Queen Mary
University of London

ECN325 Economics Project II
2018/19

Labour markets and parents' gender preferences: Infant health care in Mexico

By: Buthaina Almadhi

Supervisor: Dr Francesca Cornaglia

School of Economics and Finance

Table of Contents

1. Abstract

Background: Gender differentials against females in health care and nutrition have, with time, resulted in high female mortality rates and the missing women phenomenon. The extent of gender biases in parental investments in infants have been largely addressed in the context of countries where gender inequality prevails in many aspects of a females lifetime, with great focus on India. However, it has scarcely been addressed in a context where the gender gaps within a country are not so pronounced and apparent.

Methodology: We test for gender preferences in parental provisions of nutrition and health care to infants in Mexico. Using the Mexican Family Life Survey, we check if parents alter the vaccination provisions and breastfeeding patterns for their infants solely based on their gender.

Results: No significant relations were found between being fully vaccinated and the child's gender in Mexico. Similar results were found for owning a vaccination card, ever being breastfed and breastfeeding durations.

Conclusion: The study suggests an initial indication of gender parity within the study of gender inequality in nutritional and health care provisions by parents in developing countries that show less evident signs of gender inequalities in other areas such as the labour markets.

2. Introduction

As of 2000, there were approximately 1.5 million missing women in Sub-Saharan Africa and 1.7 million missing women in India and China each (Ray, 2010). The missing women phenomenon was introduced by Sen (1990) and is agreed to be existent due to gender differential treatments against women. Researches have been able to prove gender differentials through all the stages in the lifetime of a female, with some cases starting even before birth through sex-selective abortion. Once the child is born, they also go through more gender discrimination from their parent's investments in their nutrition, healthcare, childcare time and education. A factor which influences parents to develop gender preferences among their children could be thought of in the economical context of how much return the parents would get from a child when they grow up. A utility maximization model can be used to represent the factors which affect parent's investments in sons or daughters. The factors influencing investments based on gender include the cost of investment, rate of return on the investment and the wages for either gender in the labour market. In countries with high gender differentials in the labour market, we see parents developing a gender preference towards the gender favoured in the labour market, as they are more likely to get a job and care for the parents in the long run. Many evidences of gender biases in parent investments have been reported particularly in developing countries with high gender discrimination in the labour force such as India, China and Bangladesh. Nevertheless, there is very limited research on gender preference amongst parents in countries with less evident gender biases in the labour market. This study aims to fill the gap in the literature by testing for gender differentials in parental investments in a country with less apparent gender inequality in labour market, thus allowing us to test whether the future value of an individual affects the parent's investments in them. Mexico was chosen in this research as it fits the criteria of a developing country with less gender discrimination in the labour markets compared to India, China and Bangladesh. By testing for gender differentials in investments in vaccinations and breastfeeding patterns by parents in Mexico, we will be able to either support the theory of investments depending on future return achievable by the child, or provide proof against it. To the best of my knowledge, no research has taken place before to check for gender differentials in vaccination attainments in Mexico.

2.1 Aims and Objectives

The aim of this study is to further understand the relation between a child's gender and their parents' investments on the health of that child, particularly in a country with less apparent gender differences to be faced in the future of a child. Another aim of this study is to help fill the gap in the literature regarding gender differentials in vaccination is Mexico. These aims will be achieved by adopting the methodology drawn by Barcellos and colleagues (2014), but instead used to test for gender differentials in parental investments in Mexico. By running regressions to estimate the effect of gender on completing vaccination schedules and breastfeeding patterns, we will be able to check if health and nutritional provisions are sacrificed due to the gender of a child in Mexico.

The rest of the paper is structured as follows. It begins with an insight into the decision process that parents go through when deciding how much to invest on their child using the utility maximising theory. This will help in understanding why parents may develop gender preferential between children. The next step covers the existing literature regarding gender differentials in parental differences in countries with very apparent gender biasness in their respective labour markets, with a large focus on India. This is followed by an overview of the literature within this field in Mexico and reasons that make Mexico a good country for this research. After theoretically contextualising the topic of gender inequality in parental investments, the data set that will be used in this paper will be described in detail followed by the methodology that is adopted to answer the question of the existence of gender discrimination in parental investments in Mexico. Finally, the paper presents the results found along with a discussion of the results and possible limitations confronted. Possible areas for further research are also suggested.

3. Literature Review:

Living in the 21st century, many programmes have come about to target gender inequalities in employment, education and health. Such programmes include GEWE (Gender Equality and Women's Empowerment Programme) by the United Nations (MDG Fund, n.d) and the Gender Equality Programme targeting the garment industry in India by the British High Commission (GOV.UK, 2017). However, despite these efforts, gender inequality has been proven by the Global Gender Gap Index to be increasing in 82 countries around the world (World Economic Forum, 2017). This is a growing issue due to its increasing negative impact as gender inequality can start by taking effect in micro-households and eventually grow to affect macro-economic factors. To understand the magnitude of its negative impact, we consider the following example: Imagine individual households that prefer their sons over their daughters. Due to such gender preference, the parents' investments on their daughters are relatively more fluctuating in response to shocks in the family's income and hence the daughters welfare is sacrificed (Rose, 1999). The continuation of such events disturbs the nutrition and health care provisions for females and with time leads to higher female mortality rates (Waldron, 1987; Visaria, 1985; Basu, 1989; Sen and Sengupta, 1983; Kishor, 1993). Further spiraling of this example in real life has resulted in the "missing woman" phenomenon (Milazzo, 2018). This phenomenon addresses imbalances in female to male ratios where there are fewer women than men in countries (Gupta, 2005). Another factor contributing towards the 'missing women' phenomenon is the use of technologies which help identify the gender of a baby during the prenatal periods, thus encouraging sex-selective abortion (ibid). Having fewer woman in the world is a problem because the world would not have unlocked the full potential of its population. It is estimated that global GDP can be increased by $12 trillion if gender gaps are closed and the world worked towards equality (Woetzel et al., 2015).

3.1 Parental investment decision making process

Understanding what encourages parents to gender discriminate in their investments can be done with the help of the utility maximization theory and a zero-sum allocation process (Beaulieu and Bugental, 2008). By looking at parental investments as a zero-sum allocation process, we are referring to the total amount available for the parents to invest on children. By investing more on one child, the parents have less to invest on the other (Trivers, 1974). Therefore, parents will naturally try to maximise the gains from their investments and this process can be understood using the following general utility maximising model. We adopt the utility maximization model used by Pasqua (2005) to portray the decision making process of parents. Although Pasqua's model is to test for gender differential treatment in education, we adapt it to health care by replacing investment in education with investment in healthcare provision in the utility function. Under the utility maximization theory, parents would aim to maximise the utility gained from their spending, given their income constraints.

$$U^p = \max_{c_1^m, c_1^f, c_2^m, c_2^f, x_1^b, x_1^g} U^p(c_1^m, c_1^f, c_2^m, c_2^f, x_1^b, x_1^g)$$

Subject to: $M = I^i + w^{-1}(1 + n^i x_0^{-1})$

Where c_t^i is the consumption of the parent i (male or female) in period t (1 or 2). x^b is their investment on boys and x^g is the investment on girls. M is their budget from both labour income ($w^{-1}(1 + n^i x_0^{-1})$) and non-labour income (I^i) gained. By maximizing the utility function subject to the income constraints, Pasqua arrived at the following equation which determines the parents expenditure on boys and girls:

$$\frac{x_1^g}{x_1^b} = \frac{p_1^b - \lambda^b w^{-m} n^b}{p_1^g - \lambda^g w^{-f} n^g} \frac{\varepsilon}{\delta}$$

This equation shows that the parents investment in girls or boys is affected by p_1^b and p_1^g which refer to the cost of the investment for a boy or a girl. In the context of this study, health care systems set the same cost of healthcare for boys and girls for the same health services. Parents investments in boys or girls also depend on the rate of return of providing healthcare for a boy and a girl, given by n^b and n^g. When boys and girls are equally cared for, girls survival rates are said to surpass boys' (Sen, 1990), thus indicating that the rate of return (n) from health care is higher for girls than boys. Another factor influencing parental investment decisions is the wages of females and males in the labour market, given by w^{-m} and w^{-f}. If the wages of males are higher in the labour market, the investment in boys by their parents is also likely to be higher. Overall, such models suggest that if the child's earning opportunities are low in their adult life, parents would adjust their investments on the child accordingly. This prompts us to look at the inequalities faced by a child in later points in time to help explain the inequalities they face at earlier stages of their life.

3.2 Evidence from a country with high labour market gender biases: India

One of the countries with most research indicating gender inequality and son preference is India. The existence of gender inequality in parental investments has been noted in India where female infants were less likely to get vaccinated than male infants (Borooah, 2004) and breastfeeding durations were lowest for female infants (Jayachandran and Kuziemko, 2011). Gender differentials favouring boys in childcare time and vitamin supplementation have also been documented (Barcellos, 2014). There has also been proof by Rose (1999) that the resources allocated to girls are more responsive and fluctuate more to exogenous shocks to a family's income as compared to resources allocated to boys. For example, when a family goes through a crisis and cannot afford food for everyone, they tend to sacrifice the welfare of their girls first. As expected, these factors combined have contributed to 1.7 million missing women in India (Ray, 2010).

By considering the labour force participation rates between males and females in India, we see gender inequality prevailing. 82% of the working-age men population in 2008 were part of the labour force compared to only 32% of women population (Table 1). Moreover, female workers are paid a lower wage rate than their male counterparts in each employment category as per 2011-12, with the overall raw wage gender gap being 34% (International Labour Organization, 2018). For a parent in India, this means that their son has higher chances of not only getting a job but also receiving higher income than a daughter, thus encouraging parents to invest more on sons than on daughters. In fact, the effect of the labour market on parental investments in India has been studied by Rosenzweig and Schultz (1982) where they explained that child mortality is related to the parent's expectation of how 'economically productive' the child will be when he/she grows up as parents consider this factor when trying to optimise the household's resource allocation process. Tracing back through the life span of children in India to check if this truly is the case, we consider the youth literacy rate amongst 15-24 year old boys and girls. It is found that while 88.4% of youth males are literate in India, only 74.4% of youth females are literate (UNICEF, 2013). This coincides with the expected flow of events as it shows that parents are more likely to educate their sons than their daughters in India where sons are more 'economically valuable'. This helps explain all the previously mentioned evidences of gender differentials favouring sons in parental investments on childcare time, vaccinations and even breastfeeding patterns.

Interestingly, not all literature in India shows gender preferences in parental investments. A publication by Ryan (1982) found that boys were the ones with the disadvantage in smaller households with limited resources in India. Another paper by Basu (1989) was not able to find large variances in receiving immunizations amongst females and males. It is noticed that the findings on gender discrimination depend largely on the region in which gender-discrimination is being tested. This is perhaps due to the structure of immunization programs in India where larger decision are taken on a national scale but the implementation is left for state-level bodies to decide on, thus resulting in different gender-bias outcomes across states and regions (Pande & Yazbeck, 2003). Another observation is found in a book review by Poffenberger (1982) which summarises B.D.Miller's book "The endangered sex" where she noticed that the differential in preferences may arise due to girls in the south being more 'economically valuable' as they help with agriculture when compared to North. For whichever reason, it is generally observed that the girls in North of India face more gender-discrimination than those in the South of India (Basu, 1989).

Other countries with reported gender preferences include China where female children have a lower likelihood of getting the necessary immunizations and this likelihood is further reduced if they have an older sister but not an older brother (Li, 2004; Graham, Larsen and Xu, 1998). In South Korea (Choi and Hwang, 2015) and Bangladesh (Chen, Huq and D'Souza, 1981) as well, further evidences are found of parental investments being limited because the child is a female. In all of these reported cases, gender gaps in the labour market are evident as seen in table 1 and son preference is reported.

Table 1: Gender discrimination in labour force participation rates between males and females in 2008

Country	Male labour force participation rate (2008)	Female labour force participation rate (2008)
India	82%	32%
China	79%	65%
Bangladesh	85%	29%

Sources: https://data.worldbank.org/indicator/SL.TLF.CACT.MA.ZS?locations=IN-CN-BD

: https://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS?locations=IN-CN-BD

Nonetheless, there are countries where it is agreed that gender-differentials are favouring girls over boys such as in Nigeria, where female children were reported to have higher chances of receiving full immunizations rather than male children (Antai, 2012). There have also been interesting research showing daughter-preference in Japan (Fuse, 2013). In addition, despite it being evident that there are more reported cases proving son preferences across various regions as opposed to daughter preferences, a policy paper by Duflo (2005) has contradicted the aforementioned literature by expressing that even when son preference is greatest, there is difficulty in finding gender inequality where girls are less cared for. In fact, Duflo was not the only one to find that the gender of a child has no effect on the parental investments and care towards the child. Angus Deaton (1989) found no evidence of gender bias favouring sons in Cote d'Ivoire in Africa and statistically insignificant evidence of gender preference in Thailand.

3.3 Evidence from a country with lower labour market gender biases: Mexico

After reviewing the relevant literature for countries with pronounced gender inequality in labour markets, the focus is now shifted to the literature for countries with less apparent, yet still existent, gender-biased labour markets. Research in the field of gender bias in parental investments is very limited in such countries. This project aims to fill the gap in the literature by testing for gender differentials in parental investments in nutrition and health care in a country with less observable gender gaps.

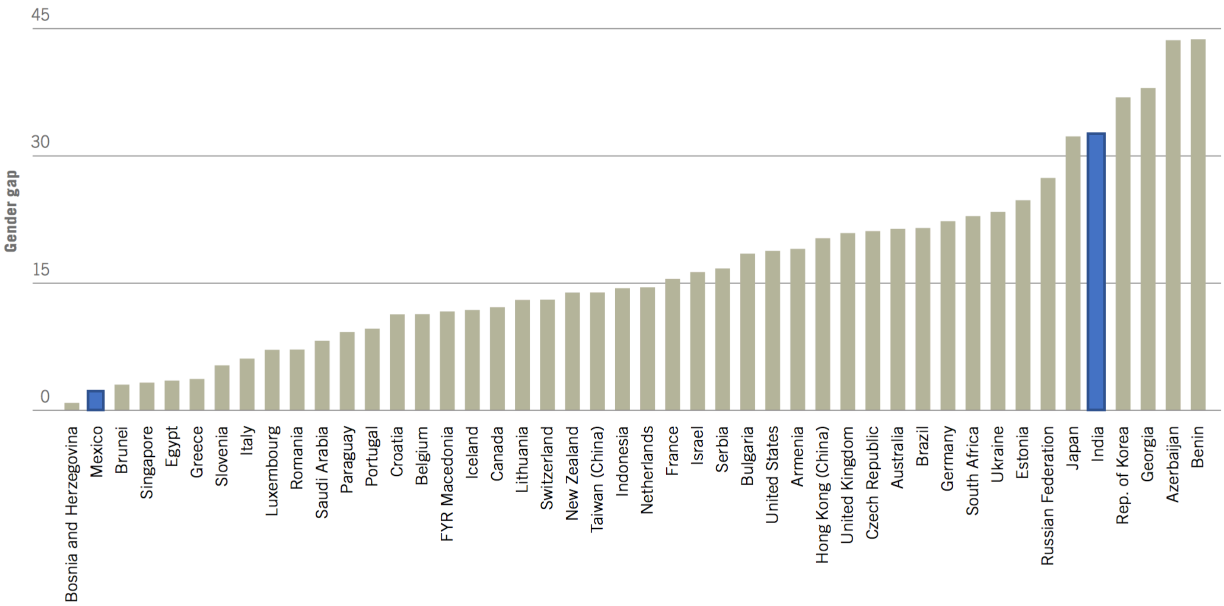
Mexico was chosen as the ideal country for this research as the once heightened levels of social inequality in Mexico allowed it to be a good representation of "the gamut of health problems affecting the world" but has recently managed to improve its healthcare system to overcome many of the issues (Frenk, 2006). It is also a good country to consider as it has gender biased approaches in its labour market, however less evident than those found in India and no gender gaps in its youth literacy rates compared to India.

Table 2: Difference in gender inequalities for various indicators in India and Mexico		
Indicator	India	Mexico
Labour force participation rate – % of the **male** population in 2008	82%	81%
Labour force participation rate – % of the **female** population in 2008	32%	43%
Youth literacy rate (15-24 years) (2008-2012) **male**	88.4%	98.4%
Youth literacy rate (15-24 years) (2008-2012) **female**	74.4%	98.5%

Sources: https://www.unicef.org/infobycountry/india_statistics.html
https://www.unicef.org/infobycountry/mexico_statistics.html
https://data.worldbank.org/indicator/SL.TLF.CACT.MA.ZS?locations=MX-IN
https://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS?locations=MX-IN

As seen in table 2, gender inequality does exist in Mexico's labour force where 81% of the men working-age population is participating in the labour force compared to 43% of the female working-age population, but is less apparent than the gender inequality present in India's labour force participation rates. Moreover, graph 1 shows that Mexico has a much lower gender gap in wages in their labour market:

Graph 1: Graphical illustration of the gender gap in hourly wages of 47 countries. Highlighted in blue are the hourly wages in India and Mexico



Note: Hourly gaps built with data on average/median wages and average/median hours worked received by the ILO from the respective national statistical institutes or equivalent institutions of the countries included; 94.6 per cent of the data refer to 2013 or a more recent year.

Source: ILO Global Wage Report 2016/17

As for the youth literacy rate, no gender gap is observed for Mexico unlike in India. This makes Mexico a good fit for this, as it is a developing country in which we do not observe big gender gaps similar to India but some gender gaps are still apparent. Moreover, to the best of my knowledge, there was no research done before to test gender bias in one of the main health care determinants, vaccinations, in Mexico. Vaccines are proven to be very effective in Mexico where 53% of deaths caused by influenzas were cut after the introduction of an influenza vaccine to the National Immunization Program (Sánchez-Ramos, Monárrez-Espino and Noyola, 2017), hence making it particularly important to test for any gender preferences in parents' investments in vaccinations in Mexico. Moreover, vaccinations are considered to be the most cost-effective factor used to tackle preventable diseases, particularly in countries with high mortalities caused by such preventable illnesses (Pande & Yazbeck, 2003), With that being said, most of the deaths between 2005 and 2016 for children between the ages of 5-14 years in Mexico were due to preventable conditions (Fadel et al., 2019).

In this study, we will also be exploring gender differentials in breastfeeding, as it is a vital factor in the nutrition of children, particularly in developing countries. As explained by Jones and colleagues, in 2003, breastfeeding can prevent at least 10% of child deaths in low to middle income countries, including Mexico (Jones et al., 2003). Therefore, it is a crucial factor to identify any gender biases in its provision to children as it has serious future implications.

There is also very limited research about the gender differentials in breastfeeding patterns in Mexico. Despite various researches took place to test the determinants of breastfeeding in different cities of Mexico (Perez-Escamilla et al., 1997), most of them are outdated and all took place before the introduction of the national breastfeeding program in Mexico in 1991 (ibid). Even when considering more recent papers which undertook analysis after the introduction of the National Breastfeeding Program (Segovia, Suárez and Aguirre, 2009), the most common determinants analysed were the mothers' ages, occupations and education levels, while none of them tested the breastfeeding patterns against the gender of the infant. With exception to one paper by Perez-Escamilla (1996), which did not find a relation between the child's gender and breastfeeding patterns. Even on a more general note, research on the sole effect of gender on parental investments is uncommon in Mexico. There has been evidence within Mexico where mothers are more likely to enrol their sons in secondary schooling (Chakraborty and De, 2015) or are more likely to spend more on their son's schooling expenditure (Sovero, 2017), however these cases were in the end attributed to the mothers level of autonomy and risk preferences, not exclusively due to gender preference.

Nevertheless, there is more gender parity in Mexico as compared to India, which suggests that there will be considerably fewer gender preferences in parental investments in earlier stages of a child's life. This is tested in this project by considering similar variables to those used in Barcellos paper (2014) where she tests for differentials in parental investments in India.

4. Data Set:

The data set used for this project is a household survey conducted by MxFLS (Mexican Family Life Survey). The survey is Longitudinal and was collected in three points in time: 2002 (MxFLS-1), 2005-2006 (MxFLS-2) and 2009-2012 (MxFLS-3). This project uses the final survey, MxFLS-3, to provide an up-to-date and recent insight on the gender inequality situation in Mexico. The population is surveyed at a household and individual level across 10,200 households; 46,300 individuals from 30 states in Mexico. Given that the MxFLS is a multi-thematic survey, the gathered data available on the website is grouped into different 'books' under household or individual databases. Each book is used for a subgroup of the population, for example, Book V holds individual information about the characteristics of children younger than the age of 15 years and Book IV holds information about the reproductive health of the women in Mexico. Despite the data being grouped into different books, each individual is given an Individual ID number and a Household ID number which allows for separate books to be merged. This feature is used in this project to create a final data set that links children with their respective mothers and to combine the control variable listed in Section 4.2. The final dataset consists of 748 observations with only one observation per household. Moreover, it includes data relating to the youngest child in the family between the ages of 0-15 months as parents would not have had time to have another child as a reaction to the youngest child's gender (Barcellos, 2014). A detailed 'do-file' including all steps taken to create the final data sets is available in Appendix II. The first data set is used for running the regressions (Appendix II – (2a)) and the second is used when testing for son-biased stopping rule in Mexico (Appendix II – (2b)).

4.1 The Dependent Variables:

I consider a number of outcomes to represent the health and nutritional investments that parents decide for their children. The main nutritional investment considered in this research is breastfeeding and the main health investment is vaccination, hence the following four outcomes are considered as the dependent variables:

a. <u>Ever Breastfed</u>: takes 1 if the child was ever breastfed and 0 if not.
b. <u>Breastfeeding Duration</u>: Time in months spent breastfeeding child i.
c. <u>Vaccination Card</u>: takes 1 if the child owns a vaccination card and 0 if the child does not own a vaccination card.
d. <u>Full Vaccination dummy</u>: takes 1 if child is fully vaccinated and 0 if not.

In order to decide whether a child is fully immunized or not, we use the criteria set by the New Mexico Department of Health (Cdc.gov, 2019) to identify the required vaccines for the age group 0-15 months. It is worth noting the following changes and introductions to the immunization schedule:

- Pentavalent vaccine: In Aug 2007, the Pentavalent vaccine was changed to the Pentavalent Acellular which now includes coverage of the Polio vaccine and excludes the Hepatitis B vaccine (Cdc.gov, 2018).
- Rotavirus vaccine: Introduced in 2008 into the immunization program. (Who.int, 2013).
- Pneumococcal Conjugate Vaccine: Introduced in 2009 to the immunization schedule (Apps.who.int, 2013).

The MXFLS survey takes these changes into account by recording the vaccines of children with the previous vaccination card (2005-2008) and those with the current vaccination card (2009 onwards).

'Previous Card' immunization schedule:

Vaccine	Doses
BCG	1
Pentavalent	3
Triple Viral	1
Polio (SABIN)	3

‘Current Card’ immunization schedule:

Vaccine	Doses
BCG	1
Pentavalent Acellular	3
Triple Viral SRP	1
Hepatitis B	3
Rotavirus*	2
Pneumococcal Conjugate	3

*Children in Latin America were vaccinated for Rotavirus using Rotarix, which requires only 2 doses (Who.int, 2009)

4.2 Background/Control Variables:

The control variables used in this project are similar to those used in Barcellos’ publication which are particularly chosen as they do not correlate to the gender of the children who are ‘young enough’ (Barcellos, 2014). For the purpose of this study, we will assume the same age range as Barcellos, which is 0-15 months old (ibid.). This is necessary in order for the data set to meet the assumptions of the methodology, which is further explained in the Methodology section. The control variables used are the following:

a. Total Siblings: shows the number of siblings ever born for child i.
b. Total Brothers: shows the number of brothers ever born for child i.
c. Total Sisters: shows the number of sisters ever born for child i.
d. Mothers’ Age: age in years at the time the survey was taken.
e. Mother’s Education Dummies:
    i. “elem” dummy: takes the value of 1 if the mother’s highest level of education reached is elementary level or below (preschool or no education).
    ii. “sec” dummy: takes the value of 1 if the mother’s highest level of education attained is secondary schooling.

iii. <u>“hs” dummy:</u> takes the value of 1 if the mother’s highest level of education attained is high school or “normal basic” education.

iv. <u>“college” dummy:</u> takes the value of 1 if the mother’s highest level of education attained is college level.

v. <u>“graduate” dummy:</u> takes the value of 1 if the mother has received a diploma/graduated.

f. <u>“Speakspan” dummy:</u> equals to 1 if the mother can speak Spanish.

g. <u>Age first got married:</u> age at which the mother got married the first time.

h. <u>Age first got pregnant:</u> age at which the mother got pregnant the first time.

i. <u>#check-ups during pregnancy:</u> the number of times the mother made check-up visits during pregnancy period.

j. <u>“TetanusShots” dummy:</u> takes the value of 1 if the mother was given tetanus vaccine during pregnancy period.

k. <u>“Homebirth” dummy:</u> takes the value of 1 if the child was born at home.

l. <u>“IndigenousGroup” dummy:</u> takes the value of 1 if the mother belongs to an indigenous group.

5. Methodology:

The following are the alternate hypotheses considered for this project:

Hypothesis 1: Parents with sons are more likely to fully immunize their child.

Hypothesis 2: Mothers are more likely to breastfeed sons than daughters.

Hypothesis 3: Parents with sons are more likely to own a vaccination card for their child.

Hypothesis 4: Mothers with sons are more likely to breastfeed for longer durations.

5.1 Empirical Strategy:

The main task in this dissertation is to assess whether a correlation exists between the gender of the child and the levels of parental investment in health and nutrition in Mexico. In this project, parental investments are represented by 4 variables: Ever breastfed and Breastfeeding durations as the nutritional input and Fully vaccinated and Owning a vaccination card representing the investment in health care provision. In order to test the effect of gender on the above variables, we will be using the following regression model:

$$I = \alpha_0 + \alpha_1 G_{ih} + \rho X_{ih} + u_{ih}$$

Where I indicates the parental investment in the form of breastfeeding duration, full vaccination, having a vaccination card and if the child was ever breastfed. G_{ih} is a dummy variable which equals to 1 if child i from household h is a male, and 0 if female. u_{ih} represent the error term and α_1 is the estimated coefficient which will indicate the relation between the child's gender (G_{ih}) and parents' investments (I). With time, household characteristics begin to be affected by the child's gender, for example, household income and living area/region. Therefore, we limit the age range of the children to the ages before a correlation between the gender and the household characteristics begins to develop, which is 0-15 months (Barcellos, 2014). As explained in the data set section, we control for variables which could affect the parent's investments in the child, but which are not affected by the child's gender. We control for these characteristics through X_{ih} where X can take the form of any of the control variables (listed in section 4.2) related to child i from household h. For example, the age of child i's mother in household h or the total number of siblings that child i from household h has. By limiting the age range, we are allowing the child's gender to be exogenous which follows the OLS assumptions such that the OLS estimate of α_1 is an unbiased estimator; $Cov(G_{ih}, u_{ih}|X) = 0$. In other words, there should be no correlation between the error term u_{ih} and the child's gender G_{ih} once all control variables are controlled for. The age range for which the child's gender is independent is not calculated in this project through joint testing, but is assumed to be 0-15 months similar to the age range calculated in the Barcellos (2014) study.

The main assumptions required for this empirical strategy to work are:

1. There is no sex-selective abortion or excess mortality for either gender (OECD, 2017)
2. $Cov(G_{ih}, u_{ih}|X) = 0$

5.2 Does Mexico follow son-biased stopping rules:

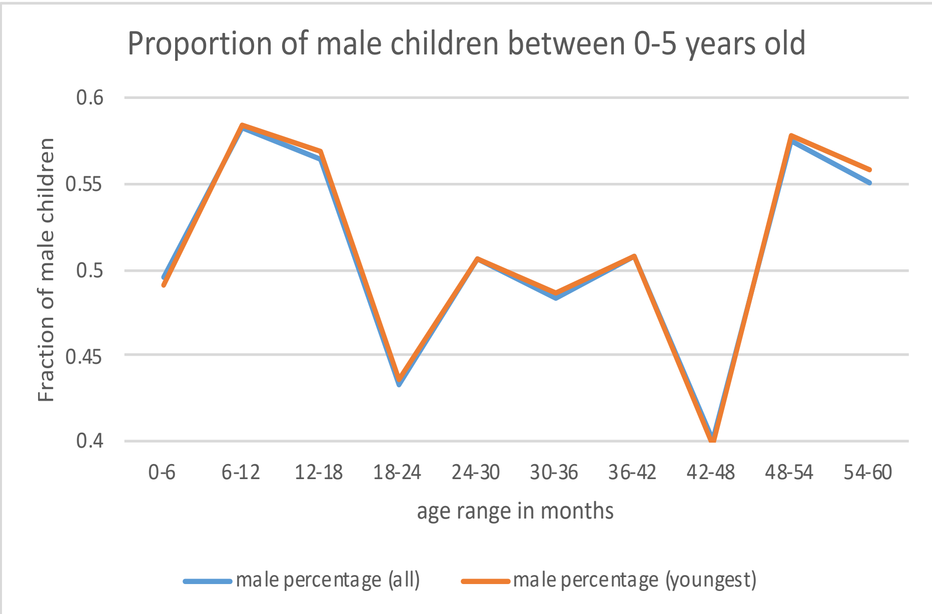
One of the methods used to identify gender preference amongst children is to check whether mothers stop having children after giving birth to their preferred gender. This can be done by finding the probability of the youngest child in the household being a male at different ages. Assuming that the gender of a child is purely determined by nature and hence the sex ratio is exogenously determined, any skewness away from this original ratio amongst youngest children suggests a gender-biased stopping rule. In other words, if families stop having children once they get a boy, the probability that the youngest child in a family is a male increases as the youngest grows older. The hypothesis is that mothers stop having children after they give birth to a son. If this hypothesis is true, then the probability of the youngest child being a male, would be higher than the natural probability of 0.5 at older ages (Barcellos, 2014)

If son-biased stopping rules exist, it creates a correlation between the gender of the child and the size of families. Therefore, we test for son-biased stopping patterns to decide whether family size should be controlled for or not.

6. Results:

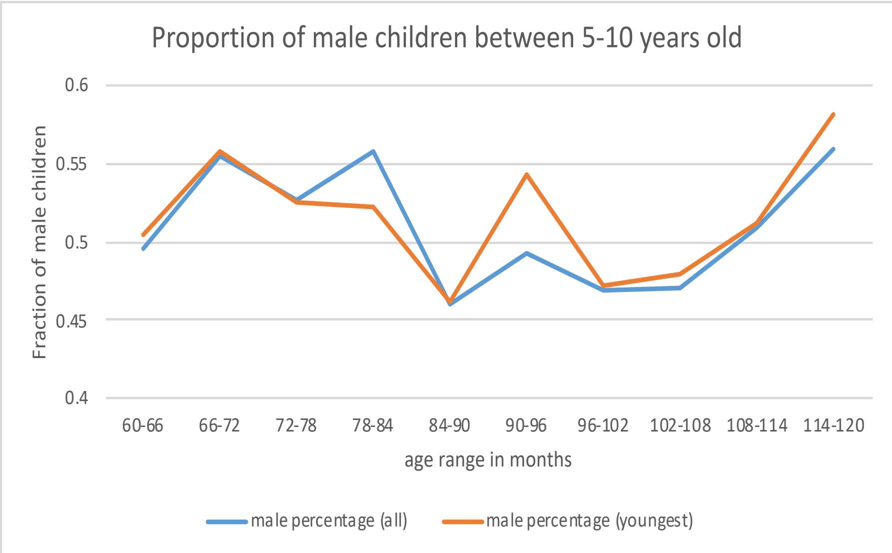
6.1 Gender-biased stopping rule:

As described in the hypothesis in section 5.2, if mothers stop having children after a son then the likelihood that a youngest child is a male would be higher at older ages. To investigate this, the fraction of youngest children who are male was plotted as a function of a 6-month age range between the ages of 0-5 years, illustrating the likelihood of the youngest child being a male at a given age (Graph A, orange line). The fraction of all male children as a function of a 6-month age range between the ages of 0-5 years, was also investigated to evaluate the likelihood of any child being a male (Graph A, blue line).



Graph (A): A graphical illustration of the fraction of any child to be male (blue line) and fraction of youngest child to be male (orange) between 0-5 years within 6-month age ranges.

As shown in Graph A, the likelihood of any child being a male between the ages of 0-5 years (blue line) is above the natural probability of 0.5 between 0-18 months and decreases below the natural probability between 18-48 months. These probabilities are exactly in line with the probabilities of the youngest child being a male (orange line). These results indicate no evident gender preference between these periods.



Graph (B): A graphical illustration of the fraction of any child to be male (blue line) and fraction of youngest child to be male (orange) between 5-10 years within 6-month age ranges.

When looking at older ages (between 5-10 years) in graph B, both the percentage of any child to be male, and the youngest child to be male fluctuate less around the 0.5 natural probability. Another observation is that the fraction of males at a given age between 5-10 years begins to differ between the 'youngest child' and the 'total children'. This could be considered sign of gender preference as it shows that the youngest child's gender begins to shape a trend independently. Nonetheless, it is difficult to confidently conclude the existence of any gender-biased stopping practices as the data is constantly fluctuating and no clear trend can be drawn.

6.2 Regressions:

The control variables discussed in the data set section have been included into the regressions in groups and hence the following five columns are found in all regression tables:

Var(1): A regression between the gender of a child and the dependent variable in question only.

Var(2): Var(1) plus control variables that are personal to the child – birth month and age in months.

Var(3): Var (2) plus control variables related to the characteristics of the mother – mother's age, mother's education, whether the mother speaks Spanish, whether the mother belongs to an indigenous group, the mother's ages at first marriage and first pregnancy.

Var(4): Var(3) plus prenatal information control variables – tetanus shots, number of check-ups during pregnancy and home birth.

Var(5): Var(4) plus sibling related controls variables – number of brothers and number of sisters.

6.2.1 Full Vaccination and Child Gender:

Table 1 shows the relation between a child's gender and its effect on the likelihood of being fully vaccinated. Starting at the simple regression between being fully vaccinated and the child's gender only, [fullvac (1)], there seems to be a negative relation between gender and being fully vaccinated, in other words, males are less likely to be fully vaccinated. However, this result is not statistically significant (t=-0.0542) and thus suggests that there is no relation between a child's gender and their likelihood of completing all required vaccinations. Nevertheless, when we include the personal control variables to the regression [fullvac(2)], the estimated coefficient increases and becomes positive to 0.015. This shows that a male child is more likely to be fully vaccinated. With a higher estimated coefficient (even in absolute terms) and a relatively lower standard error, the t-statistic increases (t=0.4) and skews more towards statistical significance, but does not reach the critical values therefore remaining statistically insignificant. However, once the mothers characteristics [fullvac(3)], prenatal information [fullvac(4)] and sibling characteristics [fullvac(5)] are included, the estimated coefficients are negative again but remain statistically insignificant (t=-0.96, t=-1.06 and t=-1.02 respectively). Therefore, we do not have enough statistical evidence to reject our null hypothesis which states that there is no gender preference amongst parents on investing for their child to receive all the vaccines listed in the immunization schedule.

Moreover, the percentage of the variation in the likelihood of being fully vaccinated explained by this study's regression model, given by the R-squared indicator, is very low. Child gender explains 0% of the variation in the likelihood of being fully vaccinated [fullvac(1)] and when all sets of control variables are included [fullvac(5)], our regression model is only able to explain 11.6% of the variation. This confirms that a child's gender is not a major factor which explains differentials in the likelihood of a child being fully vaccinated. Other variables included in the regression such as the child's age in months, the mother's education and whether or not the mother speaks Spanish are all statistically significant, meaning they have an effect on the likelihood of the child being fully immunised, but they do not inform us about the relation between child gender and being fully vaccinated.

Table 1: Tabulated OLS regression outcomes showing the effect of child gender on being fully vaccinated by parents

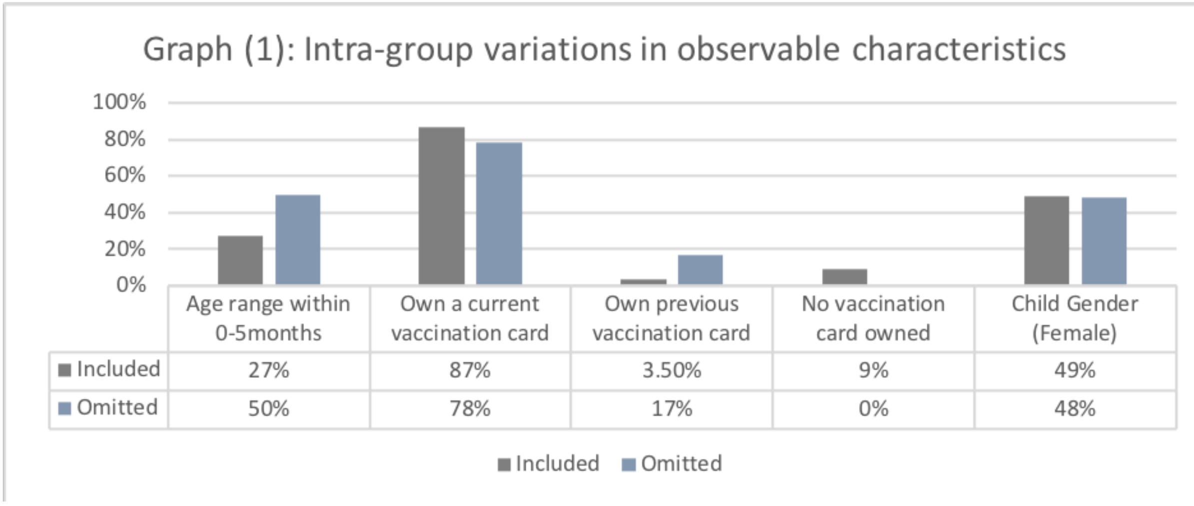
	(1)	(2)	(3)	(4)	(5)
VARIABLES	fullvac	fullvac	fullvac	fullvac	fullvac
childgender	-0.00214	0.0150	-0.0594	-0.0669	-0.0642
	(0.0395)	(0.0378)	(0.0616)	(0.0632)	(0.0629)
birthmonth		-0.0119**	-0.00689	-0.00460	-0.00409
		(0.00545)	(0.00827)	(0.00851)	(0.00852)
ageinmonths		0.0336***	0.0348***	0.0318***	0.0289***
		(0.00465)	(0.00760)	(0.00787)	(0.00817)
motherage			0.00734	0.0108	0.0275**
			(0.00927)	(0.00872)	(0.0129)
elem			-0.385***	-0.344***	-0.327***
			(0.0857)	(0.0886)	(0.0898)
sec			-0.418***	-0.391***	-0.384***
			(0.0713)	(0.0763)	(0.0769)
hs			-0.418***	-0.399***	-0.396***
			(0.0861)	(0.0885)	(0.0882)
college			-0.356***	-0.345***	-0.317***
			(0.0992)	(0.110)	(0.109)
o.graduate			-	-	-
speakspan			-0.380***	-0.375***	-0.372***
			(0.0741)	(0.0781)	(0.0782)
indigenousgroup			-0.0764	-0.105	-0.0658
			(0.0885)	(0.0903)	(0.0910)
agemotherfirstmarried			-0.00319	-0.00288	-0.00461
			(0.0130)	(0.0134)	(0.0139)
agemotherfirstpreg			-0.00389	-0.00607	-0.0212
			(0.0132)	(0.0130)	(0.0161)
tetanusshots				0.0694	0.0670
				(0.0996)	(0.101)
homebirth				0.132	0.125
				(0.260)	(0.242)
numbercheckupsduringpreg				-0.00215	-0.00245
				(0.00753)	(0.00749)
totalbrothers					-0.128**
					(0.0504)
totalsisters					-0.0198
					(0.0707)
Constant	0.661***	0.460***	1.206***	1.111***	1.110***
	(0.0280)	(0.0587)	(0.137)	(0.179)	(0.176)
Observations	579	579	245	240	240
R-squared	0.000	0.086	0.098	0.096	0.116

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

6.2.1.1 Demographic Analysis:

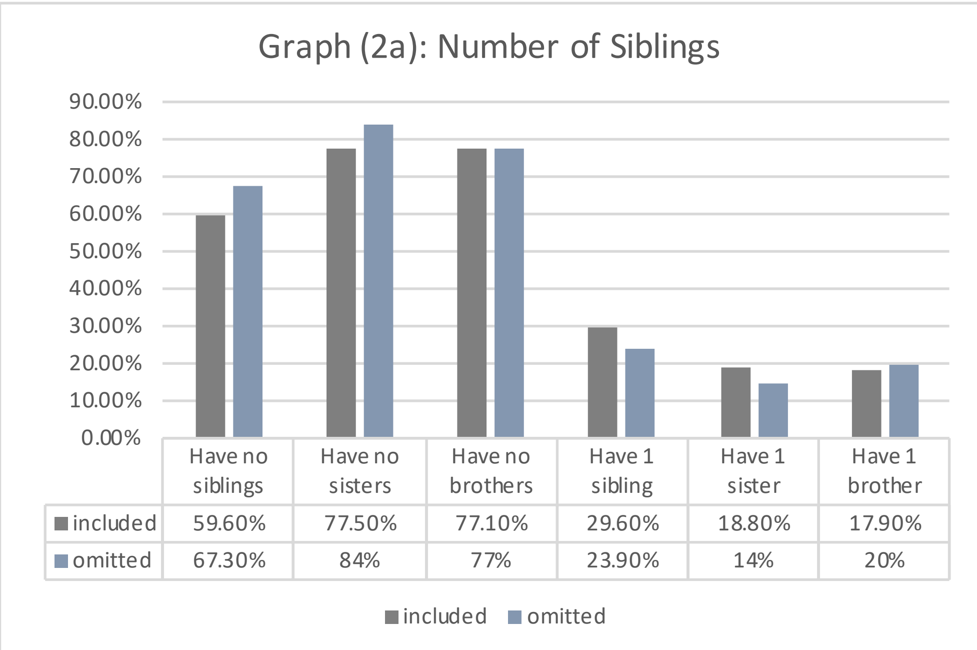
It is also noticed that the number of observations included in the simple regression [fullvac(1)] are less than the total observations included in the dataset (579 included from 748). This is due to some observations having missing information about their vaccination schedule. We compare the observable characteristics of those 169 omitted observations against the characteristics of the included observations to check if there is a trend which could explain why this group of children do not have recorded information regarding their vaccinations.



Graph (1): The intra-group variations found in 5 observable characteristics represented by percentages of the included 579 observations (grey) and the omitted 169 observations (blue).

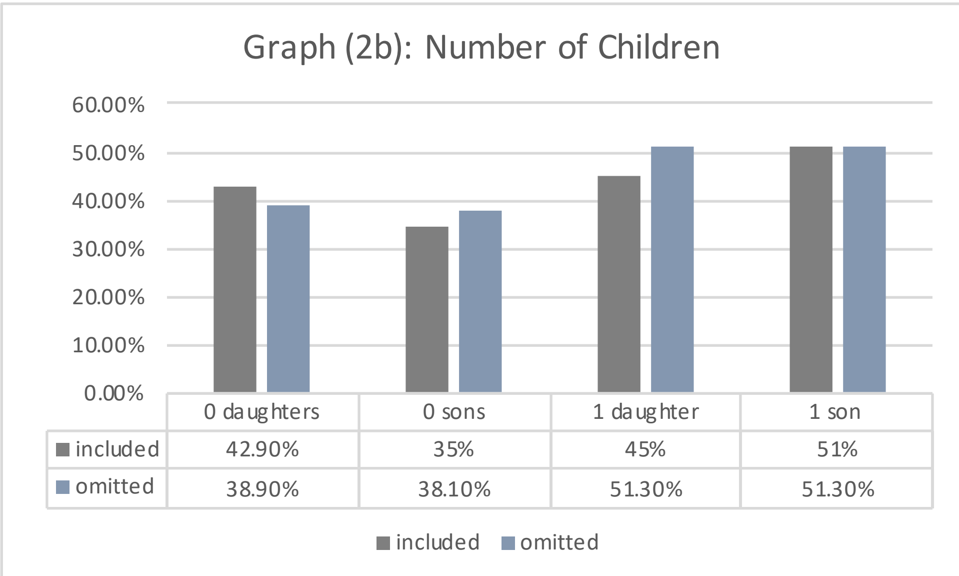
Graph (1) shows that half of the children with incomplete or missing information recorded about their vaccinations are between 0 to 5 months old compared to only 27% of the observations included in the regression. Moreover, 78% of the omitted observations own a current vaccination card compared to 87% of the included observations. As for those who own a previous vaccination card, they make up 17% of the omitted group and 3.5% of the included group. This says that more people have current vaccination cards in the included group compared to the omitted group and more people have previous vaccination cards in the omitted group than the included group. This is an unexpected outcome since we would expect to find a trickledown effect from having more younger children in the omitted group to having more current vaccination cards in the omitted group too. Moreover, everyone in the omitted group owns a vaccination card where as 9% of the included group do not have vaccination card. Finally, the proportion of boys and girls within the omitted group and included group are equal, indicating that the gender of the child does not affect the decision of the parents to record their infants' vaccination schedule or not. This further supports that a child's gender is not a significant contributor to them being fully vaccinated.

The above analysis refers to the omitted observations under the regression without any control variables. Once all controls are included [fullvac(5)], more observations are omitted. The regression includes only 240 observations from 748. Table 1 shows most observations are lost once the mother's characteristics are included as control variables (as seen in [fullvac(3)]), with the number of observations included in the regression dropping from 579 to 245. A similar demographic analysis of observable characteristics is conducted to show in what ways the omitted variables differ from those included in the regression and to investigate possible reasons that may have led to many missing observations.



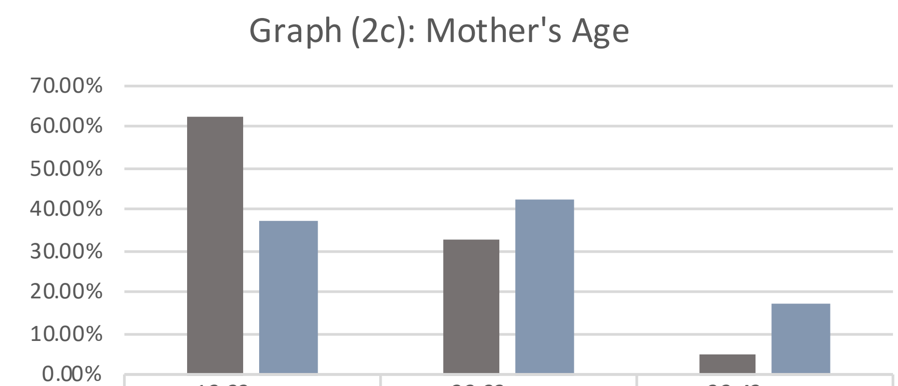
Graph (2a): The variations found in the number and gender of siblings represented by percentages of the included 240 observations (grey) and the omitted 508 observations (blue).

Graph (2a) shows that there is a higher percentage of children who do not have siblings in the omitted group compared to the included group. This could be due to parents having no experience in filling surveys related to children as they only have one child. Deeper investigation shows that the omitted group has higher percentage of children with no sisters (84%) compared to the included group (77.5%), as opposed to having no brothers, which is the same across both groups (both approximately 77%). When increasing the number of siblings, it is noticed that there are more children with 1 sibling in the included group (29.6%) as opposed to the omitted group (23.9%). By considering the gender of the siblings, we notice that there are more children with 1 sister in the included group than the omitted. As for having 1 brother, the opposite is true, but at a smaller percentage difference. This suggests that the effect of sisters is higher than the effect of brothers on parents' decisions to provide information on surveys. We further investigate this from the parents' point of view through the number of children they have:



Graph (2b): The variations found in the number and gender of children the parents have, represented by percentages of the included 240 observations (grey) and the omitted 508 observations (blue).

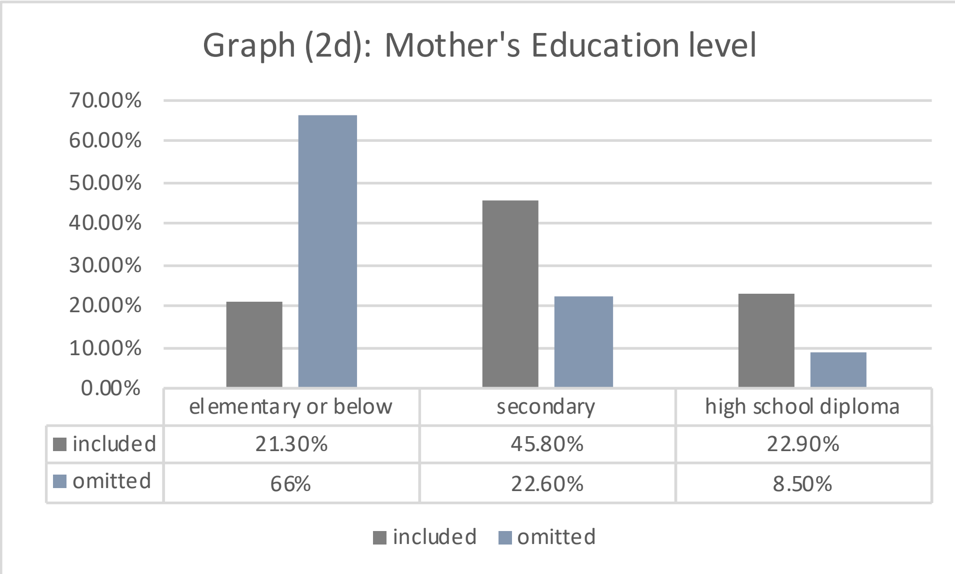
Graph (2b) shows the difference between the omitted group and the included group with respect to the number of children the parents have. We can see that there is a higher proportion of the included group who have no daughters (42.9%) compared to the omitted group (38.9%), while the opposite is true for having no sons. Once again, the effect is lower for having no sons than having no daughters. This suggests that parents who have no daughters were more likely to provide full information about their child's immunization schedule than having no sons as they are more likely to be included in the regression. When increasing the number of children the parents have, it is noticed that there is a higher probability for an omitted observation to belong to a household with only 1 daughter (51.3%) compared to an included observation (45%), while the likelihood of having 1 son seems to be unchanged for either group. This shows that it is possible for the number of daughters to have higher effects on the likelihood of a parent filling in surveys as compared to the number of sons, which has small to no effect.



	13-23 years	23-33 years	33-42 years
included	62.50%	32.50%	5%
omitted	37.10%	42.40%	17.30%

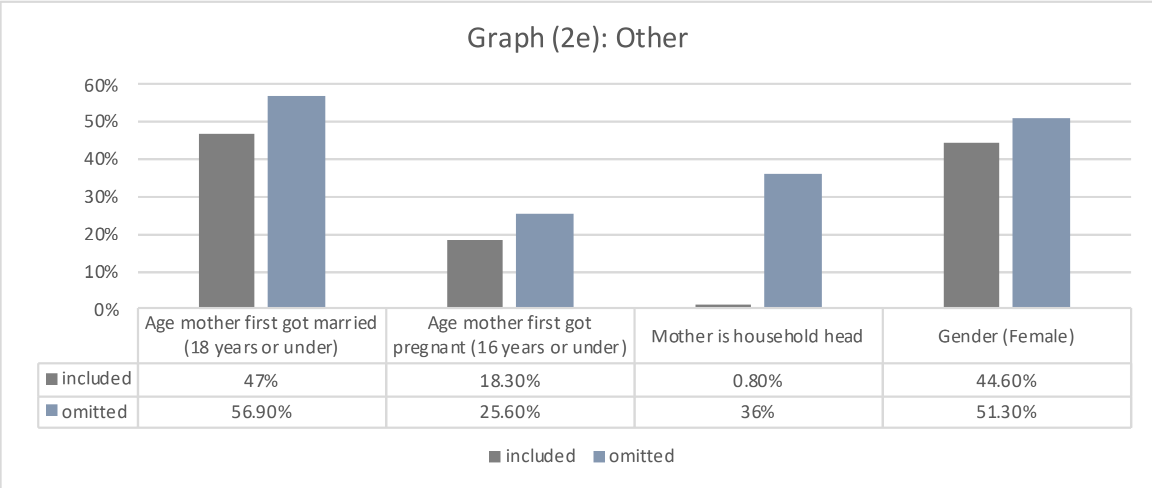
Graph (2c): The variations found in the mothers' age ranges represented as percentages of the included 240 observations (grey) and the omitted 508 observations (blue).

Since most of the observations are lost once mother characteristics and included, we check for evident trends in the characteristics of mothers. An obvious trend is spotted when it comes to the age of the mothers, where younger mothers are more likely to provide complete survey responses and therefore are more likely to be part of the included group in the regression (Graph (2c)). 62.5% of the included observations had mothers between the age of 13-23 years, compared to only 37.1% of the omitted group. As the mother's age increases, it is clearer that mothers become less likely to provide full or accurate survey records. For example, 32.5% of the included group's mothers belong to the age range of 23-33 compared to a larger 42.4% of the omitted group. At the age range of 33-43 years old, these percentages change to only 5% of included observations and 17.3% of omitted groups.



Graph (2d): The variations found in the mothers' education levels represented as percentages of the included 240 observations (grey) and the omitted 508 observations (blue).

The level of the mother's education also plays a role in determining whether or not mothers complete the surveys. 66% of the mothers in the omitted group have attained a level of education only up to elementary or even no education compared to only 21% of the included group. When looking at mothers with secondary level of education attained, the figures change to 22.6% and 45.8% respectively. It is noticed that higher education levels attained by mothers results in more complete survey responses.



Graph (2e): The variations found in other observable characteristics represented as percentages of the included 240 observations (grey) and the omitted 508 observations (blue).

Graph (2e) shows relatively more females in the omitted group (51.3%) compared to the females within the included group (44.6%). This could mean that parents with the youngest child being a daughter are more likely to not provide full or accurate information on surveys. However, the difference is not substantial. Moreover, mothers who got married for the first time at early ages (18 years or less) and those who got first pregnant at early ages (16 years or less) are more likely to have incomplete surveys. Another observation, and perhaps the most striking, is that 36% of mothers of omitted observations are considered to be the heads of their household, compared to only 0.8% of those from the included observations. This could largely be attributed to the amount of other responsibilities these mothers have to take care of as opposed to filling in full surveys.

6.2.2 Ever breastfed and Child gender

The regression output shown in table 2 shows a negative, yet not statistically significant, estimated coefficient of the relation between a child being ever breastfed and child gender. In other words, there are no signs of gender bias in the mother's decision of breastfeeding or not breastfeeding her child. With the increase of control variables included in the regression, the t-statistic increases in absolute value towards statistical significance, however, never reaching the level of statistical significance under the model specification used in this study (from t=-0.15 to t=-0.96 for everbreastfed(1) and everbreastfed(5) respectively). Similarly, the R-squared variable increases across the columns with each set of controls added but reaching a maximum of 6%. In other words, when all sets of control variables are included, the regression model is only able to explain 6% of the difference between a child who has been breastfed and another who has not. Child gender explains 0% of the variation in the mother's decision of breastfeeding her child or not. Therefore showing that gender is also not a major explanatory factor for differentials in breastfeeding decisions.

Table 2: Tabulated OLS regression outcomes showing the effect of child gender on being ever breastfed by mother

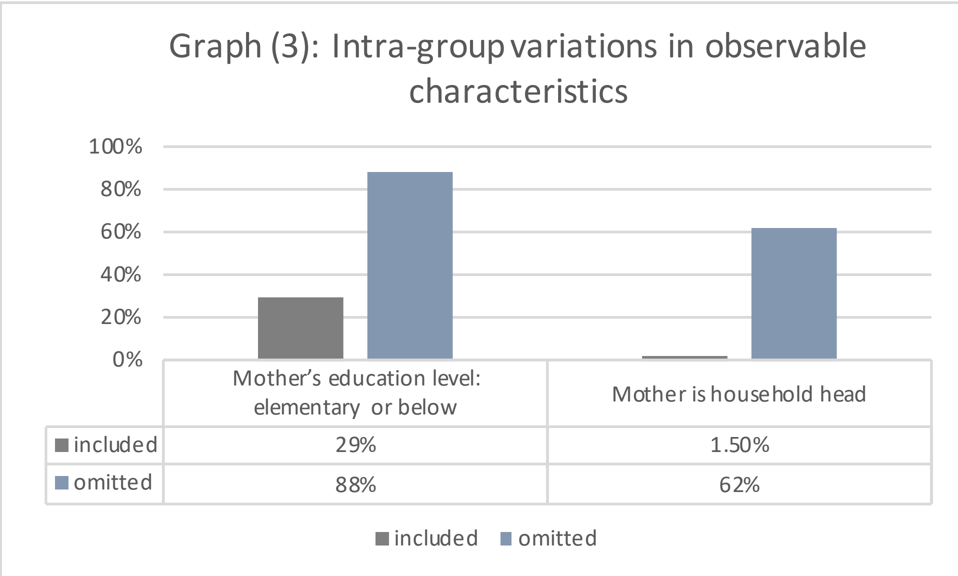
	(1)	(2)	(3)	(4)	(5)
VARIABLES	everbreastfed	everbreastfed	everbreastfed	everbreastfed	everbreastfed
childgender	-0.00346	-0.00150	-0.0125	-0.0237	-0.0242
	(0.0234)	(0.0235)	(0.0269)	(0.0253)	(0.0253)
birthmonth		0.00108	-0.000210	0.00181	0.00151
		(0.00334)	(0.00398)	(0.00378)	(0.00382)
ageinmonths		0.00273	0.00231	0.000240	0.000396
		(0.00274)	(0.00305)	(0.00280)	(0.00283)
motherage			0.000259	0.00558**	0.00181
			(0.00537)	(0.00244)	(0.00244)
elem			-0.0323	3.49e-05	-0.00364
			(0.0323)	(0.0282)	(0.0285)
sec			-0.0607*	-0.0418	-0.0433
			(0.0311)	(0.0338)	(0.0336)
hs			-0.0418	-0.0279	-0.0297
			(0.0401)	(0.0446)	(0.0445)
college			-0.0723	-0.0534	-0.0581
			(0.0540)	(0.0489)	(0.0495)
o.graduate			-	-	-
speakspan			-0.0296	-0.0612**	-0.0616**
			(0.0333)	(0.0299)	(0.0300)
indigenousgroup			0.0325	0.0148	0.00915
			(0.0284)	(0.0301)	(0.0315)
agemotherfirstmarried			0.00879	0.0109	0.0111
			(0.00705)	(0.00676)	(0.00678)
agemotherfirstpreg			-0.00779	-0.0149**	-0.0112
			(0.00887)	(0.00733)	(0.00707)
tetanusshots				0.103*	0.104*
				(0.0591)	(0.0593)
homebirth				0.0934**	0.0960**
				(0.0459)	(0.0454)
numbercheckupsduringpreg				-0.00248	-0.00223
				(0.00400)	(0.00398)
totalbrothers					0.0200
					(0.0145)
totalsisters					0.0213
					(0.0134)
Constant	0.934***	0.905***	0.990***	0.921***	0.922***
	(0.0164)	(0.0371)	(0.0550)	(0.0635)	(0.0634)
Observations	461	461	309	302	302
R-squared	0.000	0.003	0.015	0.059	0.062

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

6.2.2.1 Demographic Analysis:

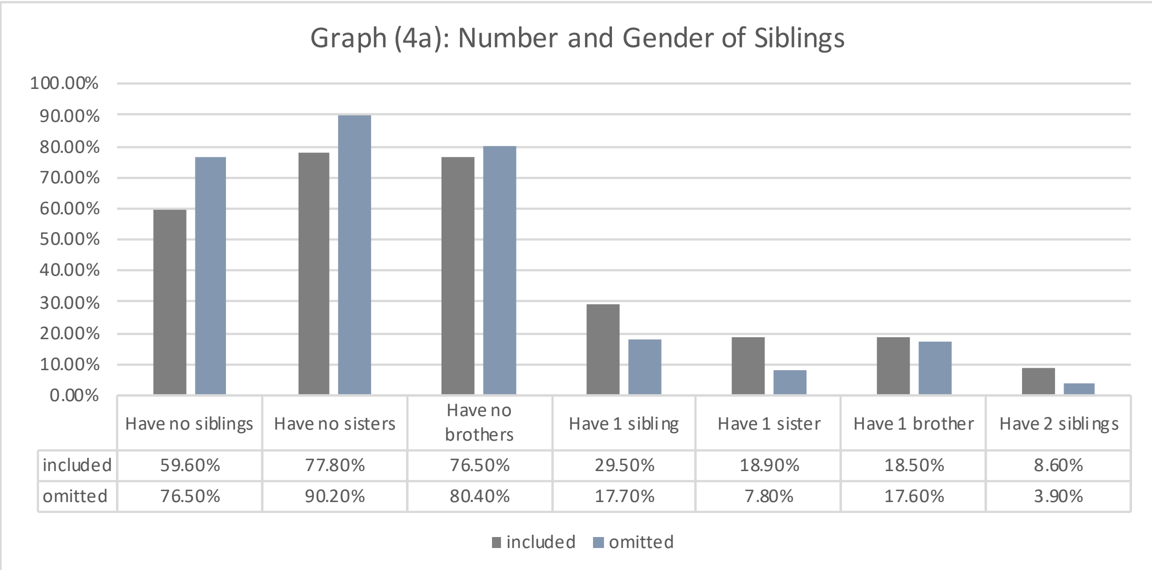
Similar to the case of full vaccinations, the regression [everbreastfed(1)] omits a number of observations due to missing information about whether or not the children have ever been breastfed. By running a test on the characteristics of these omitted observations, the following differences were found:



Graph (3): The intra-group variations found in 2 observable characteristics represented by percentages of the included 461 observations (grey) and the omitted 287 observations (blue).

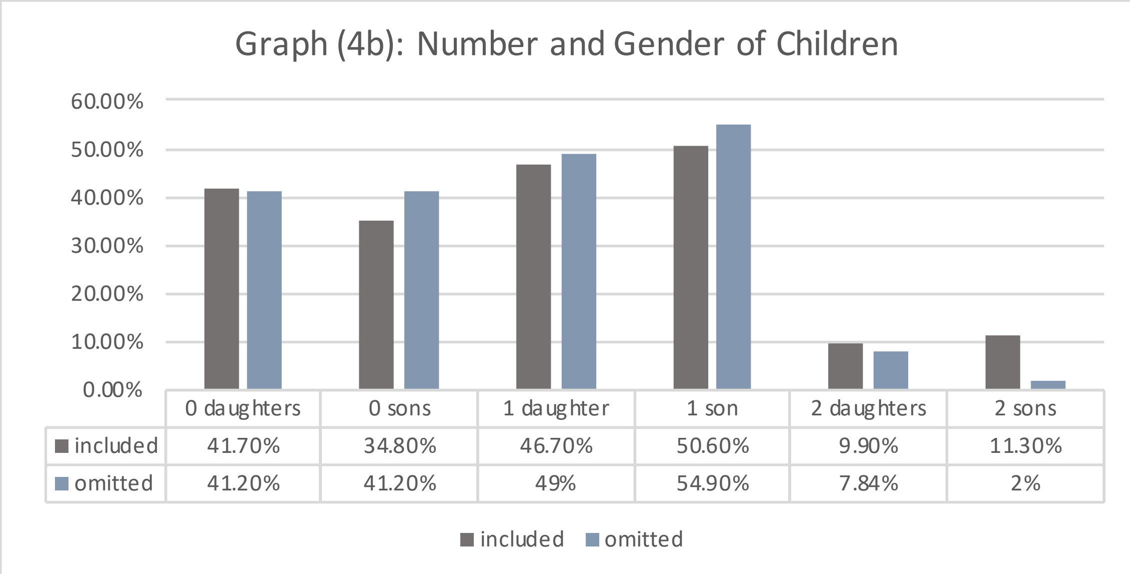
Graph (3) shows that 88% of the mothers who did not record whether they ever breastfed their child or not have only received the elementary level of education or even below compared to only 29% of those with recorded information. Moreover, 62% of the children with missing information about being ever breastfed happen to have mothers who take on the role of a household head compared to only 1.5% of those with recorded information. Perhaps this is due to these mothers having relatively more responsibilities and hence less time to fill in the survey as accurately and completely as others who take on the role of a spouse/partner of the household head.

As with the case of Full Vaccinations, adding more control variables leads to lower observations included in the regression. The largest drop in observations occurs with the introduction of the mothers characteristics as control variables with 152 observations (from 461 to 309 observations). Similarly, comparison graphs of characteristics for the omitted observations against the included observations takes place, given all control variables are added to the regression:



Graph (4a): The variations found in the number and gender of siblings represented by percentages of the included 302 observations (grey) and the omitted 446 observations (blue).

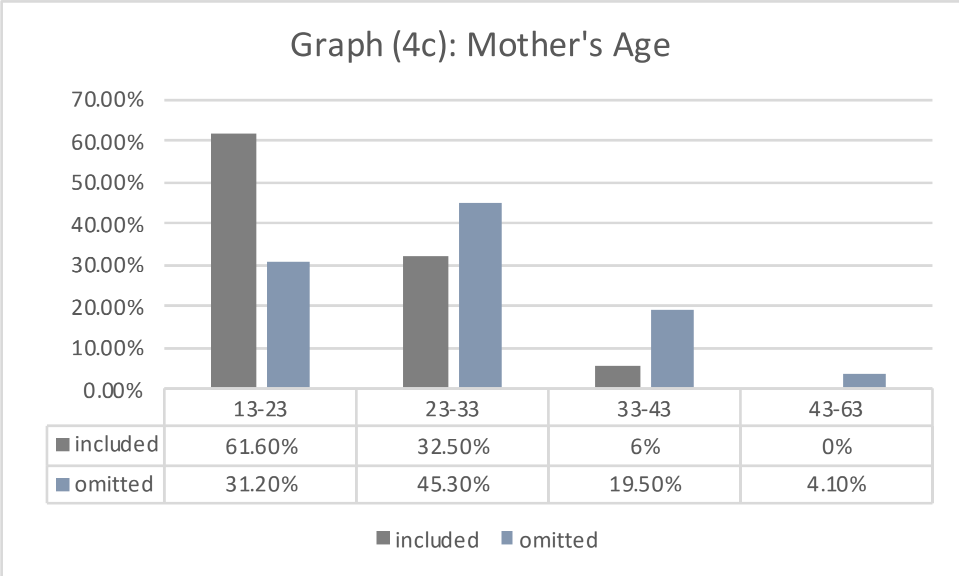
Graph (4a) shows the comparison of the number of siblings a child has between the included and omitted observations. The percentage of children who have no siblings is higher among the omitted observations. By checking the changes to the percentages when a child has 1 sibling, we see that the percentage for the omitted observations falls below the percentage for included observations (17.7% for omitted compared to 29.5% for included). Another outcome shows that having no sisters seems to have higher effect than having no brothers. Children who do not have sisters are more likely to be omitted from the regression than children who have no brothers. This hints towards some gender bias when it comes to the parents providing information about their children. Graph (4a) also shows that a child who has 1 brother has equal chances of being within an omitted or included into the regression. While there is a higher chance for a child from the included group to belong to a household with 1 sister as compared to a child within the omitted group. To further analyse this, we check for gender bias in recording information from the parent’s perspective:



Graph (4b): The variations found in the number and gender of children the parents have, represented by percentages of the included 302 observations (grey) and the omitted 446 observations (blue).

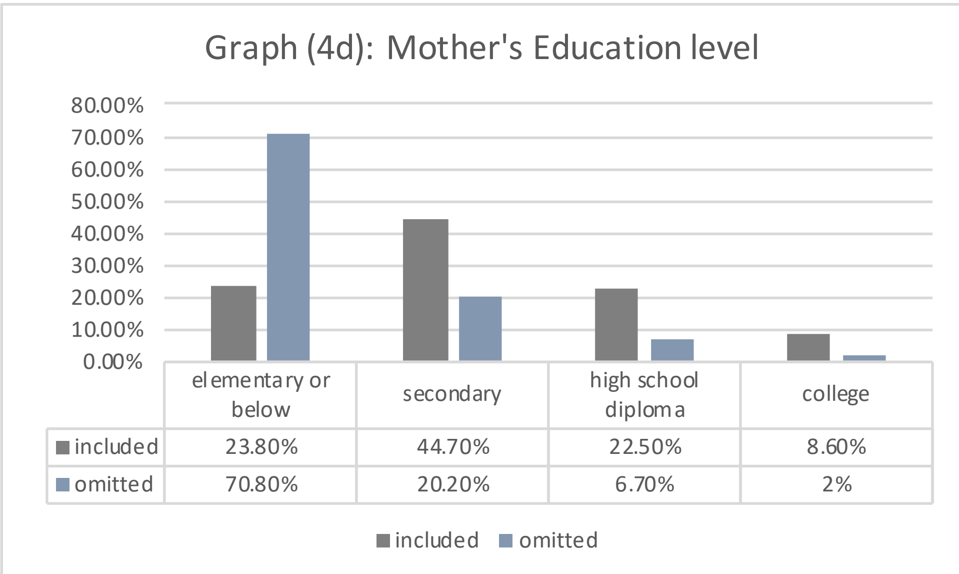
Graph (4b) shows the probability for a child to be omitted or included into the regression based on the number of daughters or sons the parents have. As already established, observations are omitted due to missing or incomplete information, therefore, we analyse the percentages from Graph (4b) as the likelihood of a parent providing incomplete information. The data shows that having no daughters has no effect on the parents decision to complete the survey whereas having 0 sons lowers the probability of completing a survey. Moreover, having 1 son has a higher impact on making parents provide incomplete surveys as compared to having 1 daughter. As we increase the number of sons to 2, parents are more likely provide full and complete information for the child to be included in the regression. As for having 2 daughters, the same applies but at a smaller scale. Although a clear trend cannot be drawn, it is safe to say that having a son affects parental decision more than daughters when it comes to filling in complete surveys.

Other trends were found relating to the mother characteristics such as age, level of education and age at first pregnancy:



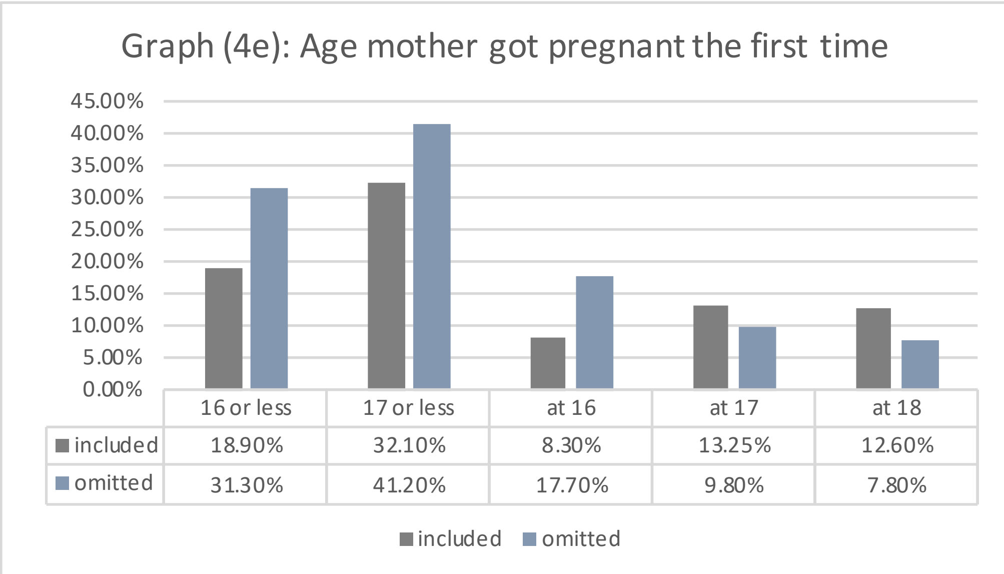
Graph (4c): The variations found in the mothers' age ranges represented as percentages of the included 302 observations (grey) and the omitted 446 observations (blue).

Graph (4c) shows that younger mothers are more likely to provide complete and accurate information on the surveys.



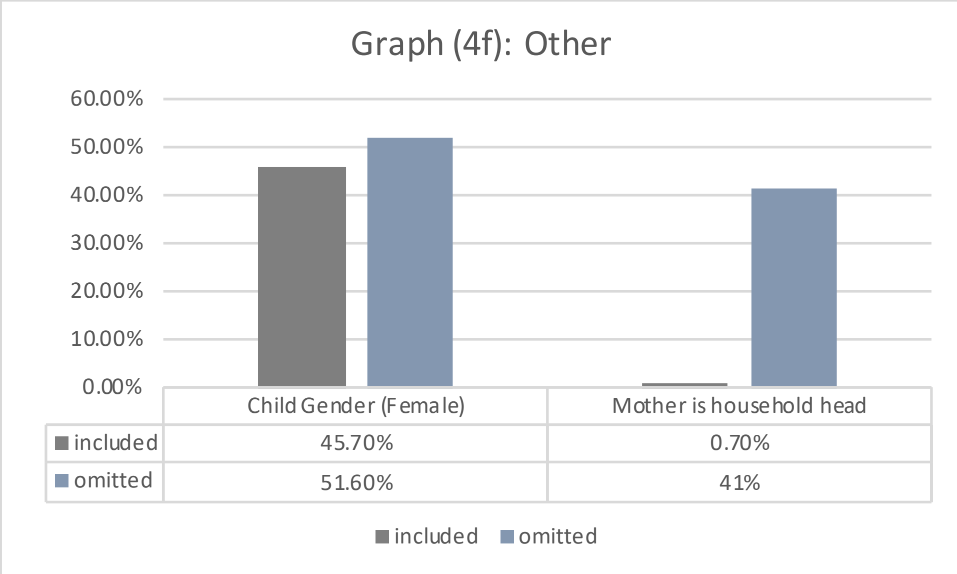
Graph (4d): The variations found in the mothers' education levels represented as percentages of the included 302 observations (grey) and the omitted 446 observations (blue).

As for Graph (4d), it is seen that mothers with higher levels of education are more likely to complete surveys. Beginning at elementary level of education, approximately 71% of the mothers from the omitted observations have elementary level or below compared to 23.8% of the included observations. However when this is increased to secondary level of education, we see a higher percentage of mothers with this level in the included observations (44.7%) as opposed to those amongst the omitted (20.2%).



Graph (4e): The variations found in the mothers' age at her first pregnancy, represented as percentages of the included 302 observations (grey) and the omitted 446 observations (blue).

The outcome in Graph (4e) shows that mothers who got pregnant the first time at younger ages are more likely to not provide sufficient information through surveys.



Graph (4f): The variations found in 2 other observable characteristics (child gender and mother as a household head) represented as percentages of the included 302 observations (grey) and the omitted 446 observations (blue).

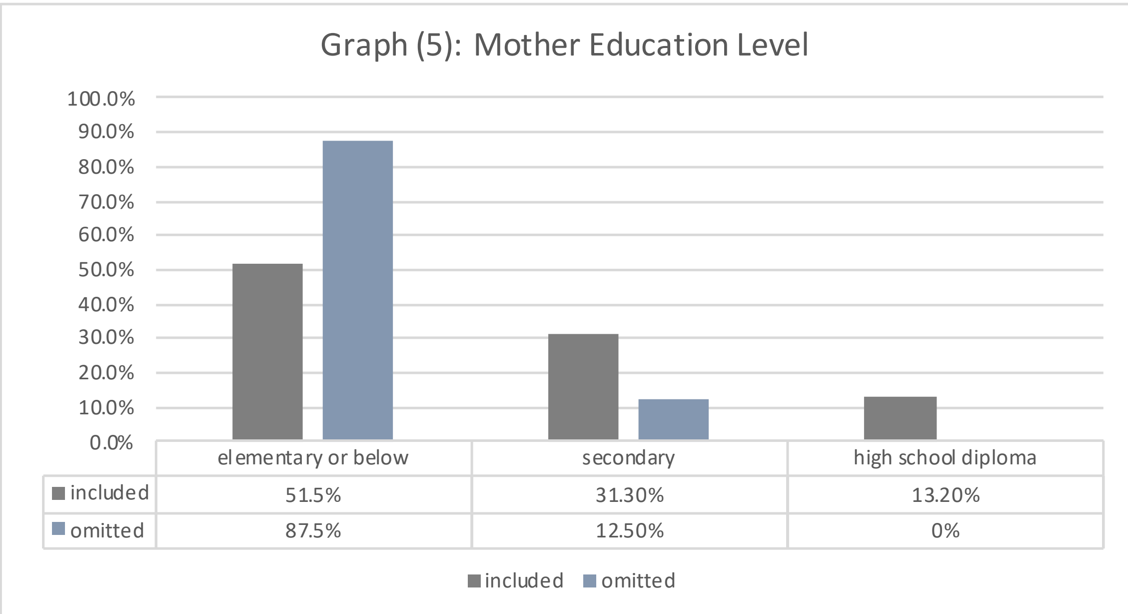
Other outcomes noticed are included in Graph (4f). The first shows that there are more females in the omitted group (51.6%) as compared to the included group (45.7%). However, this is a very small difference suggesting very small effect of gender on the inclusion or exclusion of the observation from the regression. The second outcome is the position of the mother within the household. 41% of the omitted observations are from households where the mother is considered as the head of the family whereas only 0.7% of the included group come from such households. A possible explanation for this could be that mothers who are the head of their households have more responsibilities to take care of and hence have less time to fill in complete surveys.

6.2.3 Vaccination Card and Child gender

Similar to being fully vaccinated and ever breastfed, there is no statistically significant relation between a child's gender and the ownership of a vaccination card, as shown in table 3. The t-values decrease from 0.97 for VacCard(1) to 0.66 for VacCard(5), thus remaining insignificant. The regression model, with all control variables, is only able to explain 8% of the difference between the children who own a vaccination card and those who do not. As for the gender of the child, it only explains to 0.1% of a child's likelihood of owning a vaccination card.

Unlike the previous regressions, the majority of the observations were not missing, with only 8 observations omitted from the regression. When analysing the 8 missing observations, we find the following trend:

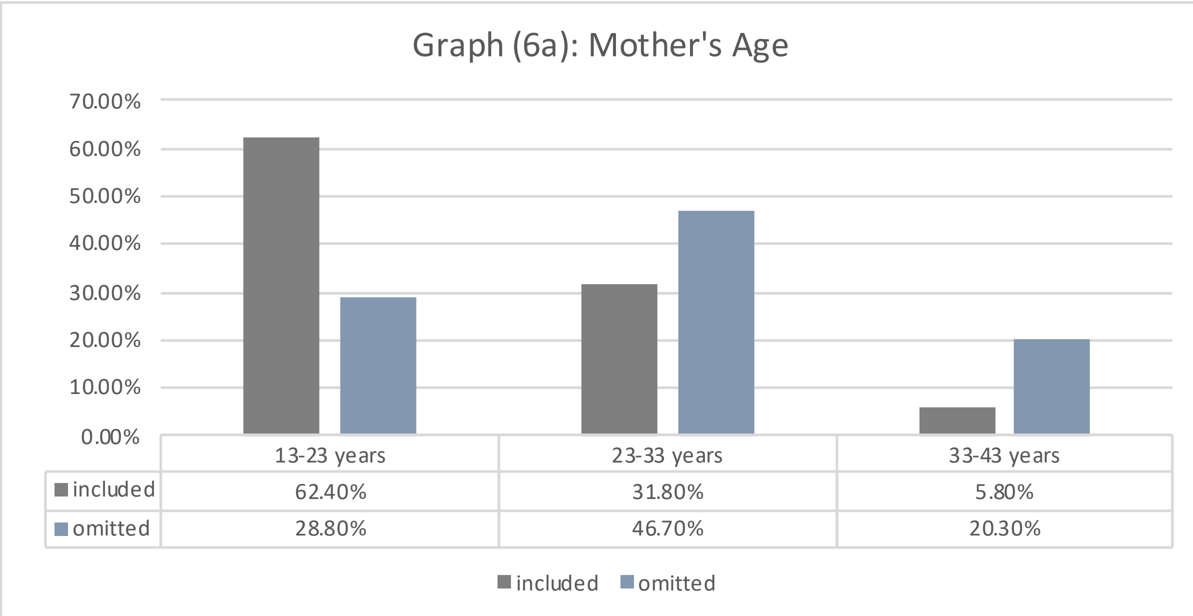
6.2.3.1 Demographic Analysis:



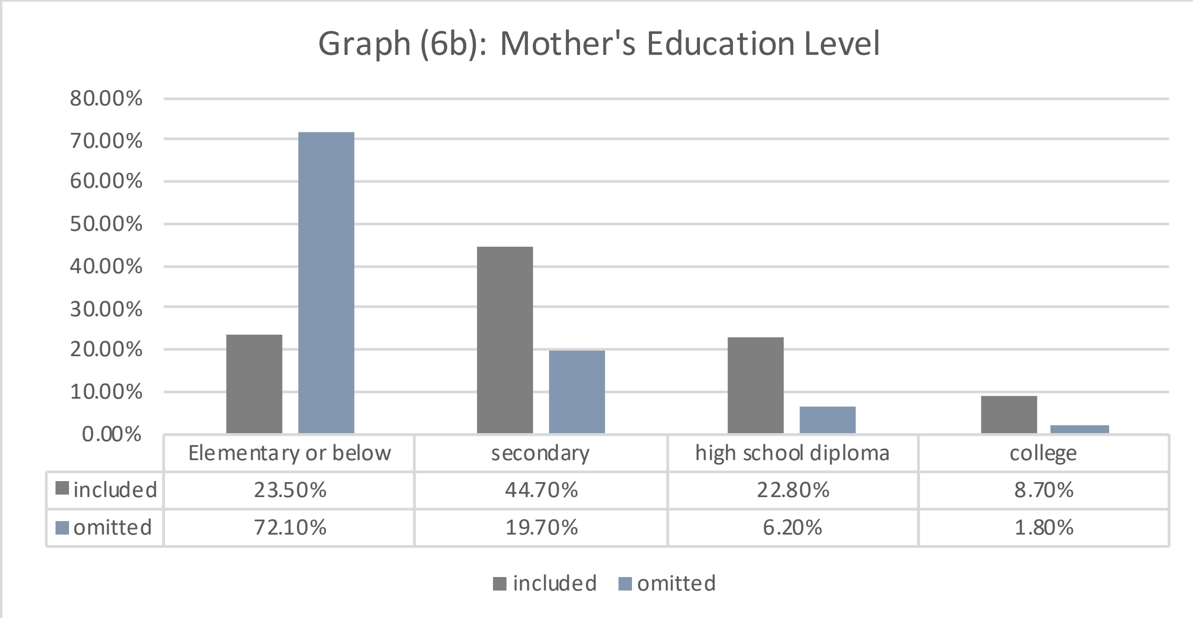
Graph (5): The intra-group variations found in one observable characteristic, mother's education level, represented by percentages of the included 740 observations (grey) and the omitted 8 observations (blue).

Graph (5) shows that, from the 8 omitted observations, 7 mothers have elementary level education or below, with 6 of the 8 mothers not receiving any level of education. Moreover, only one mother from the omitted group has attained secondary level of education. This could explain why they did not record information within the survey as they have relatively low literacy which could lead to low comprehension of survey questions and hence not answering them.

The above analysis related to the simple regression between having a vaccination card and child gender. However, when looking at the multiple regression with all control variables included [VacCard(5)], we witness a large fall in the number of included observations going from 740 observations (almost all observations in the dataset) to only 311 observations. By analysing the characteristics of the observations omitted in the final regression against the included observations, we find the following:

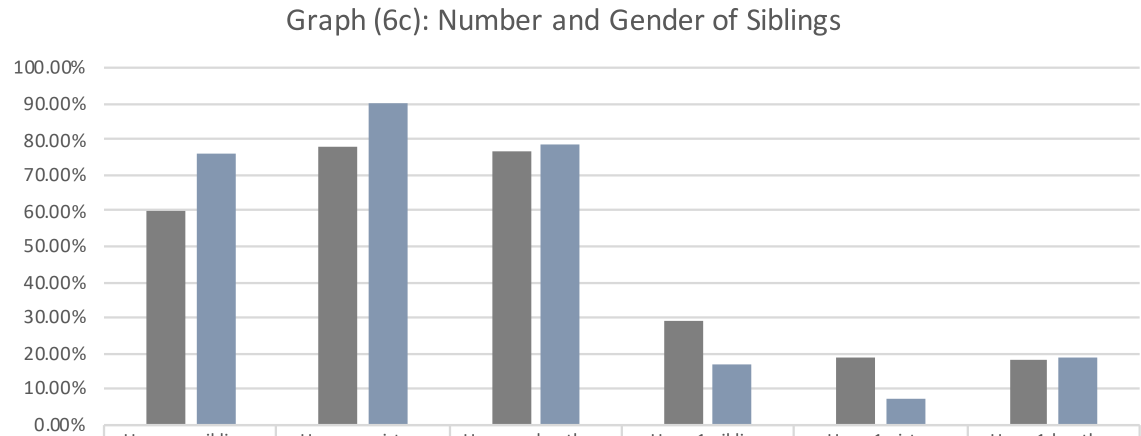


Graph (6a): The variations found in the mothers' age ranges represented as percentages of the included 311 observations (grey) and the omitted 437 observations (blue).



Graph (6b): The variations found in the mothers' education levels represented as percentages of the included 311 observations (grey) and the omitted 437 observations (blue).

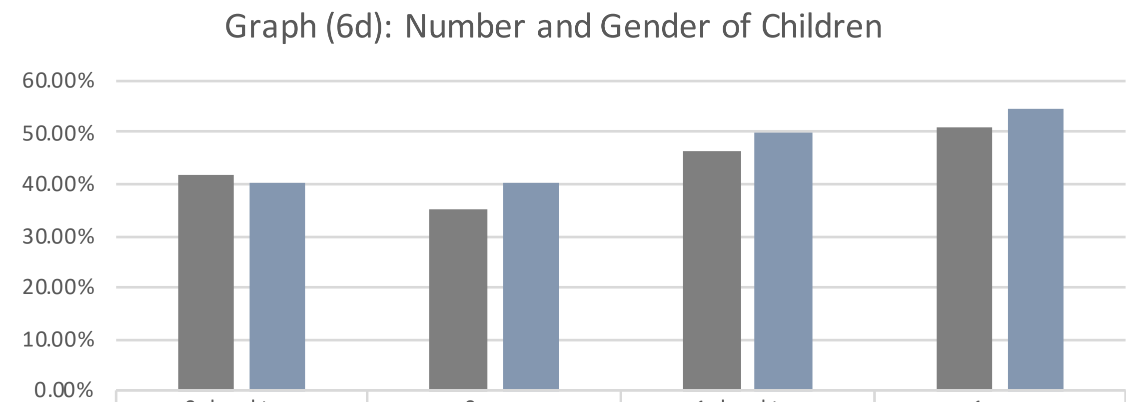
The above two graphs (6a and 6b) show similar outcomes to those established in the previous two regressions; Younger mothers are more likely to provide complete survey information and lower level of mother's education make up higher proportions within the omitted groups.



	Have no siblings	Have no sisters	Have no brothers	Have 1 sibling	Have 1 sister	Have 1 brother
included	60.10%	78.10%	76.90%	29.30%	18.70%	18.30%
omitted	76.20%	90.50%	78.60%	16.70%	7.10%	19.10%

included omitted

Graph (6c): The variations found in the number and gender of siblings represented by percentages of the included 311 observations (grey) and the omitted 437 observations (blue).

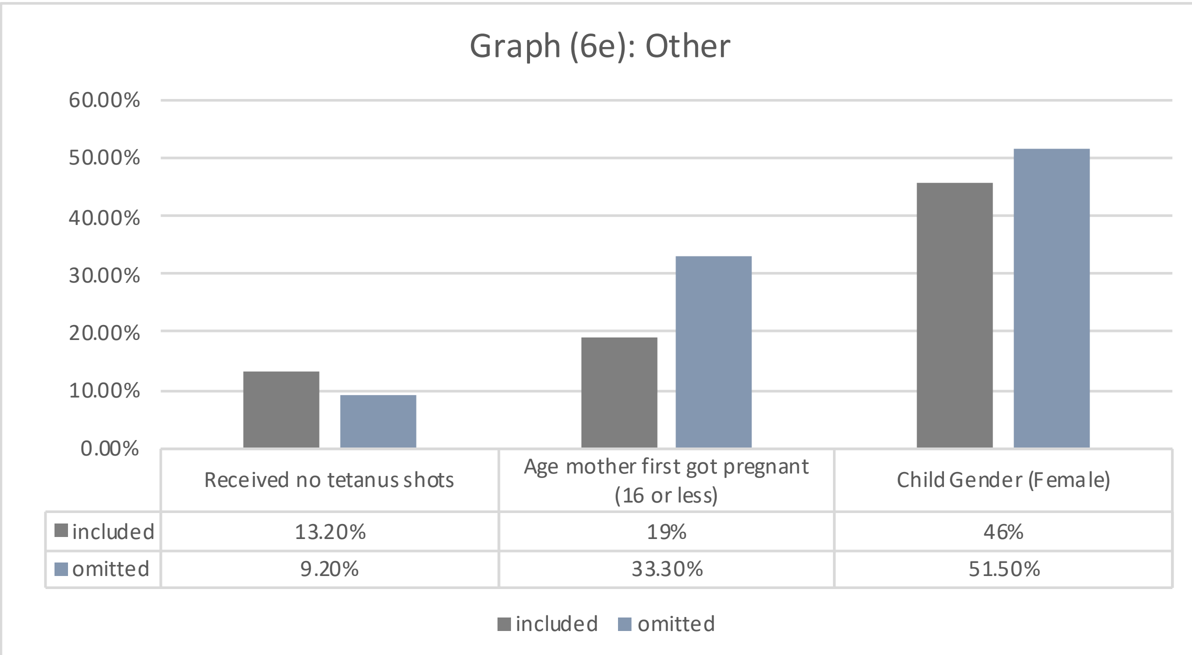


	0 daughters	0 sons	1 daughter	1 son
included	41.80%	35.1%	46.6%	50.8%
omitted	40.50%	40.50%	50.00%	54.80%

included omitted

Graph (6d): The variations found in the number and gender of children the parents have, represented by percentages of the included 311 observations (grey) and the omitted 437 observations (blue).

Applying the same reasoning as in previous sections, Graphs (6c) and (6d) show that daughters can have more impact on parental decisions than sons. This is similar to the result arrived under the full vaccination demographic analysis (Graph (2b)) that daughters have higher effects on parents' decisions to complete surveys; while the opposite was found under the Everbreastfed regression from Graph (4b).



Graph (6e): The variations found in 3 other observable characteristics represented as percentages of the included 311 observations (grey) and the omitted 437 observations (blue).

Graph (6e) shows a higher proportion of female children amongst omitted group compared to the included group. Nevertheless, it is a very small difference. The graph also shows that a lower proportion of mothers from the omitted group who have not received tetanus shots as part of prenatal care, 9.2%, as compared to 13.2% of the group included in the regression. Finally, similar to the demographic analysis of the previous two regressions, the omitted group has a higher proportion of mothers who first got pregnant at a young age (33.3%) compared to the included group (19%).

Table 3: Tabulated OLS regression outcomes showing the effect of child gender on owning a vaccination card

VARIABLES	(1) VacCard	(2) VacCard	(3) VacCard	(4) VacCard	(5) VacCard
childgender	0.0186	0.0236	0.0213	0.0195	0.0210
	(0.0192)	(0.0191)	(0.0315)	(0.0318)	(0.0319)
birthmonth		-0.00549**	-0.00532	-0.00471	-0.00381
		(0.00274)	(0.00447)	(0.00453)	(0.00446)
ageinmonths		0.00982***	0.0129***	0.0120**	0.0118**
		(0.00287)	(0.00475)	(0.00477)	(0.00472)
motherage			-0.00308	0.00202	0.0119
			(0.00555)	(0.00408)	(0.00800)
elem			-0.0856*	-0.0911*	-0.0823
			(0.0509)	(0.0537)	(0.0540)
sec			-0.109**	-0.109**	-0.106**
			(0.0421)	(0.0442)	(0.0447)
hs			-0.0434	-0.0494	-0.0460
			(0.0415)	(0.0441)	(0.0446)
college			-0.0898	-0.0934	-0.0837
			(0.0557)	(0.0590)	(0.0529)
speakspan			-0.107***	-0.0981***	-0.0971***
			(0.0380)	(0.0332)	(0.0335)
indigenousgroup			-0.0132	-0.0213	-0.00929
			(0.0473)	(0.0490)	(0.0460)
agemotherfirstmarried			0.00754	0.00671	0.00628
			(0.00767)	(0.00796)	(0.00816)
agemotherfirstpreg			-0.00204	-0.00656	-0.0163*
			(0.00812)	(0.00740)	(0.00886)
tetanusshots				0.0853	0.0827
				(0.0564)	(0.0564)
homebirth				0.109**	0.104*
				(0.0490)	(0.0534)
numbercheckupsduringpreg				0.00151	0.000940
				(0.00315)	(0.00324)
totalbrothers					-0.0413
					(0.0465)
totalsisters					-0.0685*
					(0.0356)
Constant	0.918***	0.878***	1.008***	0.910***	0.908***
	(0.0144)	(0.0329)	(0.0948)	(0.111)	(0.112)
Observations	740	740	320	311	311
R-squared	0.001	0.030	0.059	0.067	0.080

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

6.2.4 Breastfeeding duration and Child gender

The regression for the breastfeeding duration includes only 142 observations and eventually drops down to 98 once all the control variables are included. This is due to the fact that many of the mothers are still breastfeeding and hence cannot provide a fixed number of months as yet. Therefore, the outcome is not very representative due to a small number of observations and even so, the relation is not statistically significant. The t-values are -1.21 for monthsbreastfed(1) and it decreases in absolute value to -0.68 for monthsbreastfed(5). Moreover, due to a small number of observations and a large amount of dummy variables, we face an unexpected and unplanned collinearity between homebirth and the control variables of total brothers, total sisters and age mother first got pregnant, thus omitting homebirth variable from the regression. The results are shown in table 4.

In order to address this issue more accurately, we run another regression using the Tobit regression model instead of the OLS. This is mainly because our dependent variable (Months breastfed) is only censored and limited to those who have stopped breastfeeding and not those who are still breastfeeding while our independent variable (child gender) is not. The results from this regression are shown in table 5. Nevertheless, the Tobit regression provides the same outcome as the OLS regression model, hence confirming no statistically significant relation between the duration of breastfeeding and the child's gender.

Table 4: Tabulated OLS regression outcomes showing the effect of child gender on breastfeeding duration

	(1)	(2)	(3)	(4)	(5)
VARIABLES	monthsbreastfed	monthsbreastfed	monthsbreastfed	monthsbreastfed	monthsbreastfed
childgender	-0.656	-0.494	-0.436	-0.506	-0.496
	(0.541)	(0.559)	(0.685)	(0.723)	(0.729)
birthmonth		-0.0445	0.0415	0.0228	0.00699
		(0.0721)	(0.0981)	(0.101)	(0.106)
ageinmonths		0.127	-0.00678	0.0302	0.0412
		(0.103)	(0.151)	(0.168)	(0.178)
motherage			0.163	0.163	0.0328
			(0.157)	(0.162)	(0.222)
elem			0.498	0.446	0.358
			(1.195)	(1.355)	(1.366)
sec			0.846	1.063	0.993
			(0.765)	(0.882)	(0.914)
hs			1.347	1.555	1.379
			(0.982)	(1.060)	(1.065)
college			2.754**	2.964**	2.729**
			(1.069)	(1.196)	(1.309)
speakspan			-0.609	-0.525	-0.536
			(0.828)	(0.931)	(0.946)
indigenousgroup			1.237	1.178	1.046
			(1.396)	(1.420)	(1.385)
agemotherfirstmarried			-0.225*	-0.226	-0.226
			(0.132)	(0.140)	(0.140)
agemotherfirstpreg			-0.0562	-0.0636	0.0655
			(0.176)	(0.183)	(0.238)
tetanusshots				0.636	0.728
				(0.767)	(0.794)
o.homebirth				-	-
numbercheckupsduringpreg				-0.0564	-0.0495
				(0.0460)	(0.0484)
totalbrothers					0.490
					(0.733)
totalsisters					0.877
					(0.929)
Constant	4.926***	3.918***	5.925***	5.537***	5.578***
	(0.399)	(1.048)	(1.487)	(1.604)	(1.580)
Observations	142	142	101	98	98
R-squared	0.010	0.027	0.103	0.116	0.127

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Tabulated TOBIT regression outcomes showing the effect of child gender on breastfeeding duration

	(1)	(2)	(3)	(4)	(5)
VARIABLES	monthsbreastfed	monthsbreastfed	monthsbreastfed	monthsbreastfed	monthsbreastfed
childgender	-0.656	-0.494	-0.436	-0.506	-0.496
	(0.536)	(0.543)	(0.671)	(0.692)	(0.688)
birthmonth		-0.0445	0.0415	0.0228	0.00699
		(0.0735)	(0.0906)	(0.0928)	(0.0937)
ageinmonths		0.127	-0.00678	0.0302	0.0412
		(0.0826)	(0.108)	(0.117)	(0.118)
motherage			0.163	0.163	0.0328
			(0.128)	(0.129)	(0.180)
elem			0.498	0.446	0.358
			(3.345)	(3.385)	(3.371)
sec			0.846	1.063	0.993
			(3.288)	(3.324)	(3.305)
hs			1.347	1.555	1.379
			(3.325)	(3.359)	(3.346)
college			2.754	2.964	2.729
			(3.388)	(3.435)	(3.421)
speakspan			-0.609	-0.525	-0.536
			(3.289)	(3.316)	(3.296)
agemotherfirstmarried			-0.225*	-0.226*	-0.226*
			(0.127)	(0.131)	(0.131)
agemotherfirstpreg			-0.0562	-0.0636	0.0655
			(0.161)	(0.162)	(0.202)
indigenousgroup			1.237	1.178	1.046
			(1.043)	(1.053)	(1.055)
numbercheckupsduringpreg				-0.0564	-0.0495
				(0.0827)	(0.0824)
tetanusshots				0.636	0.728
				(1.154)	(1.150)
o.homebirth				-	-
totalbrothers					0.490
					(0.818)
totalsisters					0.877
					(0.795)
Constant	4.926***	3.918***	5.925	5.537	5.578
	(0.387)	(0.979)	(4.702)	(4.931)	(4.902)
Observations	142	142	101	98	98

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Results under Logit regression model:

Since the dependent and independent variables used in this study are all binary variables (with the exception of breastfeeding duration), Logit regression models were also run (Appendix III, tables 6 – 8). This regression model was used to test the relation of child gender with being fully vaccinated, having a vaccination card and being ever breastfed. The results generated were similar to the OLS regression results and therefore supporting the findings of no statistically significant relations between any of the dependant variables with the gender of the child.

7. Discussion and Conclusion:

The results obtained are in line with what was expected; no gender differentials in the provision of nutrition and healthcare variables by parents in a country with less gender inequality in its labour force. In others words, in a developing country where children of both genders have relatively more equal work opportunities in their future, parents do not adjust their nutritional inputs and their child's health care access based on their gender. Looking at it in economic terms, when the probability of later returns generated by a child in their future is more equalized, the parents' utility of investing in a son or daughter converges and thus not incentivizing the parents to develop gender preferences among their children. Research addressing gender differentials in developing countries with relatively more gender parity have not been widely adopted, therefore the results obtained cannot confirm but can act as evidence and indication of how the gender differentials in the labour force affects parental investments in developing countries with less gender discrimination. More research into other countries would help further clarify and confirm if parents do not discriminate in their investments based on the gender biases in the child's future economic worth. Moreover, further research into the effect of the gender of the child on parent's provision of full information on surveys would help determine whether the results found in the demographic analysis are due to causation or not. Just because the omitted group's differed in the characteristic of number of daughters/sons, does not necessarily mean that they are solely omitted because of number of daughters or sons. As widely known, correlation does not necessarily mean causation and hence further research into this relation would be favourable. Despite reaching to this conclusion, a few factors may have affected the final results and are addressed below:

<u>Small Sample Size:</u> A large part of our results were affected by missing or unrecorded information, hence reducing the sample size used in the regression. Smaller number of observations make it less likely to find a relation. Moreover, for those observations which have completed the full survey, there were still certain inaccuracies spotted in the information recorded. For example, different dates of birth were recorded in different files for the same individual or contradicting genders were recorded for the same individual in different parts of the survey. In a few cases, the personal ID number given to each individual did not match the surveys description of equating the household ID plus the individuals line number within that respective household. Observations with such inaccuracies were dropped from the sample,

resulting in a further smaller sample size. Another limitation for using particularly the MxFLS survey is that it is originally in Spanish but was translated to English, where some inaccuracies may have risen during the translation.

<u>Age range considered:</u> Another possible factor affecting the results would be the age range of 0-15 months studied. In this project, 0-15 months was assumed to be the age range for which the child's gender was exogenous and therefore would not be correlated with the error term, as calculated by Barcellos (2014). However, it would be more accurate to calculate this age range specifically for the sample used in this paper. Due to this assumption, it is possible for the results to be biased by probable dependence between other family characteristics, which were uncontrolled for, and the child's gender.

Despite the limitations incurred, this study can still be used to incentivize further research into gender discrimination in parental investments in Mexico. Future interesting scopes could include testing other variables, which represent parental investments to get a more wholesome representation of the state of gender differentials in Mexico. The original aim of the study initially included testing gender biases in the child care time allocated by parents, however it was not plausible due to unavailable variables in the MxFLS dataset used in this project. Further scope of research could test for gender differential treatment arising from the health system in Mexico as opposed to that from parents. I believe the main hurdle with such researches would be finding the data. However, if the data is available, such studies will be very helpful and informative for the field of gender discrimination in health and nutrition for children in developing countries.

8. Bibliography:

Anderson, S. and Ray, D. (2010). Missing Women: Age and Disease. *Review of Economic Studies*, 77(4), pp.1262-1300.

Antai, D. (2012). Gender inequities, relationship power, and childhood immunization uptake in Nigeria: a population-based cross-sectional study. *International Journal of Infectious Diseases*, 16(2), pp.e136-e145.

Apps.who.int. (2013). *Introduction of pneumococcal vaccine PCV13, A handbook for district and health facility staff*. [online] Available at: https://apps.who.int/iris/bitstream/handle/10665/90380/WHO_IVB_13.10_eng.pdf?sequence=1 [Accessed 14 Feb. 2019].

Barcellos, S., Carvalho, L. 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), pp.157-189.

Basu, A. (1989). Is Discrimination in Food Really Necessary for Explaining Sex Differentials in Childhood Mortality?. *Population Studies*, 43(2), pp.193-210.

Beaulieu, D. and Bugental, D. (2008). Contingent parental investment: an evolutionary framework for understanding early interaction between mothers and children. *Evolution and Human Behavior*, 29(4), pp.249-255.

Borooah, V. (2004). Gender bias among children in India in their diet and immunisation against disease. *Social Science & Medicine*, 58(9), pp.1719-1731.

Cdc.gov. (2018). *2018 Binational Immunization Resource Tool for Children from Birth Through 18 Years*. [online] Available at: https://www.cdc.gov/vaccines/schedules/downloads/child/binational-schedule-pr.pdf [Accessed 12 Feb. 2019].

Cdc.gov. (2019). *Birth-18 Years Immunization Schedule | CDC*. [online] Available at: https://www.cdc.gov/vaccines/schedules/hcp/imz/child-adolescent.html#f11 [Accessed 10 Feb. 2019].

Chakraborty, T. and De, P. (2015). Mother’s autonomy and child’s secondary schooling enrollment in Mexico. *Review of Economics of the Household*, 15(3), pp.1037-1053.

Chen, L., Huq, E. and D'Souza, S. (1981). Sex Bias in the Family Allocation of Food and Health Care in Rural Bangladesh. *Population and Development Review*, 7(1), p.55.

Choi, E. and Hwang, J. (2015). Child Gender and Parental Inputs: No More Son Preference in Korea?. *American Economic Review*, 105(5), pp.638-643.

Data.worldbank.org. (2018). *Labor force participation rate, female (% of female population ages 15+) (modeled ILO estimate) | Data*. [online] Available at: https://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS?locations=IN-CN-BD.

Data.worldbank.org. (2018). *Labor force participation rate, female (% of female population ages 15+) (modeled ILO estimate) | Data*. [online] Available at:

https://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS?locations=MX-IN [Accessed 19 Feb. 2019].

Data.worldbank.org. (2018). *Labor force participation rate, male (% of male population ages 15+) (modeled ILO estimate) | Data*. [online] Available at: https://data.worldbank.org/indicator/SL.TLF.CACT.MA.ZS?locations=IN-CN-BD.

Data.worldbank.org. (2018). *Labor force participation rate, male (% of male population ages 15+) (modeled ILO estimate) | Data*. [online] Available at: https://data.worldbank.org/indicator/SL.TLF.CACT.MA.ZS?locations=MX-IN [Accessed 19 Feb. 2019].

Deaton, A. (1989). Looking for Boy-Girl Discrimination in Household Expenditure Data. *The World Bank Economic Review*, 3(1), pp.1-15.

Duflo, E. (2005). Gender Equality in Development. *BREAD policy paper*, 11(4).

Ennvih-mxfls.org. (n.d.). *MXFLS Mexican Family Life Survey Home*. [online] Available at: http://www.ennvih-mxfls.org/english/index.html [Accessed 28 Aug. 2018].

Fadel, S., Boschi-Pinto, C., Yu, S., Reynales-Shigematsu, L., Menon, G., Newcombe, L., Strong, K., Wang, Q. and Jha, P. (2019). Trends in cause-specific mortality among children aged 5–14 years from 2005 to 2016 in India, China, Brazil, and Mexico: an analysis of nationally representative mortality studies. *The Lancet*, 393(10176), pp.1119-1127.

Frenk, J. (2006). Bridging the divide: global lessons from evidence-based health policy in Mexico. *The Lancet*, 368(9539), pp.954-961.

Fuse, K. (2013). Daughter preference in Japan: A reflection on gender role attitudes?. *Demographic Research*, 28, pp.1021-1052.

GOV.UK. (2017). *India and UK working together to improve gender equality in the workplace*. [online] Available at: https://www.gov.uk/government/news/india-and-uk-working-together-to-improve-gender-equality-in-the-workplace [Accessed 28 Jan. 2019].

Graham, M., Larsen, U. and Xu, X. (1998). Son Preference in Anhui Province, China. *International Family Planning Perspectives*, 24(2), p.72.

Gupta, M. (2005). Explaining Asia's "Missing Women": A New Look at the Data. *Population and Development Review*, 31(3), pp.529-535.

International Labour Organization (2016). *Global Wage Report 2016/17 Wage inequality in the workplace*. International Labour Organization.

International Labour Organization (2018). *India Wage Report Wage policies for decent work and inclusive growth*. [online] Available at: https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/---sro-new_delhi/documents/publication/wcms_638305.pdf.

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), pp.1485-1538.

Jones, G., Steketee, R., Black, R., Bhutta, Z. and Morris, S. (2003). How many child deaths can we prevent this year?. *The Lancet*, 362(9377), pp.65-71.

Kishor, S. (1993). "May God Give Sons to All": Gender and Child Mortality in India. *American Sociological Review*, 58(2), pp.247-265.

Li, J. (2004). Gender inequality, family planning, and maternal and child care in a rural Chinese county. *Social Science & Medicine*, 59(4), pp.695-708.

MDG Fund. (n.d.). *Delivering on commitments*. [online] Available at: http://mdgfund.org/content/genderequalityandwomensempowerment [Accessed 28 Jan. 2019].

Milazzo, A. (2018). Why are adult women missing? Son preference and maternal survival in India. *Journal of Development Economics*, 134, pp.467-484.

OECD (2017). *Latin America and the Caribbean SIGI Regional Report*. [online] OECD. Available at: https://www.oecd.org/dev/development-gender/Brochure_SIGI_LAC_web.pdf [Accessed 26 Mar. 2019].

Pande, R. and Yazbeck, A. (2003). What's in a country average? Wealth, gender, and regional inequalities in immunization in India. *Social Science & Medicine*, 57(11), pp.2075-2088.

Pasqua, S. (2005). Gender Bias in Parental Investments in Children's Education: A Theoretical Analysis. *Review of Economics of the Household*, 3(3), pp.291-314.

Perez-Escamilla, R., Lutter, C., Wickham, C., Phillips, M., Treviño-Siller, S. and Sanghvi, T. (1997). Identification of risk factors for short breastfeeding duration in Mexico city through survival analysis. *Ecology of Food and Nutrition*, 36(1), pp.43-64.

Poffenberger, T. (1982). The endangered sex: Neglect of female children in rural north India. *Child Abuse & Neglect*, 6(3), pp.369-370.

Rose, E. (1999). Consumption Smoothing and Excess Female Mortality in Rural India. *Review of Economics and Statistics*, 81(1), pp.41-49.

Rosenzweig, M. and Schultz, T. (1982). Market Opportunities, Genetic Endowments, and Interfamily Resource Distribution: Child Survival in Rural India. *The American Economic Review*, 72(4), pp.803-815.

Ryan, J. (1982). Wage functions for daily labor market participants in rural south India.

Sánchez-Ramos, E., Monárrez-Espino, J. and Noyola, D. (2017). Impact of vaccination on influenza mortality in children <5years old in Mexico. *Vaccine*, 35(9), pp.1287-1292.

Segovia, B., Suárez, J. and Aguirre, E. (2009). Factores maternos asociados a la duración de la lactancia materna en un Hospital Regional. *Pediatría de México*, 11(1), pp.15-19.

Sen, A. (1990). More Than 100 Million Women Are Missing. *The New York Review of Books*.

Sen, A. and Sengupta, S. (1983). Malnutrition of Rural Children and the Sex Bias. *Economic and Political Weekly*, 18(19/21), pp.855-864.

Sovero, V. (2017). Risk preferences and child investments: evidence from Mexico. *Review of Economics of the Household*, 16(4), pp.1027-1047.

Trivers, R. (1974). Parent-Offspring Conflict. *Integrative and Comparative Biology*, 14(1), pp.249-264.

UNICEF. (2013). *Statistics*. [online] Available at: https://www.unicef.org/infobycountry/india_statistics.html.

UNICEF. (2013). *Statistics*. [online] Available at: https://www.unicef.org/infobycountry/india_statistics.html [Accessed 22 Mar. 2019].

UNICEF. (2013). *Statistics*. [online] Available at: https://www.unicef.org/infobycountry/mexico_statistics.html [Accessed 22 Mar. 2019].

Visaria, L. (1985). Infant Mortality in India: Level, Trends and Determinants. *Economic and Political Weekly*, [online] 20(34), pp.1447-1450. Available at: https://www.jstor.org/stable/4374747 [Accessed 2 Feb. 2019].

Waldron, I. (1987). Patterns and causes of excess female mortality among children in developing countries.

Who.int. (2009). *WHO | Age of administration of rotavirus vaccines*. [online] Available at: https://www.who.int/vaccine_safety/committee/topics/rotavirus/rotarix_and_rotateq/7Aug_2009_age_administration/en/ [Accessed 27 Feb. 2019].

Who.int. (2013). *WHO | Diarrhoea-related hospitalizations in children before and after implementation of monovalent rotavirus vaccination in Mexico*. [online] Available at: https://www.who.int/bulletin/volumes/92/2/13-125286/en/ [Accessed 12 Feb. 2019].

Woetzel, J., Madgavkar, A., Ellingrud, K., Labaye, E., Devillard, S., Kutcher, E., Manyika, J., Dobbs, R. and Krishnan, M. (2015). *How advancing women's equality can add $12 trillion to global growth*. McKinsey&Company.

World Economic Forum (2017). *The Global Gender Gap Report 2017*. [online] World Economic Forum. Available at: http://www3.weforum.org/docs/WEF_GGGR_2017.pdf [Accessed 18 Apr. 2019].

9. Appendix:

I. Variable List with respective locations

Variable	Book Number	Section	Variable Name	Database
Ever Breastfed	Book 4	iv_he2	he47	Individual
Breastfeeding Duration in Months	Book 4	iv_he2	he48_4	Individual
Has a vaccination card or not	Book V	v_vac	vac02 & vac02_a	Individual
Vaccinations	Book V	v_vac	vac02-vac04h	Individual
Gender	Book S	s_sa	sa01	Individual
Age	Book S	s_sa	sa03	Individual
Birth Month	Book S	s_sa	sa02_2	Individual
Sibling Characteristics				
Number of siblings ever born	Book 4	iv_res	generated	Individual
Number of brothers ever born	Book 4	iv_res	generated	Individual
Number of sisters ever born	Book 4	iv_res	generated	Individual
Mother's Information				
Mother's Age	Book C	c_ls	Is02_2	Household
Mother's Education Level	Book 3A	iiia_ed	ed06	Individual
Mothers age when first married	Book 4	iv_ac	ac35_1	Individual
Mothers age during 1st pregnancy	Book 4	iv_he1	he07	Individual
Does the mother speak Spanish?	Book 3A	iiia_ed	ed01	Individual
Part of an indigenous group?	Book 3A	iiia_ed	ed03	Individual
Prenatal Characteristics				
Number of prenatal check-ups	Book 4	iv_he2	he21_2	Individual
Given Tetanus shots?	Book 4	iv_he2	he25c	Individual
Home/Non-home childbirth	Book 4	iv_he2	he31_1	Individual

II. Do-files

Appendix 2(a): Dataset 1

This data set is used for the regressions and the demographic analysis

- Calculating the age of children in months:

The Mexican Family Life Survey (MxFLS) records the age in terms of years only while the age in months is required for this project. We merge the date of birth from Book S section s_sa with the date of interview from the control book (Book C, section c_conpor) in order to calculate how old the child was in months at the time of the interview.

It was noticed that the dates from each book were recorded in different formats which would not allow Stata to calculate the difference between them in terms of months. Hence, the following commands are used to align the format of the interview year to the birth year:

- **gen interviewyear = 2000 + anio**

The survey only provides the month and year in which the interview took place, hence we assume that the interview took place on the 28th of every month to account for the short month of February and to simplify the study from any leap year calculations:

- **gen intday = 28**

The following commands were used to combine the day, month, and year variables into a single date variable for the interview date and birth date respectively, this would allow Stata to treat the numbers as calendar dates:

- **gen interviewdate = mdy (mes, intday, interviewyear)**
- **format %td interviewdate**

- **gen birthdate = mdy (sa02_2, sa02_1, sa02_3)**
- **format %td birthdate**

Finally, the data is suitable to calculate the age of the children in months:

- **`gen ageinmonths = interviewdate - birthdate`**
- **`drop if ageinmonths >15`**
- **`drop if ageinmonths <0 (removing discrepancies due to estimated interview date)`**

- Combining child and mother personal identification numbers (pid_link):

The first step is to merge the previous file with the control book (Book C, section c_ls) using the personal ID (pid_link) variable. A few discrepancies were noticed in the pid_link variable as some of them did not equate to household ID (folio) plus the individual sequence number within a household (ls) as described by MxFLS. A filter was done to drop any observations for which the pid_link did not meet this definition:

- **`gen pid_link2 = folio + ls`**
- **`drop if pid_link != pid_link2`**
- **`drop pid_link2`**

We save this file as “Accurate Book C” and will be using it instead of “Book C” for the rest of the steps. Continuing with the merge, we drop any observations which did not match during the merge. This would leave the data set containing variables of only the children between 0-15 months. Within this data set is a variable which identifies the individual sequence number within a household (ls) of the child’s mother; ls07. Using this we generate the pid_link of each child’s mother.

- **`gen momlineno = string ( ls07, "%02.0f")`**
- **`gen momID = folio + momlineno`**

We further filter the data set using the ls07 variable to exclude the children whose mothers are deceased, do not live in the same HH, do not personally take care of the child or are not recorded in the survey.

- **`Drop if ls07==51`** → Does not live in household
- **`Drop if ls07==53`** → Deceased
- **`Drop if ls07!=ls08`** → Does not personally care for child
- **`Drop if missing(ls07)`** → Not included in survey

Finally, we eliminate any twins in order to determine the gender of the youngest child in a household. In order to identify twins, we use the following set of commands:

- **`Sort momID ageinmonths`**
- **`By momID (ageinmonths): egen x=sum(ageinmonths)`**
- **`Browse if x!=ageinmonths`**

We then manually drop twin observations and keep only the youngest child per mother.

- Control Variable 1: Mom's age

From this point onwards, all control variables will be adjusted in their initial files and then merged into the main data set by using the 'merge' command.

To add the mothers' age, we will need to merge the Accurate Book C but using the mother's pid-link. Hence, we would need to rename the momID variable to pid-link in the main data set for Stata to match the mothers and not the children:

- **`Rename pid_link ChildID`**
- **`Rename momID pid_link`**

Next, the age variable from the existing Book C (c_ls) would need to be deleted so that Stata can introduce it again using the merge function:

- **`Drop ls02_2`**
- **`Merge 1:1 pid_link using "Accurate Book C.dta", gen(momagemerge)`**
- **`Drop if momagemerge==2`**

- Control Variable 2: Mothers Education dummy

The education levels of mothers can be found in Book 3A (section iiia_ed) which is merged to the main data set. We then create education dummy variables for the mothers. In this project, we group the education levels into the following dummies:

Education Level	Incudes	Dummy Name
Elementary or less	No education, Preschool/Kinder, Elementary	elem
Secondary	Secondary and Open secondary	sec
High School	High school, Open high school, Normal basic	hs
College	College	college
Graduate Degree	Graduate	grad

In the cases where no education level was recorded, we assume no education was attained.

```
. gen elem=1 if ed06 <=3
(377 missing values generated)

. replace elem=0 if ed06>3
(377 real changes made)

. gen sec=1 if ed06>=4 & ed06<=5
(308 missing values generated)

. replace sec=0 if ed06<4
(167 real changes made)

. replace sec=0 if ed06>5
(141 real changes made)

. gen hs=1 if ed06>=6 & ed06<=8
(442 missing values generated)

. replace hs=0 if ed06<6
(403 real changes made)

. replace hs=0 if ed06>8
(39 real changes made)

. gen college=1 if ed06==9
(507 missing values generated)

. replace college=0 if ed06!=9
(507 real changes made)

. gen graduate=1 if ed06==10
(542 missing values generated)

. replace graduate=0 if ed06!=10
(542 real changes made)
```

- Control Variable 3: "Speakspan" dummy

Whether a mother speaks Spanish or not is recorded in Book 3A (section iiia_ed). We create a speakspan dummy variable as follows:

```
. gen speakspan=1 if ed01==1
(7 missing values generated)

. replace speakspan=0 if ed01==3
(7 real changes made)
```

- Control Variable 4: Mothers' age at first marriage

A mothers age at her first marriage is found in Book 4 (section iv_ac). The variable is suitable for use as it is, so we directly merge it to the main data set without any changes.

- Control Variable 5: Mothers' age at first pregnancy

A mothers age at her first pregnancy is found in Book 4 (section iv_he1). After dropping any other irrelevant variables contained in this file, we drop all observations for which the variable 'secuencia' is greater than 1 to ensure that only the age at first pregnancy remains.

- **`Drop he07a – he15_1`**
- **`Drop if secuencia > 1`**

- Control Variable 6: Number of Prenatal check-ups

The number of prenatal check-ups the mother had during the pregnancy of the child is recorded in Book 4 (section iv_he2). Unlike the previous control variable (5), dropping observations for which 'secuencia' is greater than 1 would keep only the youngest child. Moreover, since zero check-ups are left unrecorded, we replace them with 0 to ensure an accurate regression outcome.

- **`Drop if secuencia != 1`**
- **`Replace he21_2=0 if he21_1==2`**

- Control Variable 7: Tetanus Shots dummy

The tetanus shot data is found in Book 4 (section iv_he2) too, therefore, we filter the data to eliminate observations where secuencia is greater than 1. We then create a dummy variable for whether or not the mother received a tetanus vaccine during the pregnancy.

```
. gen TetanusShots=1 if he25c==1
(60 missing values generated)

. replace TetanusShots=0 if he25c==3
(51 real changes made)
```

- Control Variable 8: Home child birth dummy

The location of where the childbirth occurred is also recorded in Book 4 (section iv_he2). For this dummy variable, we only require to know whether the childbirth took place at home or not. Therefore, any birth which took place at home, with or without doctor/midwife, is considered as a homebirth.

```
. gen Homebirth=1 if he31_1 >=10 & he31_1 <=12
(362 missing values generated)

. replace Homebirth=0 if he31_1 <10
(345 real changes made)

. replace Homebirth=0 if he31_1 >12
(17 real changes made)
```

- Control Variable 9: Mothers' ethnicity

The mother's ethnicity in this study is represented by whether or not she belongs to an indigenous group. This information is recorded in Book 3A (section iiia_ed). We generate the following dummy variable:

```
. gen IndigenousGroup=1 if ed03==1
(318 missing values generated)

. replace IndigenousGroup=0 if ed03==3
(318 real changes made)
```

- Control Variable 10: Total siblings

To find the total number of siblings we calculate the total number of children the mother ever had (i.e. both still alive or dead) and then deduct 1 to account for the child included in the regression. The total number of children the mother ever had is calculated by adding the following variables from Book 4 (section iv_res):

a. res03: Number of sons alive and living with you
b. res04: Number of daughters alive and living with you
c. res07: Number of sons alive and not living with you
d. res08: Number of daughters alive and not living with you
e. res10: Number of sons born who died
f. res11: Number of daughters born who died

- **`gen totalchildren = res03+res04+res07+res08+res10+res11`**
- **`gen totalsiblings= totalchildren - 1`**

A few discrepancies were spotted in the records where a child's gender was reported differently in different books. For example, the mother may have recorded having only one son but this child is recorded as a female in the control book and vice versa. We eliminate those observations after merging the two data sets (as it is not possible to establish their actual gender) using the following commands:

- **`Drop if totalchildren==1 & res03==1 & res03!=childgender`**
- **`Drop if totalchildren==1 & res04==1 & res04==childgender`**

- Control Variable 11: Total brothers

To find the number of brothers, we use a similar approach to that used for total siblings. We begin by generating a variable that adds the total number of sons the mother has from the variables in Book 4 (section iv_res):

a. res03: Number of sons alive and living with you
b. res07: Number of sons alive and not living with you
c. res10: Number of sons born who died

we then use the following command to find the number of brothers the child has:

- **`gen totalbrothers = totalsons - 1 if childgender==1`**
- **`replace totalbrothers = totalsons if childgender==0`**

That is, if the child included in the regression is a male (childgender==1), then we deduct 1 from the number of sons the mother has in order to find the number of brothers the child has. On the other hand, if the child is a female (childgender==0), the number of brothers she has would equal to the number of sons the mother has recorded.

- Control Variable 12: Total sisters

Applying the same logic used to calculate the number of brothers, we begin by calculating the number of daughters the mother ever had:

a. res04: Number of daughters alive and living with you
b. res08: Number of daughters alive and not living with you
c. res11: Number of daughters born who died

following through, we calculate the number of sisters the child ever had using these commands:

- **`gen totalsisters = totaldaughters - 1 if childgender==0`**
- **`replace totalsisters = totaldaughters if childgender==1`**

That is, if the child included in the regression is a female (childgender==0), we deduct 1 from the number of daughters the mother has in order to find the number of sisters the child has. On the other hand, if the child is a male (childgender==1), the number of sisters he has would equal to the number of daughters the mother has recorded.

Certain discrepancies were spotted in the data from the survey where the gender of the child does not correspond with the gender identified by the mother. This was avoided using the following commands:

- **`drop if totalbrothers>totalsiblings`**
- **`drop if totalsisters>totalsiblings`**

- <u>Generating Child Gender dummy variable:</u>

To avoid inconsistencies, we drop observations in the main data set which have contradicting genders recorded in different books:

- **`Drop if ls04!=sa01`**

The independent variable in this study is the child's gender which we will use '1' to represent a male and '0' to represent a female:

```
. gen ChildGender=1 if sa01==1
(172 missing values generated)

. replace ChildGender=0 if sa01==3
(172 real changes made)
```

- <u>Dependent variable 1: VacCard</u>

If a mother owns a vaccination card for her child but does not have it during the interview, we consider it as owning a vaccination card.

- **`Rename vac02 currentcard`**
- **`Rename vac02_a previouscard`**
- **`gen VacCard=1 if currentcard==1`**
- **`replace VacCard=1 if previouscard==1`**
- **`replace VacCard=1 if currentcard==2`**

- `replace VacCard=0 if currentcard==3`

We do not consider previouscard==2 or previouscard==3 as no observations belong to these categories.

- Dependent variable 2: FullVac

As summarised in Appendix I, the vaccination information is found in Book V (section v_vac). The vaccinations are given names such as vac03 and vac04. Therefore, we renamed the variables with their actual vaccination names followed by the dose number. For example, variable Vac03r_1 refers to dose 1 of the triple viral vaccine required for children with a current vaccination card and Vac03r_2 refers to dose 1 of triple viral vaccine required for children with a previous vaccination card. These variables were renamed to C_TRVSRP_1 and P_TRV_1 respectively. Similarly, Vac03s_1 and Vac03s_2 refer to dose 2 of the triple viral vaccine required by current and previous vaccination card holders respectively. Those variable names were changed to C_TRVSRP_2 (2nd dose of triple viral for current card holders) and P_TRV_2 (2nd dose of triple viral for pervious card holders).

The survey leaves the entries for those who did not receive a certain vaccination blank. We change those blanks to 0 to suit the final dummy variable we are trying to achieve; whether a child is fully vaccinated or not. We use the following list of command to convert blanks to zeros:

- `Replace C_BCG="0" if currentcard==1 & missing(C_BCG)`
- `Replace C_HepB_1="0" if currentcard==1 & missing(C_HepB_1)`
- `Replace C_HepB_2="0" if currentcard==1 & missing(C_HepB_2)`

.

(the same process is repeated for all vaccinations)

.

- `Replace P_TRV_2="0" if previouscard==1 & missing(P_TRV_2)`

The survey also assigns 45 to those observations who have not yet reached the required age to receive a vaccine. We change any 45 to missing in order for Stata to not consider these observations:

- **Replace C_BCG="." if C_BCG=="45"**
- **Replace C_HepB_1="." if C_HepB_1=="45"**
- **Replace C_HepB_2="." if C_HepB_2=="45"**

.

(the same process is repeated for all vaccinations)

.

- **Replace P_TRV_2="." if P_TRV_2=="45"**

The next step has to do with the format in which the observations are recorded in the survey. If a child has received a certain vaccine, it is recorded with the date at which he or she received that vaccine. Since we only want to know whether a child has received a vaccine or not, we replace those dates with "1" using the following formulas:

- **Replace C_BCG="1" if currentcard==1 & C_BCG!="." & C_BCG!="0"**
- **Replace C_HepB_1="1" if currentcard==1 & C_HepB_1!="." & C_HepB_2!="0"**
- **Replace C_HepB_2="1" if currentcard==1 & C_HepB_2!="." & C_HepB_2!="0"**

.

(the same process is repeated for all vaccination)

.

- **Replace P_TRV_2="1" if previouscard==1 & P_TRV_2!="." & P_TRV_2!="0"**

Before creating the dummy variable of being fully vaccinated, we breakdown the immunization process into the following three categories:

1. months_2_vac
2. months_4to6_vac
3. months_12to15_vac

Where each category checks whether a child has received the required vaccines in a certain age-range. For example, months_2_vac checks whether a child has received all the vaccines he or she needed to get in the first 2 months since birth. As for months_4to6_vac, it checks whether a child has received the vaccines required between the age range of 4 to 6 months since birth. Those variables are generated using the following commands:

- **gen months_2_vac=1 if C_BCG=="1" & C_HepB_1=="1" & C_HepB_2=="1" & C_PentAcc_1=="1" & C_Rota_1=="1" & C_PneumConj_1=="1"**
- **replace months_2_vac=1 if P_BCG=="1" & P_Polio_1=="1" & P_Pent_1=="1"**
- **replace months_2_vac=0 if C_BCG=="0" | C_HepB_1=="0" | C_HepB_2=="0" | C_PentAcc_1=="0" | C_Rota_1=="0" | C_PneumConj_1=="0"**
- **replace months_2_vac=0 if P_BCG=="0" | P_Polio_1=="0" | P_Pent_1=="0"**

- **gen months_4to6_vac=1 if C_PentAcc_2=="1" & C_PentAcc_3=="1" & C_HepB_3=="1" & C_Rota_2=="1" & C_PneumConj_2=="1"**
- **replace months_4to6_vac=1 if P_Polio_2=="1" & P_Polio_3=="1" & P_Pent_2=="1" & P_Pent_3=="1"**
- **replace months_4to6_vac=0 if C_PentAcc_2=="0" | C_PentAcc_3=="0" | C_HepB_3=="0" | C_Rota_2=="0" | C_PneumConj_2=="0"**
- **replace months_4to6_vac=0 if P_Polio_1=="0" | P_Polio_3=="0" | P_Pent_2=="0" | P_Pent_3=="0"**

- **gen months_12to15_vac=1 if C_PneumConj_3=="1" & C_TRVSRP_1=="1"**
- **replace months_12to15_vac=1 if P_TRV_1=="1"**
- **replace months_12to15_vac=0 if C_PneumConj_3=="0" | C_TRVSRP_1=="0" |**
- **replace months_12to15_vac=0 if P_TRV_1=="0"**

Where "months_2_vac" is equal to 1 if the child received the required vaccinations for the ages 2 months after birth, 0 if he/she hasn't and missing if the child has not reached the required age to receive all vaccines. The same applies to "months_4to6_vac" and "months_12to15_vac".

With this being done, we now generate the FullVac dummy variable using the following list of commands. A child is considered fully vaccinated if he/she received all the vaccines required for his/her age, which is why Fullvac is equal to one even if “months_4to6_vac” and “months_12to15_vac” are missing (as they have not reached the required age yet).

- **gen fullvac=1 of months_2_vac==1 & months_4to6_vac==1 & months_12to15_vac==1**
- **replace fullvac=1 if vac04a==1 & vac04b==1 & vac04d==1 & vac04e==1**
- **replace fullvac=1 if months_2_vac==1 & months_4to6_vac==1 & missing(months_12to15_vac)**
- **replace fullvac=1 if months_2_vac==1 & missing(months_4to6_vac) & missing(months_12to15_vac)**
- **replace fullvac=0 if months_2_vac==0 | months_4to6_vac==0 | months_12to15_vac==0**

Vac04a/b/d/e are the variables which indicate the vaccines received by those who own a vaccination card but did not have it during the interview. Some observations did not know whether or not the child received a certain vaccine, therefore they were recorded as “8”. For these cases we consider that the child did not receive that vaccine and hence replace all “8” to “3”

- **replace vac04a=3 if vac04a==8**
- **replace vac04b=3 if vac04b==8**
- **replace vac04d=3 if vac04d==8**
- **replace vac04e=3 if vac04e==8**

Following these changes, we amend the FullVac variable to include these changes:

- **replace fullvac=0 if vac04a==3 | vac04b==3 | vac04d==3 | vac04e==3**

- Extra variables generated:

The following variables were also generated in the main data set to help in the demographic analysis:

- **gen agerange=″0-5″ if ageinmonths>=0 & ageinmonths<=5**
- **replace agerange=″5-10″ if ageinmonths>5 & ageinmonths<=10**
- **replace agerange=″10-15″ if ageinmonths>10 & ageinmonths<16**
- **gen motheragerange=″13-23″ if motherage>=13 & motherage<=23**
- **replace motheragerange=″23-33″ if motherage>23 & motherage<=33**

- **replace motheragerange=”33-43” if motherage>33 & motherage<=43**
- **replace motheragerange=”43-63” if motherage>43 & motherage<=63**

The final mother age range is 20 years because only 9 observations are within this range and hence has no major effect on the purpose of creating this which is to find the age ranges of the mothers who were included/omitted from the regressions (demographic analysis section 6.2.1 in the dissertation).

Appendix 2(b): Dataset 2

This data set is used to analyze son-biased stopping rules in Mexico and it includes the youngest child per mother between ages of 0-120months:

Starting with the Control book (Book C, section c_ls) and merging it with Book S section s_sa:

- **`Gen interviewday=28`**
- **`Gen interviewyear=2000+anio`**
- **`Gen interviewdate=mdy(mes, intday, interviewyear)`**
- **`Format % td interviewdate`**
- **`Gen birthdate=mdy(sa02_2, sa02_1, sa02_3)`**
- **`Format % td birthdate`**
- **`Gen ageinmonths=(interviewdate-birthdate)/12`**
- **`Drop if ls04!=sa01`**
- **`Gen momlineno=string(ls07, "%02.0f")`**
- **`Gen momID=folio + momlineno`**
- **`Drop if ls07==53`**
- **`Drop if ageinmonths>120`**

The data set so far can be used to analyse ratio of all male to all female children who are 0-120 months old. The following steps are taken to filter the data to only the youngest children:

- **`Sort momID ageinmonths`**
- **`By momID (ageinmonths): egen x=sum(ageinmonths)`**
- **`Br if x!=ageinmonths`**

. (Manually drop twins)

- **`Sort momID ageinmonths`**
- **`By momID (ageinmonths): gen siblingsno=_n`**
- **`Drop if siblingno!=1`**

The data set can now be used to check the gender ratios across the youngest children in families at different ages to check if mothers stop having children after giving birth to a preferred gender.

To check for the proportion of males within a certain age range, we create an agerange variable as follows:

- **`gen agerange="0-6" if ageinmonths>0 & ageinmonths<6`**
- **`replace agerange="6-12" if ageinmonths>=6 & ageinmonths<12`**

.

(continue replacing for every 6 month interval)

.

- **`replace agerange="114-120" if ageinmonths>=114 & ageinmonths<120`**
- **`gen male=1 if sa01==1`**
- **`replace male=0 if sa01==3`**

III. Logit regressions

Table 6: Tabulated LOGIT regression outcomes showing the effect of child gender on being fully vaccinated by parents

	(1)	(2)	(3)	(4)	(5)
VARIABLES	fullvac	fullvac	fullvac	fullvac	fullvac
childgender	-0.00952	0.0769	-0.279	-0.314	-0.268
	(0.175)	(0.184)	(0.290)	(0.297)	(0.301)
birthmonth		-0.0558**	-0.0287	-0.0181	-0.0187
		(0.0269)	(0.0391)	(0.0398)	(0.0404)
ageinmonths		0.156***	0.159***	0.147***	0.134***
		(0.0234)	(0.0367)	(0.0377)	(0.0385)
motherage			0.0341	0.0537	0.186*
			(0.0459)	(0.0518)	(0.0959)
elem			-13.49	-13.31	-12.08
			(1,097)	(1,105)	(631.4)
sec			-13.65	-13.55	-12.38
			(1,097)	(1,105)	(631.4)
hs			-13.63	-13.57	-12.44
			(1,097)	(1,105)	(631.4)
college			-13.36	-13.34	-12.18
			(1,097)	(1,105)	(631.4)
o.speakspan			-	-	-
indigenousgroup			-0.360	-0.496	-0.315
			(0.406)	(0.423)	(0.444)
agemotherfirstmarried			-0.0194	-0.0203	-0.0379
			(0.0787)	(0.0807)	(0.0838)
agemotherfirstpreg			-0.0139	-0.0256	-0.140
			(0.0813)	(0.0852)	(0.113)
tetanusshots				0.325	0.320
				(0.422)	(0.427)
homebirth				0.597	0.585
				(1.258)	(1.264)
numbercheckupsduringpreg				-0.00905	-0.0118
				(0.0421)	(0.0425)
totalbrothers					-0.912**
					(0.415)
totalsisters					-0.206
					(0.400)
Constant	0.667***	-0.207	13.20	12.79	11.62
	(0.125)	(0.282)	(1,097)	(1,105)	(631.4)
Observations	579	579	244	239	239

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

note: speakspan != 1 predicts success perfectly

: speakspan dropped and 1 obs not used

Table 7: Tabulated LOGIT regression outcomes showing the effect of child gender on being ever breastfed by mother

VARIABLES	(1) everbreastfed	(2) everbreastfed	(3) everbreastfed	(4) everbreastfed	(5) everbreastfed
childgender	-0.0552	-0.0256	-0.285	-0.532	-0.607
	(0.372)	(0.373)	(0.531)	(0.641)	(0.656)
birthmonth		0.0186	-0.000838	0.0552	0.0545
		(0.0530)	(0.0742)	(0.0859)	(0.0863)
ageinmonths		0.0444	0.0441	-0.0175	-0.00714
		(0.0455)	(0.0614)	(0.0713)	(0.0713)
motherage			0.00799	0.286	0.0544
			(0.0806)	(0.195)	(0.244)
elem			-12.17	-9.996	-9.998
			(2,389)	(1,512)	(1,512)
sec			-12.73	-11.45	-11.48
			(2,389)	(1,512)	(1,512)
hs			-12.38	-11.09	-11.12
			(2,389)	(1,512)	(1,512)
college			-12.98	-11.59	-11.61
			(2,389)	(1,512)	(1,512)
o.speakspan			-	-	-
indigenousgroup			0.891	0.310	0.143
			(1.053)	(1.096)	(1.099)
agemotherfirstmarried			0.140	0.198	0.217*
			(0.107)	(0.124)	(0.127)
agemotherfirstpreg			-0.115	-0.420**	-0.206
			(0.110)	(0.204)	(0.247)
tetanusshots				1.457**	1.397**
				(0.634)	(0.640)
o.homebirth				-	-
numbercheckupsduringpreg				-0.0369	-0.0320
				(0.0686)	(0.0683)
totalbrothers					1.138
					(1.226)
totalsisters					1.049
					(1.259)
Constant	2.658***	2.190***	14.59	11.57	11.74
	(0.267)	(0.543)	(2,389)	(1,512)	(1,512)
Observations	461	461	307	296	296

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
note: speakspan != 1 predicts success perfectly
: speakspan dropped and 1 obs not used
note: homebirth != 0 predicts success perfectly
: homebirth dropped and 5 obs not used

Table 8: Tabulated LOGIT regression outcomes showing the effect of child gender on owning a vaccination card

VARIABLES	(1) VacCard	(2) VacCard	(3) VacCard	(4) VacCard	(5) VacCard
childgender	0.276	0.369	0.251	0.256	0.232
	(0.284)	(0.290)	(0.448)	(0.462)	(0.469)
birthmonth		-0.0847*	-0.0707	-0.0715	-0.0542
		(0.0448)	(0.0655)	(0.0682)	(0.0701)
ageinmonths		0.150***	0.180***	0.175***	0.181***
		(0.0365)	(0.0560)	(0.0588)	(0.0618)
motherage			-0.0338	0.0431	0.187
			(0.0571)	(0.0783)	(0.121)
elem			-11.89	-11.96	-11.49
			(1,389)	(1,389)	(1,149)
sec			-12.16	-12.13	-11.74
			(1,389)	(1,389)	(1,149)
hs			-10.89	-11.00	-10.57
			(1,389)	(1,389)	(1,149)
college			-11.90	-12.01	-11.29
			(1,389)	(1,389)	(1,149)
o.speakspan			-	-	-
indigenousgroup			-0.220	-0.331	-0.0712
			(0.602)	(0.615)	(0.672)
agemotherfirstmarried			0.0839	0.0727	0.0768
			(0.102)	(0.106)	(0.105)
agemotherfirstpreg			-0.0227	-0.0887	-0.247*
			(0.101)	(0.114)	(0.148)
tetanusshots				1.074**	1.066*
				(0.545)	(0.548)
o.homebirth				-	-
numbercheckupsduringpreg				0.0284	0.0209
				(0.0725)	(0.0747)
totalbrothers					-0.352
					(0.330)
totalsisters					-1.053*
					(0.558)
Constant	2.410***	1.992***	13.21	11.95	11.59
	(0.191)	(0.447)	(1,389)	(1,389)	(1,149)
Observations	740	740	318	305	305

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
note: speakspan != 1 predicts success perfectly
: speakspan dropped and 1 obs not used
note: homebirth != 0 predicts success perfectly
: homebirth dropped and 5 obs not used