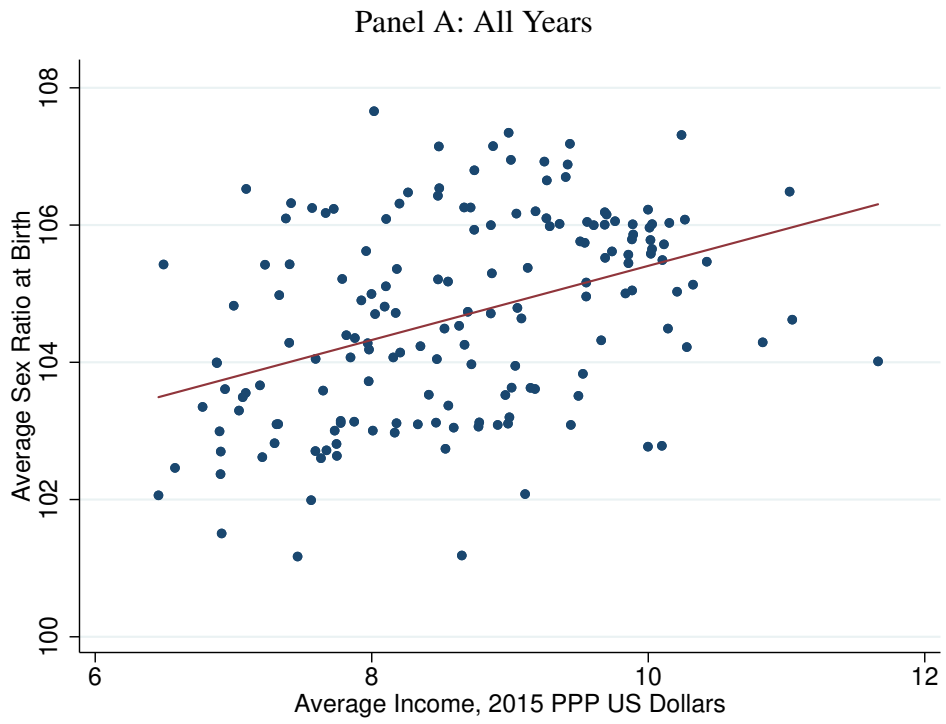


Figure 5: Cross-Sectional Relationship Between Month of Birth and Sex-Ratios

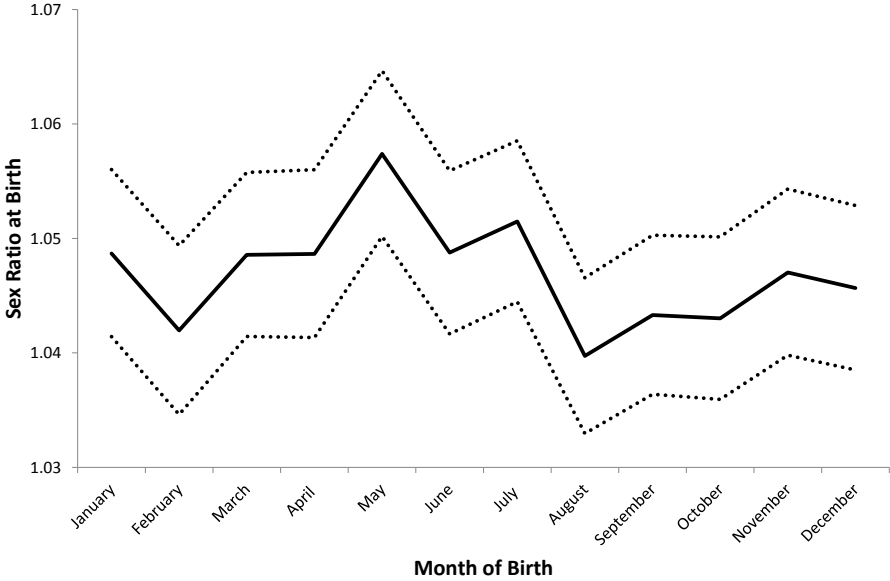
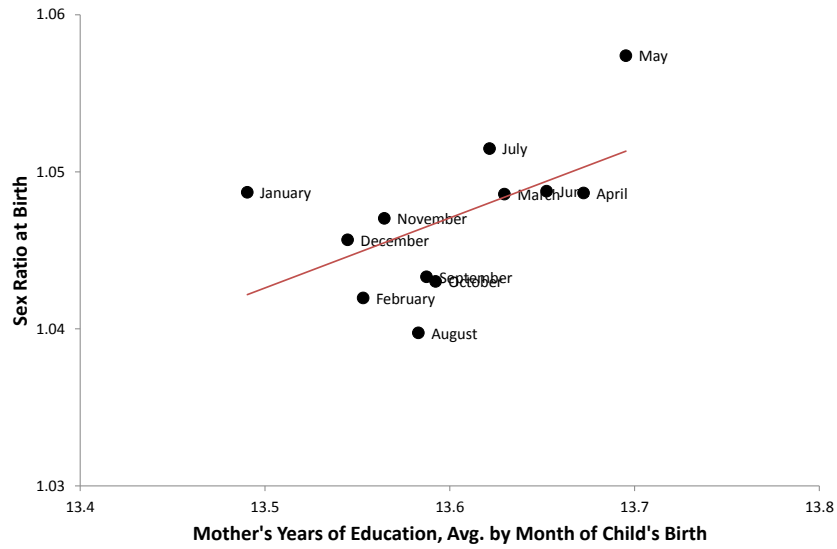




Figure 6: Sex Ratios at Birth by Month: Correlates with Education and Temperature at Conception

Panel A: Education



Panel B: Temperature 9 Months Before Birth

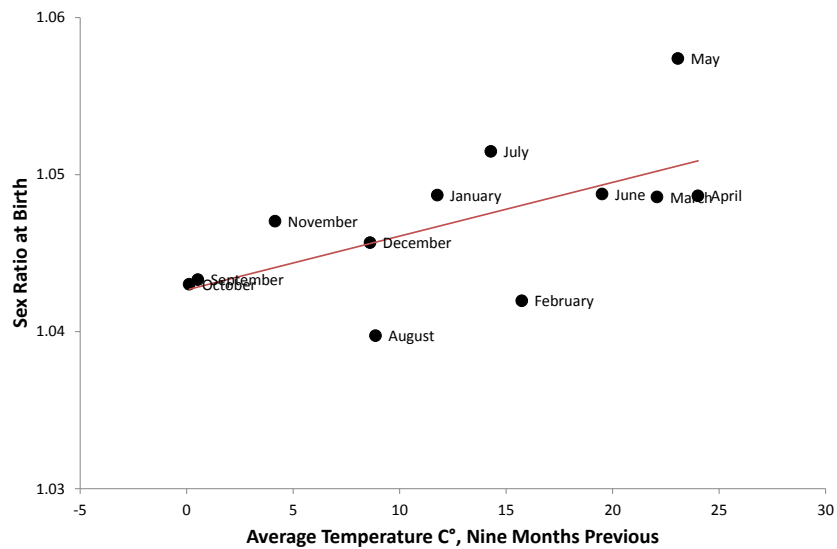


Figure 7: Cross-Sectional Relationship Between Day of Week of Birth and Sex-Ratios

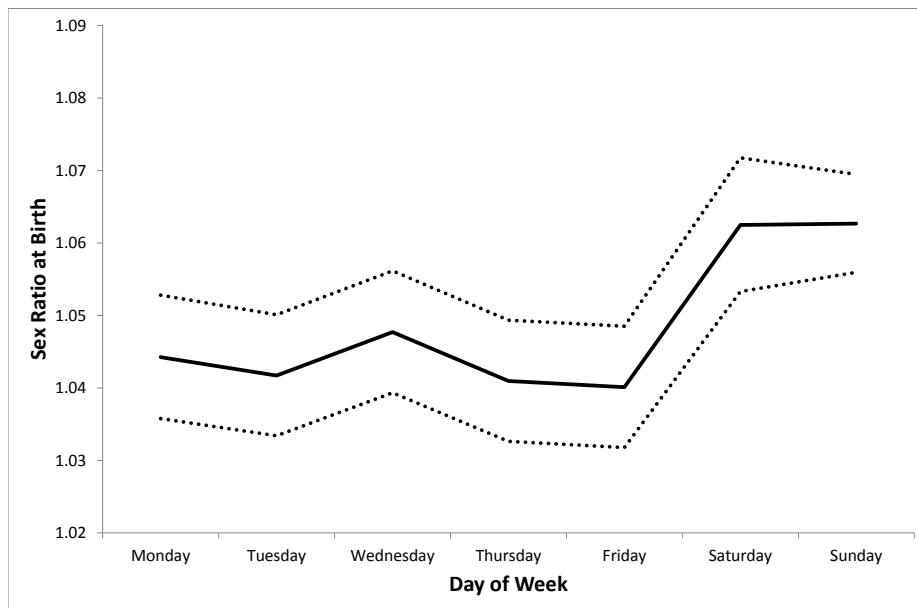
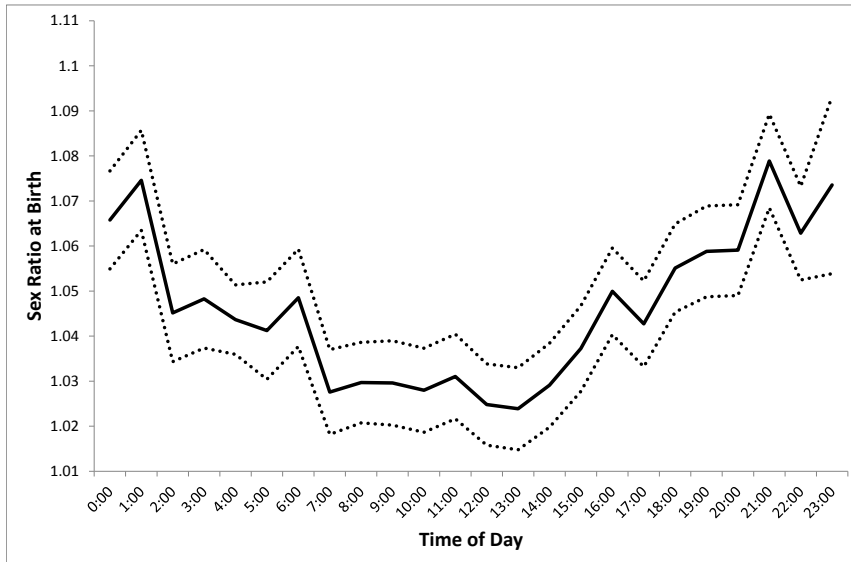


Figure 8: Cross-Sectional Relationship Between Time of Day of Birth and Sex-Ratios

Panel A: Weekday Births



Panel B: Weekend Births

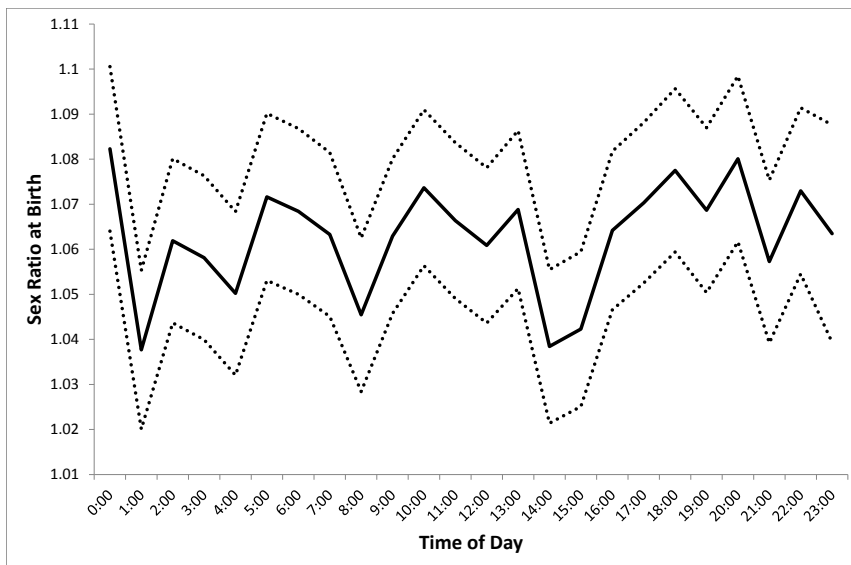


Figure 9: Racial Differences in Sex Ratios at Birth by Birth Timing

Panel A: Time of Day, Weekdays Only

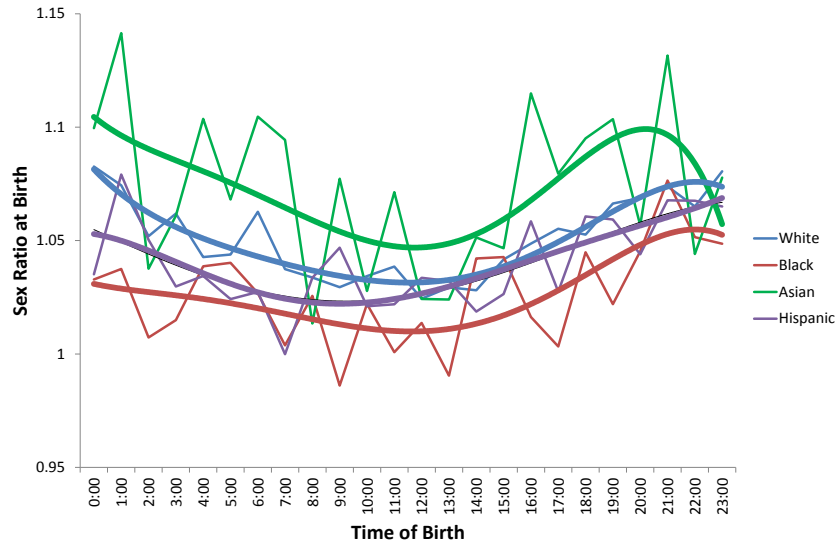
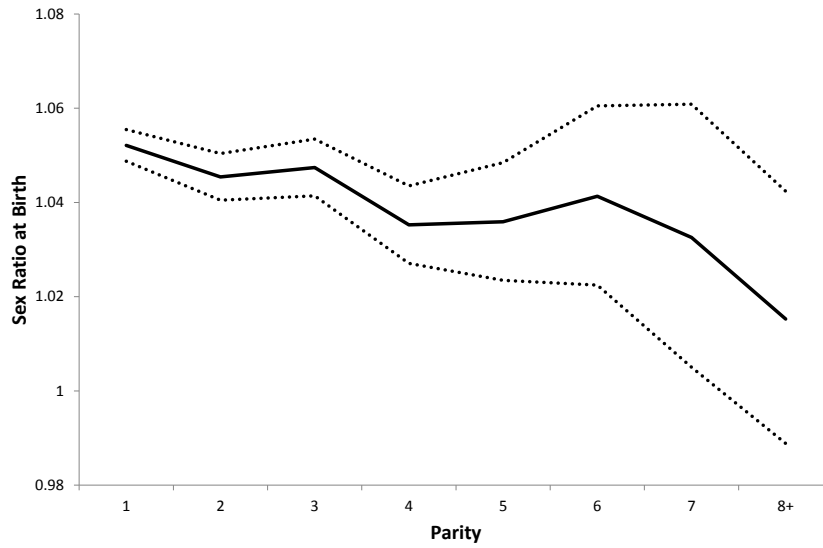


Figure 10: Cross-Sectional Relationship Between Indicators of Maternal Depletion and Sex-Ratios at Birth



Panel A: Parity

Panel B: Interval with Previous Birth

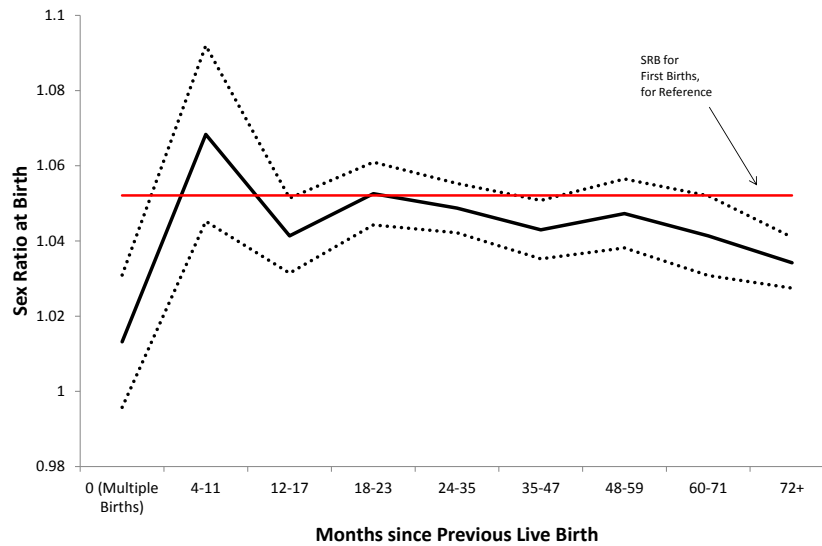


Figure 11: Cross-Sectional Relationship Between Age and Sex-Ratios at Birth

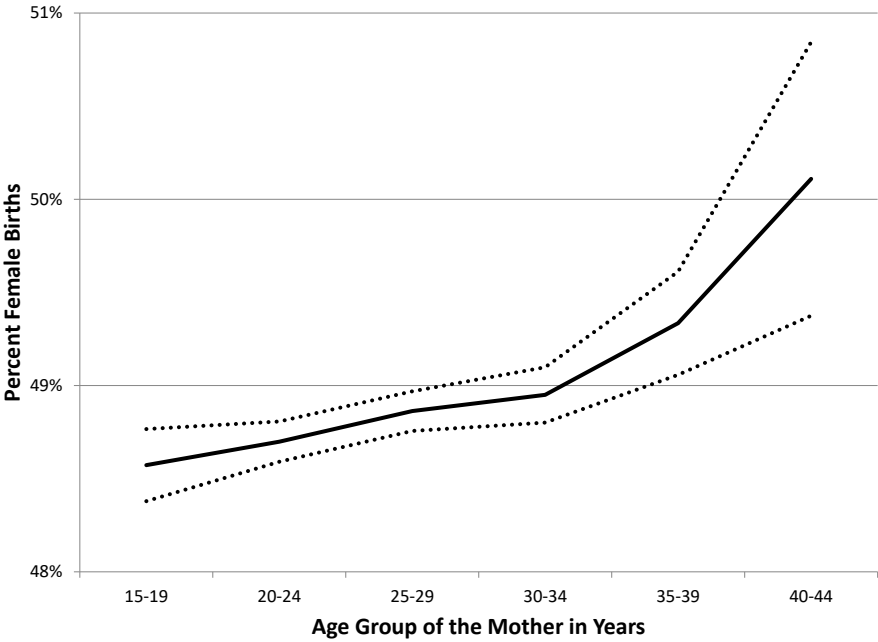


Figure 12: The Effect of Birth Order: With and Without Fixed Effects

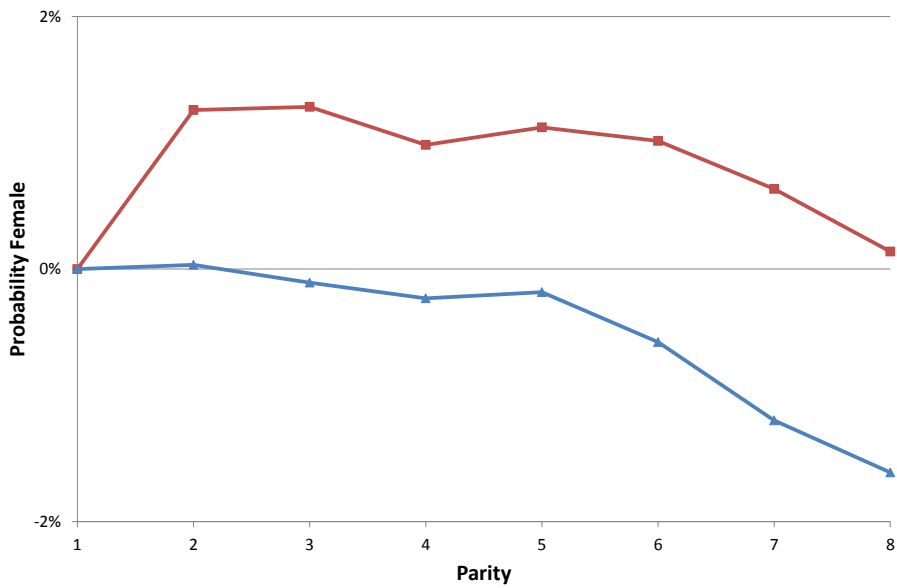


Figure 13: The Effect of Mother's Age at Birth: With and Without Fixed Effects

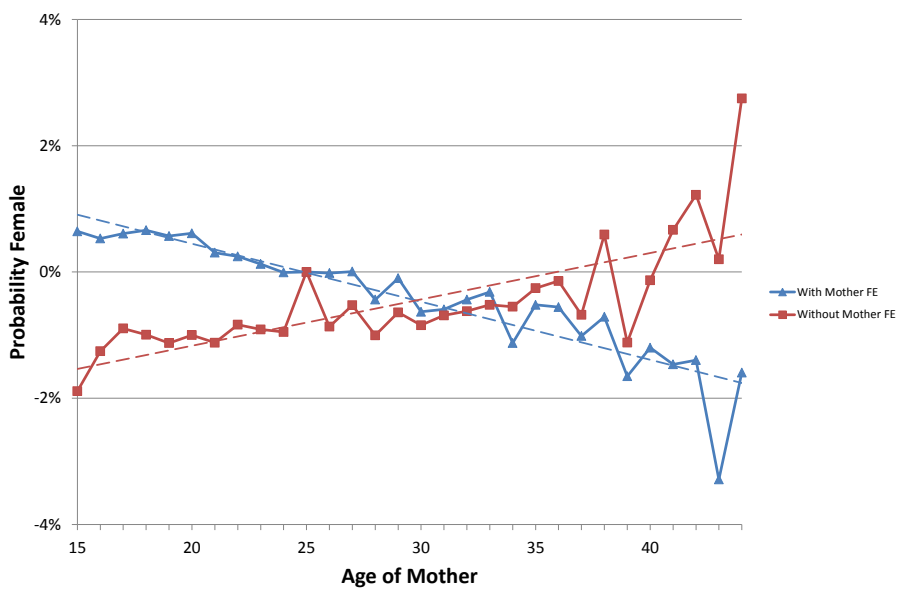


Figure 14: The Effect of Birth Intervals: With Fixed Effects

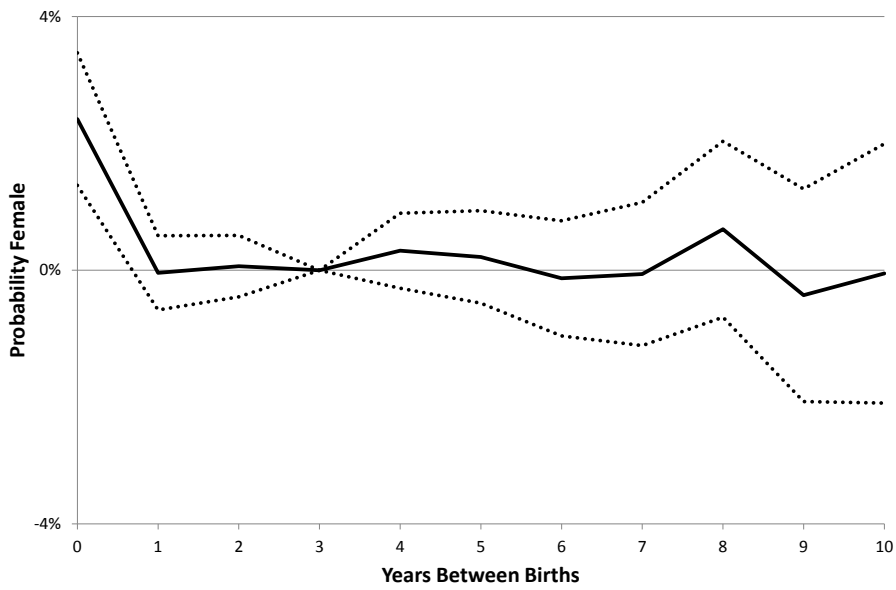


Figure 15: Parity – India DHS Sample

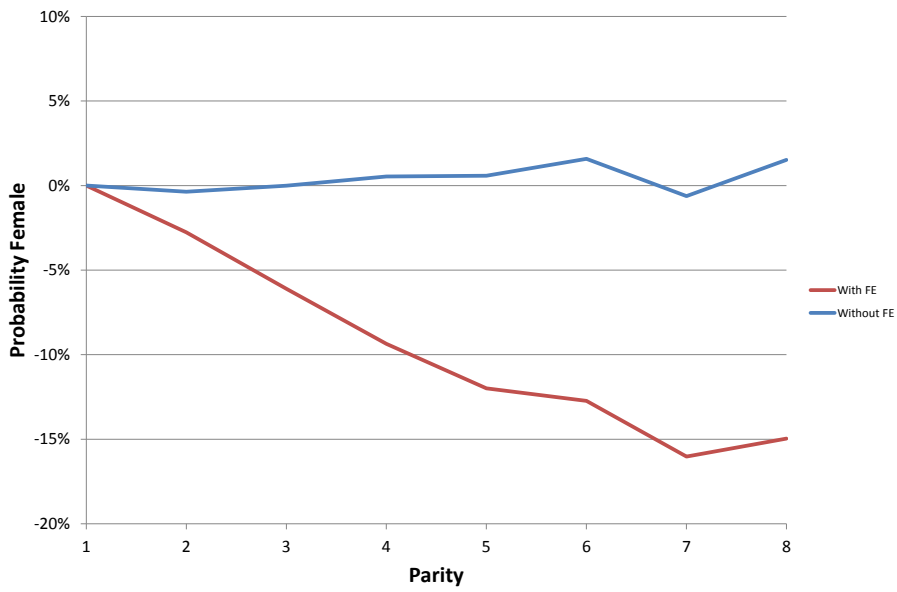




Figure 16: Age – India DHS Sample

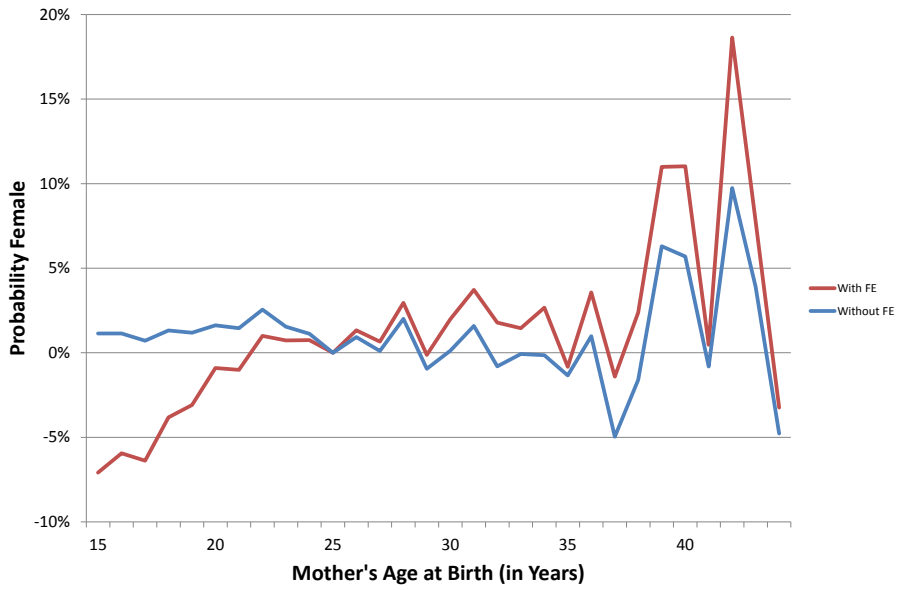


Figure 17: Interval – India DHS Sample

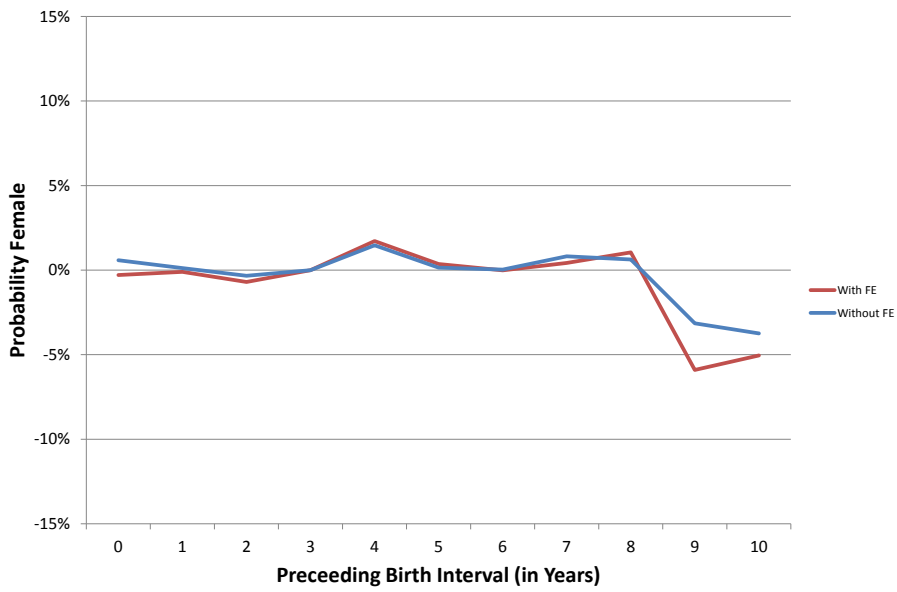


Figure 18: Parity – Sub-Saharan Africa DHS Sample

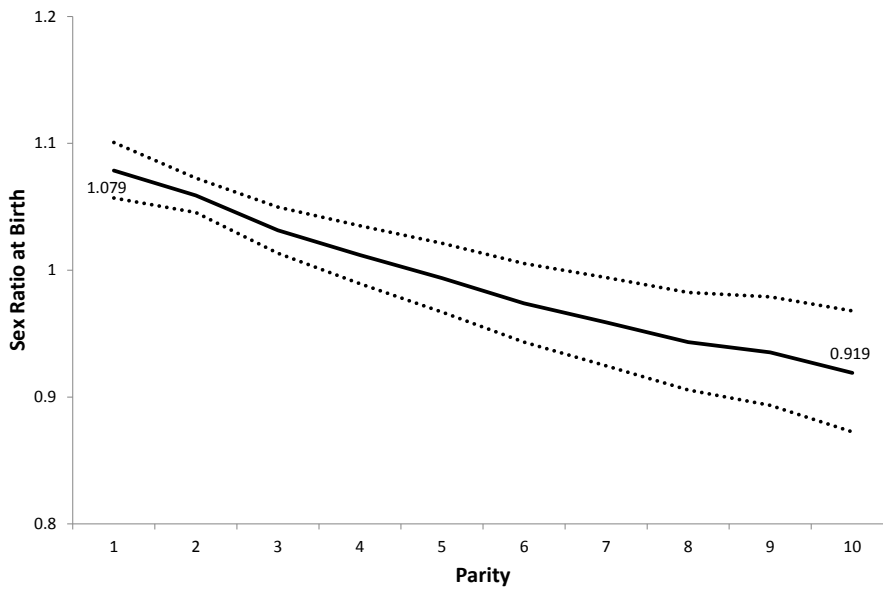


Figure 19: Education – Sub-Saharan Africa DHS Sample

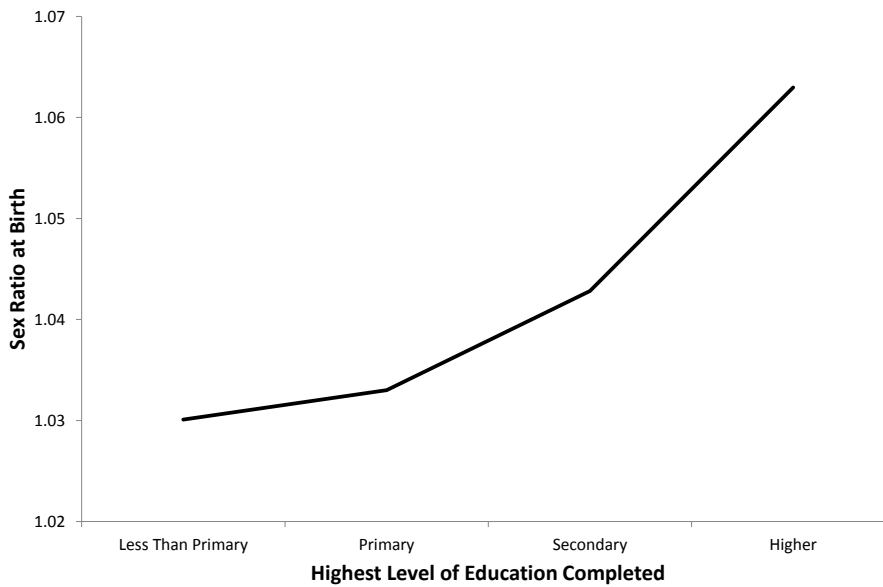


Figure 20: Sex Ratios at Birth In India, 1970-2017: Actual, Assumed, and Implied

