Son Preference and Health Disparities in Developing Countries

Le, Kien and Nguyen, My

2022

Son Preference and Health Disparities in Developing Countries

Kien Le & My Nguyen

Ho Chi Minh City Open University

Abstract: Recorded history demonstrates the preference for sons in every aspect of life. Today, despite being considered a powerful manifestation of gender inequality and discrimination against women, the preference for sons over daughters is still prevalent worldwide. In this study, we investigate the extent to which son preference influences health disparities between sons and daughters in 66 developing countries. We find that the differences in height-for-age and weight-for-age z-scores between daughters and their peers are 0.135 and 0.098 standard deviation lower compared to the analogous differences between sons and their peers due to son preference. Our heterogeneity analysis further shows that son preference disproportionately affects children of disadvantaged backgrounds such as those living in rural areas, born to lower-educated mothers, and coming from poor families.

Keywords: Son preference; health disparities; developing countries.

1 Introduction

The preference for sons over daughters is prevalent worldwide, especially in Asia and North Africa. This is one of the most persistent gender issues in many societies with sons receiving preferential treatment over daughters. Recorded history demonstrates the preference for sons in every aspect of life ranging from royal (e.g. succession laws) to peasant families (e.g. land inheritance). The deep-rooted preference for sons arises and persists until today by a variety of socioeconomic, cultural, and institutional factors. For example, births of daughters in South Asian countries are often considered as an economic liability due to the dowry system in which the bride's family has to give to the groom durable assets as a condition of the marriage. Another example is the role of ancestor worship in Sinosphere countries (e.g. China, Vietnam, and Korea) where there is a belief about the afterlife and the need for the sons to perform rituals of ancestor worship in order to ensure the welfare of not only the departed souls but also entire family line.

Prior studies have documented that the preference for sons is a major source of selective abortion among females resulting in skewed population sex ratios that substantially favor males in countries with strong son preference traditions (Hesketh & Xing, 2006; Dubuc & Coleman, 2007; Abrevaya 2009; Scharping 2013; Almond et al., 2013; Bharadwaj & Lakdawala, 2013). Given the consensus in the literature on the relationship between son preference and population sex ratios, recent analyses have begun to shift their focus on discriminatory treatments towards surviving girls. With sophisticated statistical approaches and detailed micro-data, several studies have shown that parents with a preference for sons discriminate against daughters when they distribute scarce resources such as breastmilk, sources of vitamin and protein, health care, and time spending (Jayachandran & Kuziemko, 2011; Barcellos et al., 2014; Baker & Milligan, 2016; Aurino, 2017).

In this study, we investigate the extent to which son preference influences health disparities between sons and daughters in early childhood. By doing so, the contribution of the study is three folds. First, our work complements studies identifying factors affecting child health to support policymakers in developing effective mitigation strategies. Second, we provide additional evidence on the less salient effects of son preference on early human health, whereas other studies tend to explore the more discernible effects at the aggregate level (e.g. sex ratios and marriage patterns). Finally, our study sample does not just focus on one particular country, but spreads across 66 countries covering children born between 1990 and 2018. The wide temporal and spatial coverage could make our results meaningful to policymakers in many countries where son preference is prevalent.

To quantify the impacts of son preference on early childhood health disparities between sons and daughters, we employ the Demographic and Health Surveys for information on children as well as their parents. Our empirical strategy is the household fixed effects model that exploits the differences in the health outcomes of children born to mothers living under the same roof but with different degree of son preference. The results of our study can be summarized as follows. First, we do not detect the beneficial effects of son preference on male children. However, there exists a negative and statistically significant impacts of son preference on health disparities between sons and daughters. Particularly, due to son preference, the differences in height-for-age and weight-for-age z-scores between the daughters and their peers are 0.135 and 0.098 standard deviation lower compared to the analogous differences between the sons and their peers. Second, the negative effects of son preference on health disparities tend to concentrate on those living in rural areas, born to lower-educated mothers, and coming from poor families.

Our findings highlight the serious cost of the preference for sons in terms of early human health. To the extent that poor health in early life exerts long-lasting irreversible consequences over the life cycle such as cognitive impairment, learning difficulties, higher vulnerability to chronic diseases, and decreased productivity as well as earnings (Martorell, 1999; UNICEF & WHO, 2019), son preference may impede long-term human development and gender inequality. Hence, the study calls for additional efforts in putting an end to son preference.

Our paper proceeds as follows. Section 2 reviews the literature. Section 3 describes the data. Section 4 presents the empirical methodology. Section 5 discusses the results and concludes.

2 Literature Review

The existing literature suggests at least three possible channels through which the preference for sons might arise and persist, including institution, economic incentive, and culture. First, institutional and societal rules can give rise to the preference for sons. For example, people living in societies where property and land rights favor males (e.g. Islamic laws) often prefer sons because they prefer their possessions to be passed on to their own children upon their death. Therefore, the relative demand for daughters is lower in countries that limit inheritance or bequests to females (Carranza, 2012).

Second, parents may have economic incentives to prefer sons if they expect to receive a higher return or more financial support from the sons (Pande and Astone, 2007). This expectation is particularly relevant to societies with a high economic return to physical strength where males have a comparative advantage. Parents may further reduce their investment in female children leading to wide gender gaps in various aspects such as health and human capital (Qian, 2008; Pitt, Rosenzweig and Hassan, 2012).

Third, cultural norms and religious practices can also affect parental preference for sons. For example, traditions of marital exogamy where daughters leave home upon marriage have been suggested to contribute to son preference (Dyson and Moore, 1983). Another example is the need for the sons to perform rituals of ancestor worship or funeral rituals through which son preference might arise (Pande and Astone, 2007; Jayachandran, 2015).

It is important to note that these channels do not work separately, but rather intertwine together. For example, customs such as the dowry system (i.e. the bride's family has to give to the groom durable assets as a condition of the marriage) or the eldest son responsibility (usually in supporting parents as they age) are considered as cultural aspects but also offer economic incentives to increase the demand for sons.

Our quantitative analysis of the impacts of son preference on child health is guided by the Grossman theory (Grossman, 1972). The central proposition of the theory is that health can be viewed as a durable capital stock producing an output of healthy time. It is assumed that health depreciates over time and can be increased by investment in health inputs (e.g. nutrition, vitamin supplement, medical services, etc.). Since parents with a preference for sons are more likely to rationally invest in daughters and sons differently, one might expect that the preference for sons can influence the health outcomes of children based on their gender.

Empirically, our study is related to two strands of literature. The first line of literature focuses on various outcomes of children being affected by their gender. For example, Jayachandran & Kuziemko (2011), Barcellos et al. (2014), and Aurino (2017) show that Indian parents tend to favor boys in the intra-household allocation of childcare time, breastmilk, and sources of protein as well as vitamins. Hafeez & Quintana-Domeque (2018) also confirm the existence of son-biased preferences in the duration of breastfeeding in Pakistan. Within the context of the U.S, the U.K, and Canada, Baker & Milligan (2016) find that fathers are more likely to commit more time to sons, and the son-daughter differences exist even for twins. Employing a longitudinal data set from Indonesia, Palloni (2017) reports that children born of their mother's preferred gender tend to weigh more and experience fewer illnesses.

The study can also be related to studies exploring various socioeconomic, religious, and cultural factors affecting child health. For example, it is documented that adverse economic shocks, such as economic crisis and labor demand shocks, can reduce household living standards, thus worsening health outcomes of children (Stillman & Thomas, 2008; Page, Schaller & Simon, 2019). Religious practice such as Ramadan fasting has also been shown to negatively affect child health (Almond and Mazumder, 2011). Prior studies have also established a negative relationship between various forms of political violence, such as armed conflicts and terrorism, and child health (Minoiu and Shemyakina, 2012; Le and Nguyen 2020; Shemyakina, 2021). Besides, researchers have begun to explore the impacts of climate change on early childhood health recently, including the adverse consequences of rainfall shocks and extreme temperature (Molina & Saldarriaga, 2017; Le & Nguyen, 2021).

3 Data

The data on children are obtained from the Demographic and Health Survey (DHS). The DHS is a global health and population survey that is administered in more than 90 developing nations around the world. Our analyses utilize the Woman's Questionnaire of the DHS that targets women of reproductive ages (15-49) and collects information on their background characteristics (age, age at birth, fertility, education, etc.), the characteristics of their children (gender, age, birth order, etc.), and health outcomes of the children.

Anthropometric z-scores, such as height-for-age and weight-for-age, which are collected for children under the age of five, are used to measure child health in this study. Each of the anthropometric z-score captures the number of standard deviations below or above the corresponding median value of the international reference population accounting for sex and age. A low height-for-age z-score is a result of the deficiency of growth-supporting nutrients or recurrent illnesses, and a low weight-for-age z-score reflects impaired development and vulnerability to disease as well as illness (WHO, 2008).

More importantly, the DHS asks the respondents questions about their ideal number of sons and daughters to investigate their son-biased preferences. Following the literature, our main explanatory variable, *Son Preference*, is the degree of son preference measured by the ratio of the desired number of sons to the desired number of total children. This variable takes a value of one if sons are strictly preferred, zero if daughters are strictly preferred, and 0.5 if the ideal number of sons and daughters are exactly the same.

Our final estimation sample consists of over one million under-five children spreading across 66 countries covering children born between 1990 and 2018. We report the list of countries in Table A1 of the Appendix. Descriptive statistics for the dependent (outcome) and independent (explanatory) variables are reported in Table 1. As reported in Panel A, the average height-for-age and weight-for-age z-scores are -1.256 and -1.141 standard deviations, respectively. These negative values are expected since the sample consists of mostly developing countries where child health is usually lower compared to the median of the reference population that also covers children from richer countries.

As shown in Panel B, the average value of *Son Preference* is 0.523 which is higher than the normal value of 0.5, thus confirming the overall existence of son preference in our sample. On average, the mothers are 28.14 years old at survey and 26.24 years old at birth. The mean educational years of the mothers are 5.157. Besides, the current number of own children is 3.082 and the preferred value is 3.987 on average. Around 49% of the children are female. The mean age of children is 24.52 months. The average birth order is 3.09. Approximately 1.1% of the births are plural births.

Table 1: Summary Statistics

	Mean (1)	SD (2)	N (3)
Panel A: Dependent Variables			
Height-for-age Z-score	-1.256	1.554	1,079,421
Weight-for-age Z-score	-1.141	1.321	1,079,421
Panel B: Independent Variables			
Son Preference	0.523	0.149	1,079,421
Mother's Age	28.14	6.437	1,079,421
Mother's Age at Birth	26.24	6.286	1,079,421
Mother's Education	5.157	4.918	1,079,421
Number of children	3.082	1.904	1,079,421
Preferred Number of children	3.987	2.361	1,079,421
Daughter	0.490	0.500	1,079,421
Child's Age in Months	24.52	70.11	1,079,421
Child's Birth Order	3.090	2.176	1,079,421
Being a Plural Birth	0.011	0.102	1,079,421

4 Empirical Methodology

To quantify the relationship between son preference and child health outcomes, we estimate the following regression equation,

$$Y_{ijts} = \beta_0 + \beta_1 SP_{ijts} + \beta_2 SP_{ijts} \times D_{ijts} + \delta_j + \theta_t + \lambda_s + X'_{ijts}\Omega + \epsilon_{ijts} \qquad (1)$$

where the subscripts *i*, *j*, *t*, and *s* corresponds to child, household, month-year of birth, and survey month-year, respectively. The variable Y_{ijts} represents child health outcomes measured by the anthropometric z-scores of height-for-age and weight-for-age. The variable SP_{ijts} (*Son*

Preference) presents the degree of son preference ranging from zero (daughters are strictly preferred) to one (sons are strictly preferred). The variable D_{ijts} (*Daughter*) is a zero-one indicator taking a value of one if the child is female, and zero otherwise.

We also denote by δ_j, θ_t, and λ_s household, birth month-year, and survey month-year fixed effects, respectively. The vector X'_{ijts} is a covariate of the child and mother's characteristics, including: (i) child's gender, age in months, squared-age in months, birth order, plural birth indicator, birth month-year fixed effects, and (ii) mother's age, squared-age, age at birth, squared-age at birth, years of education, the number of children, and the preferred number of children. Finally, ϵ_{ijts} is the error term. Since the source of variation in this model is within and across households, standard errors throughout the paper are clustered at the household level.

The coefficients β_1 and β_2 capture the impacts of son preference on child health. In particular, the coefficient β_1 presents the estimated impacts of son preference on the health outcomes of sons (i.e. when D_{ijts} is zero for male child, we have $\beta_1 SP_{ijts} + \beta_2 SP_{ijts} \times 0 = \beta_1$). The sum $\beta_1 + \beta_2$ reflects the estimated impacts of son preference on the health outcomes of daughters (i.e. when D_{ijts} is one for female child), and the coefficient β_2 shows the disparity in the health outcomes of sons and daughters due to son preference. In other words, β_2 quantifies the differences between the health outcomes of the daughter (with respect to the international reference, i.e. other girls at the same age) and the son (with respect to the international reference, i.e. other boys at the same age) due to son preference. In this paper, we are particularly interested in health disparities between sons and daughters due to son preference, i.e. the magnitude and statistically significant level of the coefficient β_2.

In this empirical setup, we exploit the variation in the health outcomes of children born to mothers living in the same house but having different preferences for sons. The inclusion of household fixed effects is expected to capture factors suggested by the literature that could jointly affect son preference and child health at the same time, such as economic, institutional, religious, and cultural factors discussed in Section 2.

5 Results

5.1 Main Results

The estimated impacts of son preference on health disparities between sons and daughters in terms of height-for-age and weight-for-age are provided in Panels A and B of Table 2. Here, Column 1 displays the estimates where we only control for the main explanatory variables (i.e. *Son Preference* and the interaction between *Son Preference* and *Daughter*) and child characteristics (i.e. child's gender, age in months, squared-age in months, birth order, plural birth indicator, birth month-year fixed effects). In Column 2, we additionally control for mother characteristics (i.e. mother's age, squared-age, age at birth, squared-age at birth, years of education, the number of children, and the preferred number of children). In Column 3, we introduce survey month-year and residential cluster fixed effects to the regressions (a residential cluster can be thought of as a small neighborhood). Finally, Column 4 presents our most extensive specification where we replace the residential cluster fixed effects with household fixed effects.

According to Column 1, son preference is associated with 0.149 and 0.305 standard deviation reductions in height-for-age and weight-for-age of sons. More importantly, the disparities in the health outcomes between sons and daughters are 0.322 standard deviations in height-for-age and 0.380 standard deviations in weight-for-age. However, the estimates only reflect the correlation between son preference and health outcomes as important factors that could jointly affect preference and health are not accounted for. For example, mothers with a low level of education tend to live in brawn-based societies (e.g. live in rural areas, work in the agricultural sector, etc.) and have less access to old-age pension, leading to higher demand for sons. These mothers, at the same time, are more likely to have less healthy children and more likely to sacrifice investment in daughters due to limited budget compared to highly educated mothers (Le and Nguyen, 2020).

To address such issues, we additionally control for mother characteristics in Column 2. The estimates become substantially smaller in magnitude suggesting that much of the effects found in Column 1 are actually due to the characteristics of the mothers instead of son preference. Failing to control for mother characteristics would bias our estimates. Similarly, spatial and temporal dimensions are also critical. For example, people living in conservative regions or surveyed in the 90s might have a higher demand for sons. For some unobserved reasons, their children might not be as healthy as those residing in progressive regions or surveyed more recently, and they might not have enough resources to ensure the well-being of both sons and daughters leading to the sacrifice of daughters. This issue could also bias our estimates. Therefore, we account for locational and temporal heterogeneities with the inclusion of survey month-year and residential cluster (a small neighborhood) fixed effects in Column 3. Here, we find that son preference is associated with a 0.040 standard deviation increase in height-for-age and a 0.028 standard deviation increase in weight-for-age of the sons. The disparities in the health outcomes between sons and daughters are 0.114 standard deviations in height-for-age and 0.086 standard deviations in weight-for-age. The large changes in the coefficient magnitudes suggest that a part of the son preference is actually a proxy for locational and temporal effects in determining health outcomes.

Table 2: Son Preference and Child Health - Main Results

	(1)	(2)	(3)	(4)
Panel A: Y = Height-for-age Z-score				
Son Preference	-0.149***	-0.058***	0.040***	0.038
	(0.014)	(0.013)	(0.014)	(0.041)
Son Preference x Daughter	-0.322***	-0.220***	-0.114***	-0.135***
	(0.020)	(0.019)	(0.019)	(0.034)
Observations	1,079,421	1,079,421	1,068,524	563,314
Panel B: Y = Weight-for-age Z-score				
Son Preference	-0.305***	-0.208***	0.028**	0.023
	(0.012)	(0.012)	(0.011)	(0.035)
Son Preference x Daughter	-0.380***	-0.281***	-0.086***	-0.098***
	(0.017)	(0.017)	(0.016)	(0.028)
Observations	1,079,421	1,079,421	1,068,524	563,314
Fixed Effects - Households	.	.	.	X
Fixed Effects - Clusters	.	.	X	.
Mother Characteristics	.	X	X	X
Child Characteristics	X	X	X	X

Note: *p<0.1, **p<0.05, ***p<0.01. Robust standard errors are clustered at the household level. Each column represents the coefficient in a separate regression. Child Characteristics include child's gender, age in months, squared-age in months, birth order, plural birth indicator, birth month-year fixed effects. Mother Characteristics include mother's age, squared-age, age at birth, squared-age at birth, years of education, number of children, and preferred number of children. Fixed Effects - Clusters include survey month-year and residential cluster fixed effects. Fixed Effects - Households include survey month-year and household fixed effects.

Despite an exhaustive set of child's characteristics, mother's characteristics, locational and temporal fixed effects, one might still concern that there could still exist unobservables not presented in the data but can simultaneously affect child health and son preference. For example, our world has become more diverse in the past decades. In many places, especially big cities, people have become used to neighbors from different cultural, religious, and racial backgrounds. Although the advantages of diversity are well established, there remain many challenging issues to societies. For example, people having the same level of education and living in the same area can still face discrimination based on their culture, religion, and skin color. If discrimination is correlated with such individual backgrounds (thus son preference) and child health (e.g. disparities in treatments at hospitals) simultaneously, then our estimates can still be biased. Therefore, we present our most extensive specification where we replace the residential cluster fixed effects with household fixed effects in Column 4. This specification is expected to capture all of the factors outside the family (e.g. social, institution, cultural, etc.) that could affect son preference and child health at the same time. In other words, we exploit the variation in the health outcomes of children born to mothers in the same families but having different preferences for sons to identify the impacts of interest.

According to our most extensive specification in Column 4, son preference is associated with a 0.038 and 0.023 standard deviation increase in height-for-age and weight-for-age of the sons, respectively. However, the estimates are not statistically significant suggesting that there is not enough statistical evidence to conclude the relationship between the preference for sons and their health. Most importantly, we observe statistically significant evidence for the negative association between son preference and health disparities between sons and daughters. Particularly, due to son preference, the differences in height-for-age between the daughters and their international peers are 0.135 standard deviation lower compared to the difference in height-for-age between the sons and their international peers. Analogously, the difference in weight-for-age between the daughters and their peers is 0.098 standard deviation lower compared to the difference in weight-for-age between the sons and their peers because of son preference.

5.2 Heterogeneity Analysis

So far we have detected adverse impacts of son preference on the disparities in health outcomes between sons and daughters. As discussed in Section 2, the preference for sons might arise and persist through outdated characteristics of the institution, economic incentive, and culture. These characteristics are still prevalent in families with disadvantaged backgrounds. For example, families whose lives depend on agriculture tend to live in rural areas often characterized as brawn-based societies. People with low educational attainment are more likely to work in informal sectors, thus not qualified for old-age pensions and must depend on their sons to support them as they age. These families are usually poor and do not have enough resources to ensure the well-being of their sons and daughters at the same time. Even if the preference for sons is similar between a rich and a poor family, the poor one usually has to sacrifice investment in the daughters due to limited budget. Meanwhile, the rich one, after investing enough in the sons, may still have more than enough left for the daughters. Therefore, we expect that the disparity impacts of son preference may differ between families with advantaged and disadvantaged backgrounds.

In this section, we proceed to explore the heterogeneous impacts along the lines of the mother's locational status, educational attainment, and family wealth. The estimating results are displayed in Table 3. For each panel, the panel name is the dimension of heterogeneity and each column depicts a separate regression. All estimates are from the most extensive specification (as in Column 4 of Table 2).

First, we want to examine whether the disparity impacts of son preference differ between rural and urban areas. As shown in Panel A, female children born to mothers residing in rural areas bear more serious health disparities due to son preference than those born to mothers residing in urban areas. Specifically, due to son preference, the differences in height-for-age and weight-for-age z-scores between the daughters and their peers are 0.149 and 0.127 standard deviations lower compared to the analogous differences between the sons and their peers in the rural areas. The corresponding effects are much smaller in magnitude and less statistically significant among those in the urban areas.

Table 3: Son Preference and Child Health - Heterogeneity Analysis

	Height-for-age Z-score (1)	Weight-for-age Z-score (2)	Height-for-age Z-score (3)	Weight-for-age Z-score (4)
Panel A: Heterogeneity in Location				
	Rural		Urban	
Son Preference	0.019	0.014	0.092	0.049
	(0.053)	(0.043)	(0.065)	(0.057)
Son Preference x Daughter	-0.149***	-0.127***	-0.096*	-0.031
	(0.043)	(0.035)	(0.054)	(0.046)
Observations	398,053	398,053	164,221	164,221
Panel B: Heterogeneity in Maternal Education				
	Low Education		High Education	
Son Preference	0.038	-0.006	0.083	0.066
	(0.064)	(0.053)	(0.063)	(0.054)
Son Preference x Daughter	-0.166***	-0.116***	-0.051	-0.054
	(0.048)	(0.038)	(0.049)	(0.041)
Observations	320,659	320,659	223,795	223,795
Panel C: Heterogeneity in Family Wealth				
	Poor Families		Non-poor Families	
Son Preference	0.032	0.024	-0.027	-0.038
	(0.071)	(0.058)	(0.064)	(0.054)
Son Preference x Daughter	-0.151***	-0.107***	-0.068	-0.063
	(0.053)	(0.041)	(0.057)	(0.046)
Observations	225,525	225,525	211,085	211,085
Fixed Effects - Households	X	X	X	X
Mother Characteristics	X	X	X	X
Child Characteristics	X	X	X	X

Note: *p<0.1, **p<0.05, ***p<0.01. Robust standard errors are clustered at the household level. Each column represents the coefficient in a separate regression. Child Characteristics include child's gender, age in months, squared-age in months, birth order, plural birth indicator, birth month-year fixed effects. Mother Characteristics include mother's age, squared-age, age at birth, squared-age at birth, years of education, number of children, and preferred number of children. Fixed Effects - Households include survey month-year and household fixed effects.

Second, we examine if children of mothers with high and low educational attainment are differentially affected by son preference. Mothers with low education refer to those who did not complete primary education. Mothers with high education refer to those who completed primary education and above. Evident from Panel B, the disparity impacts of son preference are larger for children born to low education mothers. Particularly, the disparity estimates indicate 0.166 and 0.116 standard deviations lower in height-for-age and weight-for-age z-scores for daughters of low education mothers. We find no effects on those of high education mothers.

Finally, we test if children from poor families are differentially affected by son preference compared to those of nonpoor families. Poor families are defined as those with the wealth index lying in the bottom and the next bottom quintiles of the within-country wealth distribution. Non-poor families refer to the remaining ones in the sample. Evident from Panel C, children from poor families are disproportionately affected by son preference. Particularly, due to son preference, the differences in height-for-age and weight-for-age z-scores between the daughters and their peers are 0.151 and 0.107 standard deviations lower compared to the analogous differences between the sons and their peers in the poor families. Nevertheless, the corresponding impacts are much smaller in magnitude and statistically insignificant among those in the non-poor families.

Taken together, the heterogeneity exercise in this section confirms our expectation that the disparity impacts of son preference differ between families with advantaged and disadvantaged backgrounds (e.g. rural, low education, and poor families).

5.3 Robustness

In this section, we employ alternative health measures and model specifications to test for the robustness of our results. In Panel A of Table 4, nutrition indicators and percentile measures are utilized in place of the z-score measures. Specifically, *Being Stunt* and *Being Underweight* are dummy variables taking the value of one if height-for-age and weight-for-age z-scores are less than -2, respectively. The -2 threshold is established by WHO (2010). *Height-for-age Percentile* and *Weight-for-age Percentile* indicate the ranking of the corresponding anthropometric measures among the reference population. We still detect the adverse relationship between son preference and the health disparities between sons and daughters. Specifically, due to son preference, the differences in the probabilities of being stunt and underweight between the daughters and their peers are 2.4 and 2.0 percentage points higher compared to the analogous differences between the sons and their peers. Similarly, the differences in the rankings of height-for-age and weight-for-age between the daughters and their peers are 1.862 and 1.948 percentiles lower compared to the analogous difference between the sons and their peers.

Table 4: Son Preference and Child Health - Robustness

	(1)	(2)	(3)	(4)
Panel A: Other Health Measures				
	Indicator Measures		Percentile Measures	
	Being Stunt	Being Underweight	Height-for-age Percentile	Weight-for-age Percentile
Son Preference	-0.001	0.006	0.376	0.891
	(0.013)	(0.012)	(0.778)	(0.736)
Son Preference x Daughter	0.024**	0.020**	-1.862***	-1.948***
	(0.011)	(0.010)	(0.644)	(0.607)
Observations	563,314	563,314	563,314	563,314
Panel B: Other Specifications				
	Weighted Regressions		Excluding Teen Mothers	
	Height-for-age Z-score	Weight-for-age Z-score	Height-for-age Z-score	Weight-for-age Z-score
Son Preference	0.058	0.025	0.069	0.055
	(0.050)	(0.042)	(0.047)	(0.040)
Son Preference x Daughter	-0.129***	-0.098***	-0.132***	-0.087***
	(0.042)	(0.035)	(0.037)	(0.030)
Observations	562,885	562,885	498,947	498,947
Fixed Effects - Households	X	X	X	X
Mother Characteristics	X	X	X	X
Child Characteristics	X	X	X	X

Note: *p<0.1, **p<0.05, ***p<0.01. Robust standard errors are clustered at the household level. Each column represents the coefficient in a separate regression. Child Characteristics include child's gender, age in months, squared-age in months, birth order, plural birth indicator, birth month-year fixed effects. Mother Characteristics include mother's age, squared-age, age at birth, squared-age at birth, years of education, number of children, and preferred number of children. Fixed Effects - Households include survey month-year and household fixed effects.

In Panel B of Table 4, we introduce the sampling weights to our most extensive regressions and report the estimated results in Columns 1 and 2. The disparity impacts on height-for-age and weight-for-age z-scores are 0.129 and 0.098 standard deviations, respectively. Applying the sampling weights only affects our main results slightly. In other words, our models are robust to the inclusion of sampling weights. However, we shy away from using the sampling weights in the main regressions because several studies criticize that weighting can lower efficiency and statistical power in estimation (Winship and Radbill, 1994; Gelman, 2007; Solon et al., 2015).

Finally, we exclude teen mothers from our sample and rerun the most extensive regressions. The motivation for this exercise is that teen pregnancy might lead to poor birth outcomes, thus child health. One might concern that the estimated disparity impacts of son preference are driven by teenage mothers. Hence, we exclude mothers aged 19 and below at childbirth from our sample. The results reported in Columns 3 and 4 (Panel B of Table 4) indicate that the issue of teen pregnancy is very unlikely to drive our main results. Taken together, our conclusion on the adverse relationship between son preference and health disparities between sons and daughters remains unchanged when other measures as well as model specifications are employed.

6 Discussion and Conclusion

Collectively, we have documented compelling evidence for the detrimental effects of son preference on health disparities between sons and daughters. Specifically, due to son preference, the differences in height-for-age and weight-for-age z-scores between the daughters and their peers are 0.135 and 0.098 standard deviation lower compared to the analogous differences between the sons and their peers. Exploring the heterogeneity in the disparity impacts of son preference, we find that son preference disproportionately affect children of disadvantaged backgrounds such as those living in rural areas, born to lower-educated mothers, and coming from poor families. The results are robust to different child health measures and model specifications.

Our findings on health disparities between sons and daughters due to son preference are consistent with the literature on various outcomes of children being affected by their gender. Specifically, parents tend to favor sons in the intra-household allocation of childcare time, breastmilk, and sources of protein as well as vitamins (Jayachandran & Kuziemko, 2011; Barcellos et al., 2014; Aurino, 2017; Hafeez & Quintana-Domeque, 2018; Baker & Milligan, 2016). However, we differ from these studies by directly looking at the output of the health production function (height-for-age and weight-for-age z-scores) instead of the allocation of inputs (e.g. breastmilk, time, nutrition, etc.)

The findings in this paper highlight an important source of heterogeneity in child health. To the extent that poor health in early life exerts long-lasting irreversible consequences over the life cycle such as cognitive impairment, learning difficulties, higher vulnerability to chronic diseases, and decreased productivity as well as earnings (Martorell, 1999; UNICEF & WHO, 2019), son preference may impede long-term human development and gender inequality. Hence, the study calls for additional efforts in putting an end to son preference. Some of the mitigating measures include changing inheritance and other similar practices to raise the value of daughters, strengthening old-age pension systems to reduce the demand for sons, promoting positive images about alternative masculinity that values gender equality, preventing misuse of technology for sex selection through the strict regulation on penalties, and conducting assessments of interventions as well as monitoring sex ratio at birth regularly. Extra attention should be given to the population from disadvantaged backgrounds such as those living in rural areas, having low educational attainment, and coming from poor families.