Selective Mortality and Malnutrition in India

Panda, Pallavi

SUNY Geneseo

29 November 2019

Selective Mortality and Malnutrition in India

Pallavi Panda*

Citation: **Panda, P.** Selective Mortality and Malnutrition in India. *J. Quant. Econ.* (2019).
https://doi.org/10.1007/s40953-019-00194-8

Abstract

India presents itself as a paradox with low infant mortality and high malnutrition. This paper provides survival bias as an explanation of the paradox. Using pooled health surveys from 1993 to 2005 and a pseudo-panel selection model, this study finds that the change in Height-for-Age Z-Scores (HAZ scores) can be explained by mortality selection. Specifically, children with sample average characteristics that survive have 17.4% less HAZ scores than a child randomly drawn from the population indicating an overestimation of malnutrition in India. This is consistent with the hypothesis of weaker children surviving due to skilled delivery which pulls down the overall HAZ scores. The results are robust to controls for unobservable characteristics of groups of women. Son preference is also apparent in the results. The selection is more evident among male children and in the states where sex selection is historically seen as a problem in India.

Keywords: Infant Mortality, Child Health, Nutrition, India.

JEL Codes: J11, J13, I15.

* State University of New York, Geneseo, 1 College Circle, 117a South Hall, Geneseo, NY 14454, email: panda@geneseo.edu.

1 Introduction

Malnutrition has presented itself as a serious burden, especially in India. Malnutrition has been shown to decrease labor productivity and intelligence and lower ability to earn future income, perpetuating a vicious cycle of poverty (Belli, 1971). Due to the high socio-economic impacts of malnutrition, reduction of malnutrition and eradication of poverty and hunger is one of the UN Sustainable Development Goals. Consequently, to develop effective and evidence-based strategies to reduce undernourishment, a clear understanding of the channels is crucial.

India presents itself as a case where despite increasing per capita incomes, it has not been able to decrease malnutrition at the same rate. India has seen a declining trend in its infant and neonatal mortality rates over the last few decades. Infant mortality is defined as a child dying before the age of 1 year and neonatal mortality is defined as child dying before the age of 1 month. Many sub-Saharan African countries have lower rates of malnutrition than India, despite having much higher infant and child mortality rates and lower income per capita.[1] India grapples with this paradox of low infant mortality and high malnutrition. This paper provides evidence of mortality selection as an explanation to the paradox of low infant mortality and high malnutrition in India. This could be the reason why we are not seeing malnutrition rates keeping pace with India's recent economic growth. Due to improved health infrastructure and neonatal care, India is able to save weaker children from dying and who by surviving, lower the average anthropometric scores leading to statistics that indicate higher rates of malnutrition prevalence in the data.

[1] For example, Chad had infant mortality rate of 124 vis-a-vis 50 for India in 2009. While, Chad had 44.8% children below the age 5 as stunted while 47.9% in India for the same period (World Health Statistics Report, 2011).

Empirical estimation of the effect is important as the selection bias in this literature can work in different ways. First, it is possible that weaker malnourished children are more probable to die in the sample resulting in a sample selection such that it pushes up the anthropometric measures. On the other hand, it is possible as well that due to improved technology and skilled birthing assistance, survival of weaker children increases which instead pushes the HAZ scores down. The dominance of the second source of bias in the sample is consistent with the paradox of low infant mortality and high malnutrition in India in recent years.

Even though this paradox is widely discussed, there are not many empirical studies analyzing the paradox. Panagariya (2013) compares the malnutrition rates across India and sub-Saharan Africa and concludes that the way malnutrition is measured could be the problem as it does not take into account micro-nutrient deficiencies and other aspects of malnutrition. Jayachandran and Pande (2017) explore the role of son preference and birth order to suggest that the exceptionally high rates of stunting observed in India vis-à-vis sub-Saharan Africa is mostly due to intrahousehold allocation of resources. There is a wide literature on the possible mechanisms to explain the high malnutrition rates in India. Spears, Ghosh, and Cumming (2013) emphasizes the role of poor sanitation and shows that open defecation can account for 35 to 55 percent difference in stunting between districts in India. This paper focuses on negative sample selection bias as another explanation to high malnutrition rates in India. I use the widely used anthropometric measure, height-for-age z-score (HAZ Scores) as a measure of malnutrition to explain the paradox of low infant mortality and high malnutrition *within* India.[2] To study this selection effect, I use three

[2] HAZ Scores are the most common anthropometric measure to track malnutrition and is used by UN and WHO. I focus on height instead of weight because height is considered as a long-run measure of an individual's health (Behrman and Deolalikar, 1988). But, I do also check for robustness of results with WAZ scores.

rounds of National Family and Health Surveys (NFHS) to examine the probability of survival of a child and link it to nutritional outcomes of the surviving children.

Using pooled national health surveys across years allows us to study how a change in infant mortality affects change in rates of HAZ scores. This is an improvement over any cross-sectional analysis where it concentrates on levels rather than changes and is unable to explain the relationship between changes in the outcome and explanatory variable. Moreover, since I can observe different cohorts of women overtime, I can control for cohort specific time-invariant unobserved heterogeneity which can produce biased estimates in both mortality and malnutrition regression. As long as women belonging to certain cohorts based on their age group, location, and socio-economic characteristics are predicted to have similar unobserved characteristics, this study does better than others. In the empirical analysis, I estimate a sample selection model where in stage one, I estimate a pooled Probit model of whether or not the child is alive for all children. For those children that are currently alive, I then estimate a least squares model with state and cohort fixed effects to examine the factors influencing child malnutrition. If infant mortality is ignored and only data from those children who are currently alive is used to study nutritional outcomes, there is the possibility of sample selection bias.

The paper's findings suggest that infant mortality selection is significant and leads to an underestimation of HAZ scores in the sample, after controlling for child characteristics and mother cohort's time invariant characteristics. This result is also robust to controlling for state-time trends, flexible specification to control for survey timing, and mother's age at birth. I observe negative mortality selection in the estimates with respect to both infant and neonatal mortality. A child with sample average characteristics who survives with controls for unobservable characteristics of groups of women, has 17.4% less HAZ scores than a child randomly drawn from the population.

The negative mortality selection points towards the survival of weaker children in the sample which would have otherwise died if not saved by superior birthing technology and skills.

The mortality selection differs by the part of HAZ or WAZ distribution the child belongs to. For stunting and being underweight that are measures of severe malnourishment, I in fact find a positive selection indicating that severely malnourished children die leaving the sample with higher anthropometric scores than the population. The negative selection is observed at the upper end of the distribution. On an average, I see that the negative selection dominates for the entire sample.

The micro level health dataset also helps in identifying the heterogeneous effects based on child characteristics like birth order and gender. I observe the familiar gender and low birth order preference pattern in India (Jayachandran and Pande, 2017). With better technology, parents put more effort in saving a male child and even a lower birth order male child. This in turn reflects in a negative selection for these groups and the average HAZ scores for these groups are lower than the overall population. This has implication on the female sibling well-being in the long run as well. If households operate under constrained budgets, as is generally the case, I may see a shift of resources from the girl child to the male child leaving the girl sibling worse off. Results also suggest a spatial heterogeneity in selection. States with historically better child sex-ratios like Kerala display no mortality selection while states like Punjab and Haryana display high negative mortality selection.

Change in HAZ can be brought about by improvement in childhood nutrition which is a causal channel. However, there are other channels through which I may observe changes in stunting data. The changes in HAZ could be brought about by changes in probability of survival due to superior birthing technology which changes latent health. Another way in which the increasing trends could be explained is by fertility selection where more educated or wealthier mothers give birth or certain kind of mothers stop giving birth. It is important to understand the channels to formulate effective policy interventions. This paper provides evidence of mortality selection being a significant channel in change in HAZ scores after controlling for demographic change of mothers and types of households.

2 Literature Review

Mortality and fertility selection have been acknowledged as a potential source of bias in the literature. A large number of studies show the effect of anthropometric indicators on child mortality indicating malnourished children are more likely to die (Caulfield et al., 2004; Pelletier, 1994; Rice et al., 2000). Boerma, Sommerfelt, and Bicego (1992) study 17 cross-sectional surveys and other longitudinal data in developing countries to find that malnutrition is more prevalent for deceased children but the survivor bias is small. These studies indicate a presence of positive selection, where healthier children survive leading to lesser malnutrition for the surviving population. However, India is experiencing an opposite phenomenon. There are other studies that have been done in sub-Saharan Africa and Bangladesh to look at selection effects as well. Pitt (1997) estimates factors affecting child mortality and child health allowing for past selective fertility and mortality behavior in the context of sub-Saharan Africa. He finds fertility selection to be a significant determinant of mortality in 14 sub-Saharan African countries but with very little change in parameters when selection is accounted for. Dancer, Rammohan, and Smith (2008) study the differences in in survival probabilities by gender and consequent differences in gender-based child nutrition in Bangladesh and find that after correcting for selection, female children were more likely to have lower HAZ and WHZ scores. This paper's focus will be on the negative selection effect explaining the paradox specific to India.

Another strand of literature looks at the effect of selective mortality and nutrition on the heights of the adult population with the premise that childhood nutrition and disease environment has a significant effect on adult height (Akachi and Canning 2010; Bozzoli, Deaton, and Quintana-domeque 2009; Moradi 2010). Deaton (2007) offers a framework of scarring and selection where he explains that positive selection effect in terms of removing shorter individuals by mortality outweighs the negative scarring effect in high mortality environments, resulting in taller adults in the case of sub-Saharan Africa. Bozzoli, Deaton, and Quintana-domeque (2009) examine adult height of 31 cohorts in England, US, and 10 European countries and show that the post neonatal mortality rate of the country predicts the average adult height of the birth cohort. Unlike developed countries where fall in infant mortality is accompanied by increasing heights, Akachi and Canning (2010) do not find evidence of the same in sub-Saharan Africa. If adult height is regarded as a measure of well-being, falling infant mortality may not be contributing to increased health in different settings.

The paradox between falling infant mortality but small improvements in health status has been documented where the reductions in child mortality are brought about by directed interventions. Epidemiology literature has documented that vaccinations and interventions aimed at reducing mortality does not reduce morbidity. Pinchinat et al. (2004) find that in southern Senegal, no nutritional improvement was found in children in 1962-1992, despite a big drop in infant mortality. With medical improvements and directed efforts at reducing infant mortality, without complementary increase in protein and other micronutrient intake by children, India may be facing a similar situation.

The paper closest to this research is by Alderman, Lokshin, and Radyakin (2011), where they construct simulated nutritional status of the child population for all the children in India who die before the age of 3 years, assuming they were alive using a proportional hazard model. This is based on matching of individuals based on observable characteristics to impute HAZ scores of children who had died. They use the three rounds of NFHS to construct the simulations and find a 5 percent difference between the counterfactual and the actual height-for-age z-scores. The problem with matching on observables is that if there are unmeasured confounding factors, the analysis could be biased. In terms of controlling for unobserved mother cohort characteristics which may affect survival and malnutrition, this study would give a better estimate of the resulting selection.

3 Model

Following Pitt (1997), I can estimate the effect of selective mortality in anthropometric measures of child health by reduced form equations of infant and neonatal mortality (M) and child health (C):

$$C = X_c\beta_c + \delta_{mc}\mu_m + \nu_c = X_c\beta_c + \varepsilon_c \quad (1)$$

$$M^* = X_m\beta_m + \mu_m + \nu_m = X_m\beta_m + \varepsilon_m \quad (2)$$

Where X_c and X_m are exogenous regressors, the error in both equations contain a heterogeneous error term μ_m which determines parental preferences over unwanted births (especially in the case of India, son preference), woman or cohort specific effects of women who observe higher mortality of children tend to also have children which are malnourished (δ_{mc}), general medical interventions and shocks which may affect mortality etc. Child health data is observed only for surviving children and hence it needs to be corrected for selective mortality.

If the errors have zero means, and ν_c and ν_m are uncorrelated, then covariance between the error terms can be written as:

$$Cov\ (\varepsilon_m, \varepsilon_c) = \delta_{mc} Var(\mu_m) \tag{3}$$

Therefore, selection bias results if δ_{mc} is not equal to 0 that is there is a feedback between mortality and child health. If medical technology is saving weaker infants, which then become malnourished due to improper care, then $\delta_{mc}>0$.

In the analysis, I define survival (s) as a child not experiencing neonatal or infant mortality. If the population errors are jointly normally distributed, the health of children conditioned on survival is:

$$E(C|X_c, S^*>0) = X_c\beta_c + Cov\ (\varepsilon_m, \varepsilon_c)\lambda \tag{4}$$

where λ is the Inverse Mill's Ratio. Omitting this λ creates a bias in the estimates of β_c.

Following Alderman et al. (2011), we can consider the HAZ scores for the whole population as,

$$Z_{pop} = Z_s(1-P_d) + Z_dP_d \tag{5}$$

Where Z_s and Z_d are average Z-scores of surviving and deceased children and P_d is the proportion of deceased children. The population Z score will change if deceased children survived:

$$\theta = Z_s - Z_{pop} = (Z_s-Z_d)P_d \tag{6}$$

According to (6), $Z_s > Z_d$ and $\theta > 0$ if more malnourished children die, leading to higher existing HAZ scores of the sample than the population. Instead, if weaker children survive due to improvements in technology and other interventions such that $Z_s < Z_d$ and $\theta < 0$, then it will result in an underestimation of HAZ-score distribution in the sample.

4 Data

This analysis uses data from three waves of India's National Family Health Survey (NFHS) – 1992/93, 1998/99 and 2005/06. The NFHS follows the pattern of a standard Demographic and Health Survey and is a large-scale survey covering a representative sample of households throughout India. For the 1992/93 survey, interviews were conducted with a nationally representative sample of 88,562 households and 89,777 ever-married women in the age group 13-49, from 24 states and the then National Capital Territory of Delhi. The 1998/99 survey covered a representative sample of about 91,000 ever-married women age 15-49 from 26 states in India. NFHS-3 conducted interviews with over 230,000 women age 15-49 and men age 15-54 throughout India.[3]

The survey is administered to ever-married females and contains detailed information about their reproductive history, asset ownership, vaccinations and preventive care, reproductive health, and educational characteristics etc. Women of reproductive age are interviewed about the date of birth and death (if applicable) of their pregnancy histories. This kind of retrospective survey gives an

[3] http://www.rchiips.org/nfhs

opportunity to build an indicator of infant and neonatal mortality and in turn the survival probabilities.[4]

The NFHS provides height and weight data for children under age of 48 months in 1992/93, under age of 36 months in 1998/99, and under age of 60 months in 2005/06.[5] The NFHS contains no anthropometric information for deceased children at the time of their death. The NFHS collects information on weight at birth in addition to weight at the time of the survey and asks mothers to categorize the weight of their children at birth as large, average, or small. It also collects information on height/length of the child and age for children up to five years of age. In the sample, I have 766364 children born to 234548 mothers. Out of these, 72116 children have non-missing data on HAZ scores and 81018 children have data on WAZ scores.

Table 1: Summary Statistics

	(1)	(2)	(3)	(4)	(5)
	Obs	**Mean**	**Std. Dev.**	**Min**	**Max**
Infant Mortality	766364	0.0818	0.274	0	1
Neonatal Mortality	766364	0.0511	0.22	0	1
HAZ Score	72116	-2.13	1.7	-6	6
Stunted (HAZ<-2)	72116	0.546	0.497	0	1
Normal (HAZ>-1)	72116	0.22	0.414	0	1
WAZ Score	81018	-1.84	1.3	-5.99	4.93
Underweight (WAZ<-2)	81018	0.444	0.496	0	1
Normal (WAZ>-1)	81018	0.252	0.434	0	1
Age in Months	90818	30.5	12.4	10	59
Height (cm)	74795	82.9	12.17	0	922.2
Weight (Kg with decimal)	81499	10.48	2.91	0.5	97
Female	766364	0.48	0.499	0	1

4 One problem that can be raised with the recall data is the measurement error problem. Since the birth histories do not go too much into the past, it is lesser of a problem in this case. Moreover, since deaths of a child are important in a mother's life, this variable should be recorded without much measurement error.
5 I also check for the mortality selection effect if I constrain the results to children born under 36 months across different survey years in Table 6. The results are unchanged.

Multiple Births	766364	0.013	0.113	0	1
Birth Order	766364	2.71	1.79	1	18
Mother's age	766364	35.4	7.74	14	49
Mother's age at birth	766364	23.3	5.12	13	48
Uneducated	766364	0.59	0.49	0	1
Rural	766364	0.68	0.47	0	1
Poor	759987	0.29	0.45	0	1
Weight of mother (Kg)	474833	48.29	10.41	24	179.2
Father Uneducated	761687	0.325	0.47	0	1
Hindu	766364	0.749	0.43	0	1
Muslim	766364	0.138	0.35	0	1
SC/ST	766364	0.294	0.45	0	1
Home Delivery	99187	0.639	0.48	0	1
Public Delivery	99187	0.198	0.40	0	1
Private Delivery	99187	0.156	0.37	0	1
Delivery, Skilled Personnel	99282	0.444	0.49	0	1
Access to piped water	766364	0.192	0.39	0	1
No access to toilet	766364	0.567	0.49	0	1
Electricity	760183	0.642	0.47	0	1

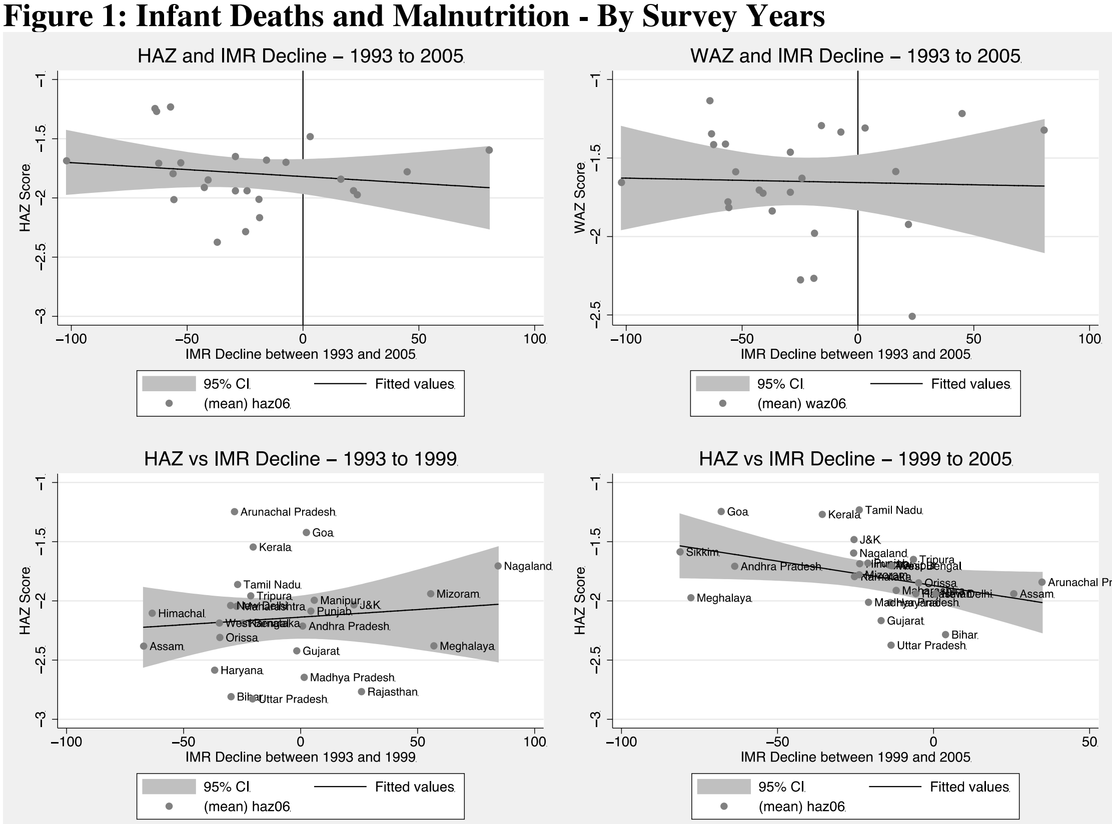
Note: Sample statistics of the variables are reported. Educated implies having attended any type of school and uneducated is defined as mother did not attend any school. Poor is defined by a mother not owning any asset as collected in NFHS. Rural and Urban are defined by the place of residence of mother during the time of interview. Female is 1 if sex of child is female. Multiple birth is a dummy variable indicating if the child is born in a multiple birth. It is 0 for a single birth and 1 for twins, triplets, or quadruplets.

In Table 1, I show the summary statistics. The average infant mortality rate is 82 children per 1000 while the average neonatal mortality rate in the sample is 51 per 1000 children. The mean HAZ score is -2.13 and mean WAZ score is -1.84. About 55% of the children in the sample are stunted (HAZ<-2) and 44% are underweight (WAZ<-2). Over the survey years, HAZ has been increasing and infant mortality has been falling (Appendix Figure 1). The infant mortality was about 92 deaths per 1000 in 1993 and has fallen to 70 deaths per thousand live births. Mean HAZ scores did not change a lot between 1993 and 1998, with it being around -2.3 but it increased by 2005 with the score at -1.89. This is shown in Fig. 4(f). The average mother's age at birth is 23 years. Around

64% mothers have delivery at home and 44% mothers have their delivery by a skilled birth attendant.

Figure 1 plots the relationship between HAZ and WAZ and decline in infant mortality. There is a strong relationship (higher confidence) between HAZ and IMR decline over the entire period and specifically in the later years of 1999 to 2005, which is the period of more technological advancements. The scatterplots show the states. Bihar, Madhya Pradesh, Rajasthan, and Uttar Pradesh lie below the fitted line with lower HAZ for the mortality decline. These states have traditionally fared worse in human development indicators in India and are termed as "BIMARU" states and exhibit lower HAZ in this sample as well.

Figure 1: Infant Deaths and Malnutrition - By Survey Years

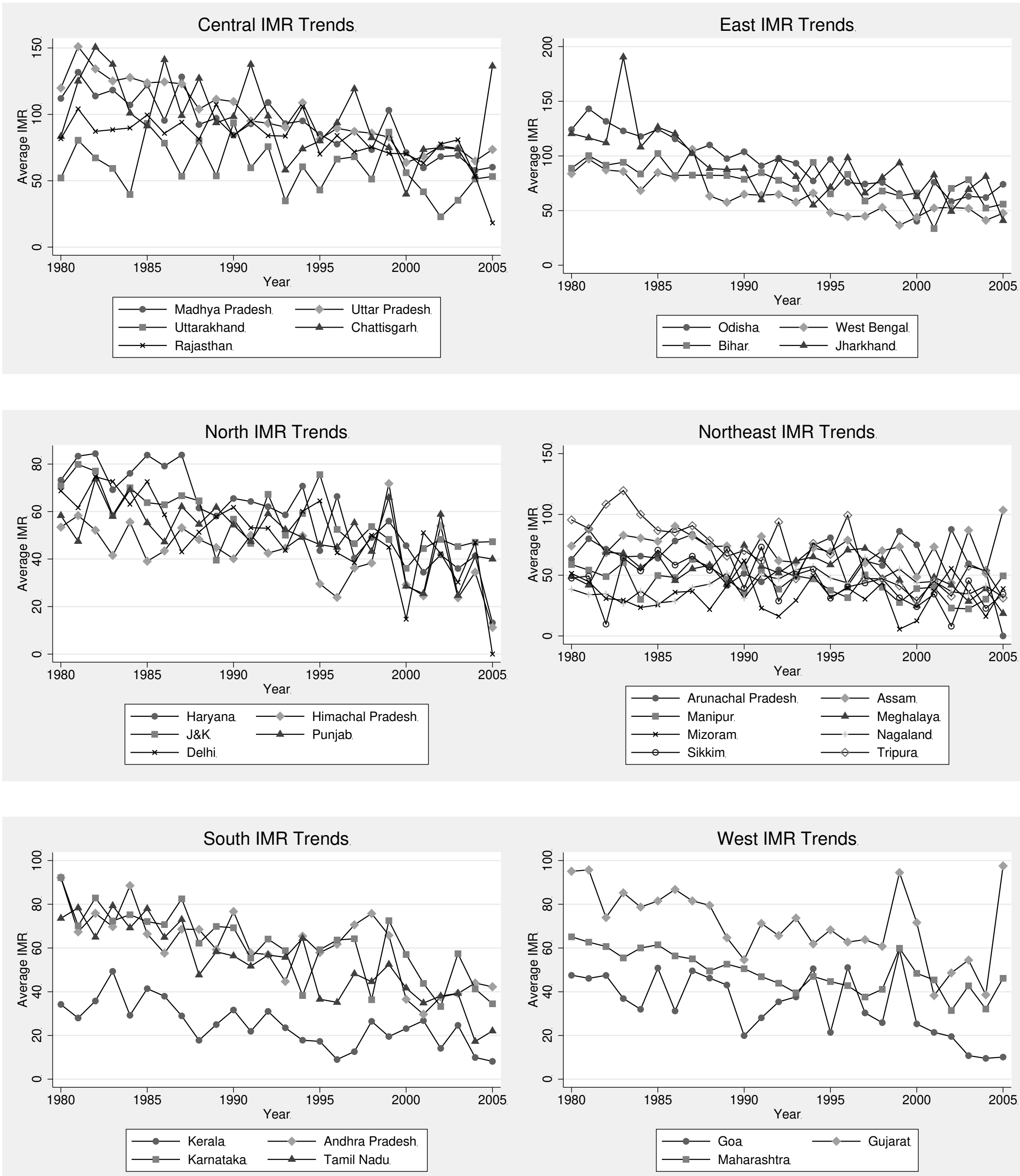


Note: In the first row, these graphs plot the decline in sample infant mortality deaths for India between 1993 and 2005. First panel plots the fitted line and the confidence interval of the relationship between IMR Decline Rate and HAZ scores. Second panel plots the relationship between WAZ scores and IMR Decline rate. In the second row, these graphs plot the decline in sample infant mortality deaths for various states between 1993 and 1999 (first and second NFHS Surveys) and 1999 and 2005 (second and third NFHS surveys). For both the panels, a fitted line is drawn to indicate the relationship between IMR Decline Rate and HAZ scores.

Figure 2 plots the sample mean infant mortality rates by year in different regions of India. Most states exhibit a declining trend in infant mortality over the years. Kerala stands out in the southern region with a distinctly lower level of mortality rate. This paper argues that continued increases in skilled delivery has been able to save more children, resulting in negative selection in child health. Therefore, we should see increases in skilled delivery rates overtime. Figure 3 plots the trends in skilled delivery rates as well as the place of delivery overtime. Consistent with the hypothesis, we see an increasing trend in skilled delivery. We also observe that there is a change in preference in the place of birth for the child – there are more deliveries taking place in public and private hospitals and a marked decline in home deliveries.

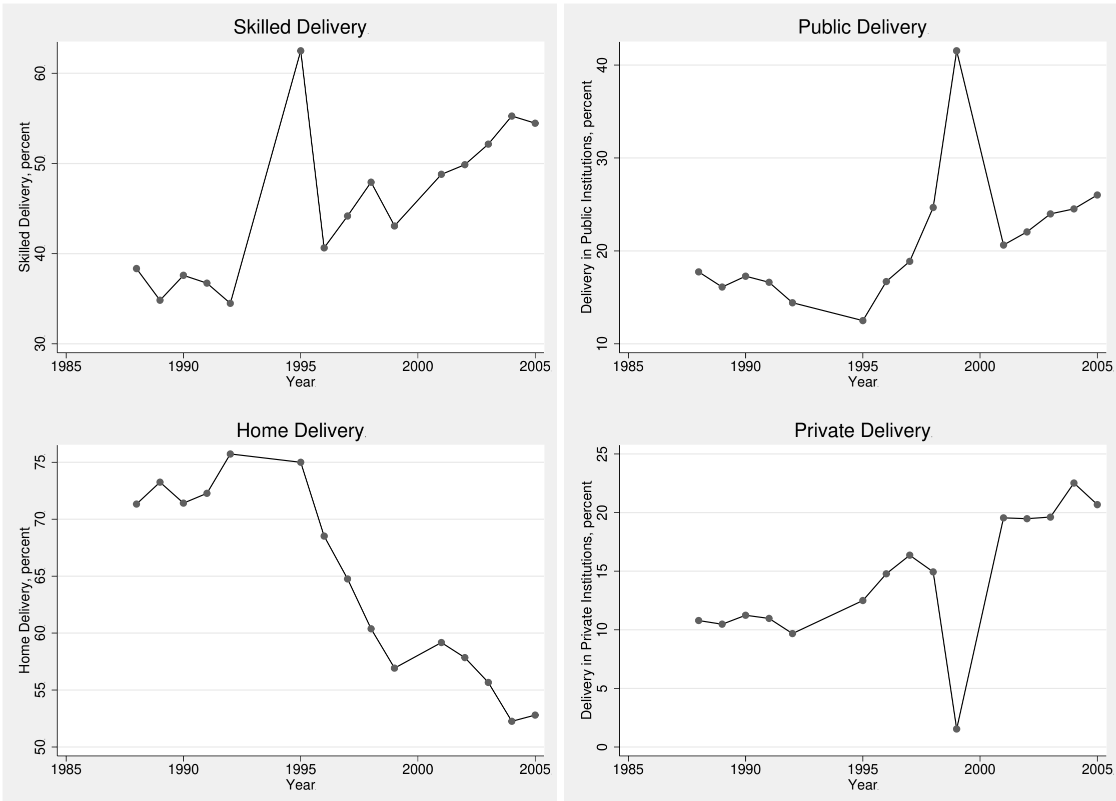
Lastly, it is imperative to look at the demographic makeup of mothers in the sample. A mother is labeled as educated if she has attended any type of school and uneducated if mother did not attend any school. Since NFHS does not have income data, definition of poor is based on possession of assets. NFHS asks easy-to-collect data on a household's ownership of selected assets, such as radio, car, television, refrigerator, and bicycles. Therefore, for this analysis, the household is categorized as poor in the event of absence of these assets in the household at the time of interview. Rural or urban are defined by the place of residence of mother during the time of interview. Around 59% of women interviewed are uneducated, 68% live in rural areas and 29% households have no assets. Figure 4 plots the difference in HAZ densities by mother characteristics. HAZ distributions for rural, uneducated, and poor women are skewed to the left. Mothers of different age groups also showcase different HAZ distributions. Keeping this in mind the mother cohorts are constructed later.

Figure 2: Infant Mortality Rates - By Region



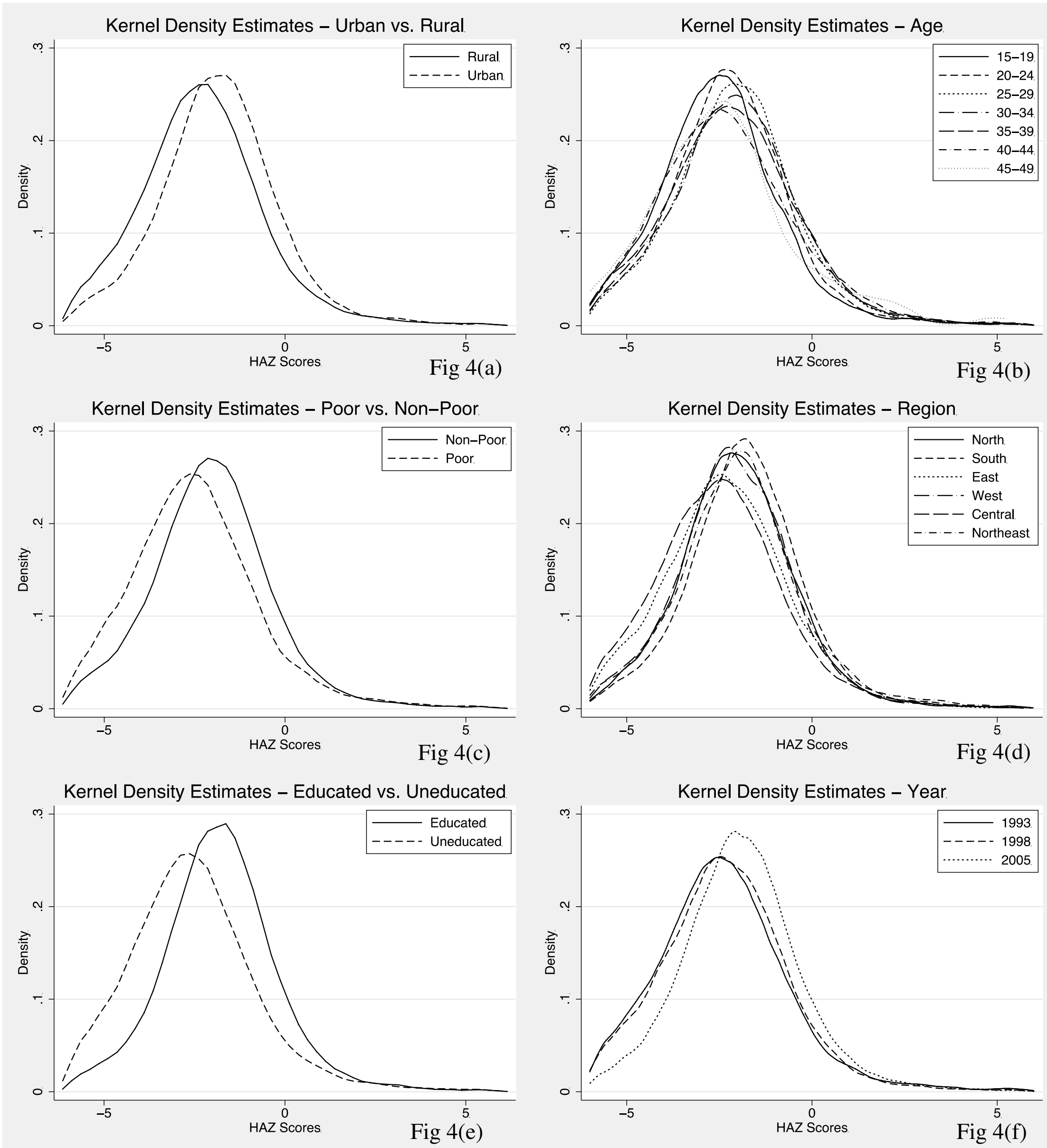
Note: These graphs plot the mean infant mortality rates over time for different geographic regions of India. North region consists of Haryana, Punjab, Jammu and Kashmir, Himachal Pradesh, and New Delhi. South consists of Kerala, Karnataka, Andhra Pradesh, and Tamil Nadu. East region has the states of West Bengal, Odisha, Bihar, and Jharkhand. West has Goa, Gujarat, and Maharashtra. Central is divided into Madhya Pradesh, Uttar Pradesh, Uttarakhand, Chhattisgarh, and Rajasthan. North-east consists of Assam, Arunachal Pradesh, Meghalaya, Mizoram, Nagaland, Sikkim, Manipur, and Tripura.

Figure 3: Trends in Skilled Delivery



Note: These graphs plot the percent of births being performed by skilled personnel (delivery being assisted by doctors, nurse/midwife, auxiliary midwife, ayurvedic doctor, and any other India-specific health professional) and the place of delivery being at public institutions, home, or private institutions using three rounds of NFHS Data between 1993 to 2005.

Figure 4: Variation in HAZ Scores Density



Note: These graphs plot the kernel density estimates of HAZ Scores by various mother, year and region characteristics. Fig 4(a), 4(b), 4(c), and 4(e) plot the kernel density by mother characteristics like place of residence, age group, possession of assets and education respectively. Fig 4(d) plots the HAZ scores by region and 4(f) plots it by year of survey.

5 Empirical Strategy

I have pooled cross-section data overtime where same individuals are not observed in different time periods and interview years. I expect individual heterogeneity to be present in the error term. Moreover, the selection process may be varying with time and individual.[6] Deaton (1985) emphasizes on using cohort to obtain consistent estimates in pseudo-panels even in the case of correlation between individual effects and explanatory variables.

The individual heterogeneity can be written as woman's cohort effect (denoted by the subscript w) plus individual deviation (denoted by the subscript i). The child health equation in our case can be written with HAZ score (HAZ_{it}) as the dependent variable:[7]

$$HAZ_{it} = X_{it}\beta + \alpha_i + \mu_{it} \tag{7}$$

$$\text{Where, } \alpha_i = \Sigma\varnothing_w\alpha_w + \sigma_i \tag{8}$$

X_{it} are the independent variables expected to affect child malnutrition that vary over individual and time. This analysis controls for gender, multiple births, month of birth of the child, survey year, birth order, interaction between being a female and birth order, caste of the household, religion of the household, mother's age at birth, whether the residence has access to piped water, has a toilet, and has electricity. The error term consists of α_i and μ_{it}. The child health equation estimated by pooled OLS will be biased if correlated individual heterogeneity (α_i) is present or selection process is nonrandom and α_i is not a random component of the error. This would be the case if there are

6 If the selection process is identical over time, then the fixed effects estimator will remove the selection bias in a panel data.

7 I use the 2006 WHO standards for HAZ, computed using the Stata package "haz06" which requires the data for age, height, and gender from NFHS. I have also run this using the HAZ scores reported in the NFHS. But, it does not change the results qualitatively.

certain woman cohort specific characteristics (α_w) like being in a rural area, being poor, being uneducated, or being a younger mother, which affects both the probability of survival of the child and nutrition that affects the error term. To eliminate these effects, fixed effects method can be used where cohort-specific dummies ($\Sigma\varnothing_w\alpha_w$) control for the cohort effect. Since deviation from cohort is independent from the selection process itself, the correlation between them now should be zero which would lead to efficient estimates.

Rodríguez and Muro (2014) develop a selection bias estimation for pseudo-panel data and show that using probit model with cohorts as instruments, a cohort level inverse Mills ratio can be derived and which can be replaced in the child health equation to derive estimates of selection bias. Accordingly, the selection equation is estimated by a probit model with controls for cohort dummies. A cohort is defined by a group of mothers belonging to the same age group (based on their year of births), residence, education, and economic condition. Since there are considerable differences in the distribution of HAZ scores by these characteristics, the mother-cohort groups are constructed keeping that in mind.[8] Controlling for mother cohorts also controls for changing demography and fertility which could be instrumental in advancing a change in HAZ, which I may erroneously attribute to improvement in nutrition.

The survival equation, indicating selective mortality is described by:

$$S_{it}^* = Z_{it}\gamma + \eta_i + u_{it}\ ;\ S_{it} = 1\ [S_{it}* > 0] \qquad (9)$$

[8] Including different characteristics with missing values, increases the number of cohorts but decreases the observations within cohorts; which is not desirable computationally and otherwise for consistency (Borjas and Sueyoshi 1994)

Survival (S_{it}) is the observation of child not dying in the data till the age of 1 year and in case of neonatal survival, till the age of 1 month. Z_{it} are the independent variables expected to affect child survival and are the same as used in the child health equation. Individual heterogeneity is present in the error term (η_i and u_{it}), which again is controlled for by using cohort-specific dummies.

For us to be able to identify the parameters of reduced form mortality selection and determinants of health, I need at least one exogenous variable that affects mortality or survival but does not affect child health and anthropometric scores. It has been noted in the literature that to avoid weak identification, there should be some variables in the selection equation which are not there in the child health equation so that the effect is not identified solely off of nonlinearity in the inverse Mills ratio (Little 1985; Vella 1998). It is also important to check for high correlations between inverse Mills ratio and the regressors in substantive equation to ensure the efficiency of the estimates (Bushway, Johnson, and Slocum 2007).[9] Moreover, not having any excluded variable may inflate the standard errors and give unreliable estimate of β in the substantive child health equation. To account for these problems, in this analysis, this excluded variable is if the woman has her delivery by a professional doctor or nurse.[10] After controlling for observable characteristics like gender of the child, being born in a multiple birth, month of birth of the child, survey year, birth order, interaction between being a female and birth order, caste of the household, religion of the household, mother's age at birth, whether the residence has basic amenities like access to piped water, a toilet, and electricity and running the fixed effects analysis for mothers within the same economic condition, area of residence, age, and education; having a delivery done by a medical

[9] Correlation of inverse Mills ratio with other independent variables are provided in the Appendix Table A2.
[10] Skilled delivery assistance is measured by delivery being assisted by doctors, nurse/midwife, auxiliary midwife, ayurvedic doctor, and any other India-specific health professional. It is not considered as skilled assistance if the baby is born with the help of trained birth attendant, traditional birth attendant, relatives, other persons, or no one.

professional affects the probability of survival of the child but does not *directly* affect the HAZ score of the child later on.

In the parametric estimation, after the estimation of selection equation over all observations, the inverse Mills ratio is constructed, which is the ratio of probability density function and the cumulative distribution function of the standard normal distribution. This ratio then is used as an additional regressor in the child health equation to consistently estimate the parameters:

$$HAZ_{it} = X_{it}\beta + \widehat{\theta\lambda_{it}} + \mu_{it} \qquad (10)$$

The equation controls for all the independent variables in the selection equation, except for the excluded variables (delivery by skilled birth attendant). If θ is statistically significant, it points towards a selection effect operating through selective mortality on health measures. If $\theta>0$, there is a positive selection meaning those who survive have a higher HAZ than a random drawing from the population with the same characteristics. If $\theta<0$, there is a negative selection meaning those who survive have a lower HAZ than a random drawing from the population with the same characteristics. Moreover since this equation is able to control for mother-cohorts, any cohort specific time invariant heterogeneity is taken care of. Any health specific shocks and survey-year differences are captured through the year dummies. Standard errors are clustered at the mother-cohort level to improve inference.

In India, especially in rural areas, studies have shown that differential health outcomes can be expected by birth order of the child and also by gender (Behrman and Taubman 1986; Horton 1988; Jayachandran and Pande 2017; Savage et al. 2013). To test if there is a differential selection effect, I run the child health specification separately for children with first and second birth order versus children of later birth order. The negative selection effect should be prominent in later birth children as they are more probable to be saved by neonatal care units but get lower nutrition inputs in a budget constrained household. Similarly, I postulate that females with higher birth order should be more affected than males or females in lower birth order. Heterogeneity is also expected by state, place of residence, and time.

6 Results

6.1 Selection Effects

The mortality selection effect is captured by the coefficient on inverse Mills ratio and is presented in Table 2. Columns (1)-(3) provide the effect of inverse Mills ratio on HAZ and WAZ scores which conditions for survival of children till 1 year of age. Probit estimates and correlations between inverse Mills Ratio and covariates are provided in Appendix Table A1 and A2. The results indicate that skilled birth assistance is significant at the 1% level and more skilled birth delivery increases the survival and therefore decreases both infant and neonatal deaths. Moreover, as shown in Appendix Table A2, the correlation between any of the independent variables and the inverse Mills ratio is never alarmingly high and much less than the thumb rule of 0.9. Table 2, column (1) finds evidence of negative selection, even in the absence of cohort fixed effects. Columns (2) and (3) control for mother cohorts and see the effect on HAZ and WAZ respectively. In both the results, the coefficient is similar to (1) and highly statistically significant at 1% significance level. With mean inverse Mills ratio in the sample being 0.116, the absolute coefficient value of 1.38 implies an average truncation effect of (0.116*1.38) = 0.16. Thus, the HAZ scores are shifted down due to the selection effect. A child with sample average characteristics who survives with controls for demographic change in fertility of women, has -(exp(0.16)-1)*100 = 17.4% less HAZ scores than a child randomly drawn from the population.

Table 2: Selective Mortality and correction for selection

	Infant Mortality			Neonatal Mortality			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable	**HAZ**	**HAZ**	**WAZ**	**HAZ**	**HAZ**	**WAZ**	**HAZ**
Inverse Mills Ratio	-1.42** (0.52)	-1.38*** (0.45)	-1.18*** (0.37)	-1.99*** (0.33)	-1.21* (0.62)	-0.82* (0.47)	
Female	0.048** (0.021)	0.045* (0.02)	0.028 (0.02)	0.04* (0.021)	0.05** (0.025)	0.039* (0.019)	0.077*** (0.022)
Multiple	0.221 (0.22)	0.21 (0.17)	0.121 (0.127)	0.34** (0.15)	0.091 (0.19)	-0.05 (0.14)	-0.27*** (0.06)
Birth order 2	-0.22*** (0.025)	-0.24*** (0.027)	-0.19*** (0.022)	-0.24*** (0.024)	-0.24*** (0.026)	-0.18*** (0.02)	-0.18*** (0.022)
Birth order 3	-0.37*** (0.028)	-0.40*** (0.032)	-0.29*** (0.028)	-0.44*** (0.032)	-0.40*** (0.034)	-0.29*** (0.03)	-0.33*** (0.026)
Birth order 4	-0.39*** (0.033)	-0.41*** (0.040)	-0.35*** (0.030)	-0.49*** (0.037)	-0.42*** (0.042)	-0.35*** (0.031)	-0.37*** (0.033)
Birth order 5	-0.50*** (0.057)	-0.50*** (0.047)	-0.42*** (0.032)	-0.64*** (0.058)	-0.53*** (0.048)	-0.43*** (0.034)	-0.49*** (0.046)
Birth order 6	-0.62*** (0.072)	-0.60*** (0.053)	-0.49*** (0.042)	-0.74*** (0.068)	-0.62*** (0.054)	-0.52*** (0.04)	-0.65*** (0.052)
Birth order 7	-0.72*** (0.086)	-0.67*** (0.066)	-0.48*** (0.049)	-0.88*** (0.078)	-0.71*** (0.065)	-0.51*** (0.05)	-0.76*** (0.065)
Birth order 8	-0.74*** (0.091)	-0.67*** (0.089)	-0.52*** (0.066)	-0.88*** (0.093)	-0.68*** (0.09)	-0.55*** (0.06)	-0.82*** (0.08)
Birth order 9	-0.86*** (0.145)	-0.73*** (0.095)	-0.62*** (0.08)	-1.04*** (0.153)	-0.79*** (0.92)	-0.68*** (0.077)	-0.95*** (0.10)
Birth order>10	-0.87*** (0.166)	-0.70*** (0.126)	-0.56*** (0.102)	-1.03*** (0.162)	-0.77*** (0.126)	-0.61*** (0.10)	-0.99*** (0.116)
Piped water	0.15*** (0.018)	0.13*** (0.021)	0.13*** (0.014)	0.19*** (0.017)	0.14*** (0.021)	0.14*** (0.014)	0.17*** (0.018)
No toilet	-0.21*** (0.032)	-0.21*** (0.02)	-0.25*** (0.014)	-0.29*** (0.031)	-0.22*** (0.021)	-0.26*** (0.014)	-0.24*** (0.017)
Electricity	0.117*** (0.029)	0.13*** (0.019)	0.13*** (0.015)	0.19*** (0.027)	0.14*** (0.02)	0.14*** (0.016)	0.15*** (0.02)
State FE	YES	YES	YES	YES	YES	YES	NO
Cohort FE	NO	YES	YES	NO	YES	YES	NO
Number of groups	27	233	233	27	223	223	
Observations	68345	67886	76662	67532	67532	76274	68422

Note: Mortality selection effect is captured by the Inverse Mills Ratio. The procedure of deriving the ratio is described in the text. Mother cohort is defined by mothers grouped by rural or urban residence, education, poverty and age. Not all control variables included are listed in the table for brevity. The other control variables included in the specifications are age of child, mother's age at birth, female-specific birth order, birth order, birth month, year of

survey, caste, religion, and state dummies. Standard errors clustered at the group level (cohort/state) are reported in brackets. Column (7) provides OLS estimates without any selection correction.
*** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

In Table 2, (4)-(6), first stage probit estimates the probability of survival till 1 month of age. The resulting inverse Mills ratio is calculated and included in these regressions. With only the state fixed effects, the ratio is highly significant and similar in magnitude as before. With the inclusion of cohort effects, the coefficients are now significant at 10% level. But, the sign of the coefficients is still negative and magnitude similar, albeit a little bit smaller. The negative inverse Mills ratio points towards the direction of a negative selection, implying that the HAZ and WAZ scores of the sample are lower than a population taken at random. This supports our hypothesis that weaker children are surviving due to skilled delivery that pulls down the sample anthropometric scores.

It is interesting to note the coefficients on some other important variables that are historically deemed to be important in determination of malnutrition in children. I run the OLS regression of HAZ score on all the control variables in Table 2, (7) without accounting for selection, where being a female or born in a multiple birth significantly affects HAZ. However, after controlling for selection, being born in a multiple birth is not statistically significant in almost all specifications and the coefficient of female dummy reduces. Birth order, access to clean drinking water, having a toilet and access to electricity all remain significant and similar in magnitude, even after accounting for selective mortality, corroborating the important roles of these variables in explaining malnutrition (Jayachandran and Pande 2017; Spears et al. 2013). With skilled delivery and medical inputs, the complications arising due to multiple births are taken care of, resulting in no differentiation in HAZ between multiple birth children and single birth children conditional on surviving.

According to WHO Standards, both HAZ and WAZ score use the cutoff of -2 to measure moderate and severe under nutrition.[11] Having a low HAZ score is termed as stunted growth while a low WAZ score leads to the child being underweight. Stunted growth refers to a child below 5 years being short for his/her age and is an indicator of chronic malnutrition. Low WAZ can be either due to the child being thin or short for his/her age. This is a combination of chronic and acute malnutrition. Table 3 checks for the presence of mortality selection at various points of the HAZ and WAZ distribution. For stunting and underweight children, I find effects of positive selection whereas for HAZ and WAZ scores greater than -2 negative mortality selection I observe negative selection.

This is consistent with our hypothesis. At the lower end of the distribution, the children are severely under-nourished. This would mean that they are at a higher risk of dying of less nourishment leaving the sample HAZ scores higher than the whole population. On the other hand, with HAZ and WAZ scores greater than -2, if medical intervention is able to save the children and they do not get enough nourishment later, they survive but the presence of these children lowers the HAZ score than the anthropometric score for the population.

[11] A z-score of zero indicates the median of gender and age specific reference population, -1 is 1 standard deviation below and +1 is 1 standard deviation above the reference median population.

Table 3: Selection by HAZ and WAZ Profile

	HAZ			WAZ		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent Variable	**HAZ<-2**	**-2<HAZ<-1**	**HAZ>-1**	**WAZ<-2**	**-2<WAZ<-1**	**WAZ>-1**
Inverse Mills Ratio	1.05*** (0.087)	-0.493*** (0.073)	-0.562*** (0.071)	0.578*** (0.085)	-0.424*** (0.079)	-0.432*** (0.092)
Explanatory Variables	YES	YES	YES	YES	YES	YES
Cohort FE	YES	YES	YES	YES	YES	YES
Interview Dummy	YES	YES	YES	YES	YES	YES
Number of groups	233	233	233	233	233	223
Observations	67886	67886	67886	76662	76662	76274

Note: Mortality selection effect on HAZ and WAZ scores are evaluated at different cutoffs. The other control variables included in the specifications are sex of child, whether born in a multiple birth, mother's age at birth, birth order, female interacted with birth order, birth month, year of survey, whether the household has access to piped water, electricity and toilet, caste, religion, and state dummies. Mother cohort is defined by mothers grouped by rural or urban residence, education, being poor and age. Standard errors clustered at the mother cohort level are reported in brackets.

*** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

6.2 Heterogeneity in Selection

India presents a case of strong gender preferences and birth order differences in health outcomes (Behrman 1988; Jayachandran and Pande 2017). As plotted in Appendix Figure A2, the changes in infant mortality also differ by birth order and gender.[12] Hence, I expect to see differential selection outcomes based on gender, birth order, and an interaction between the two. This hypothesis is tested and results are presented in Table 4. For both infant and neonatal survival

12 Similarly, percentage change in HAZ also differs by gender and birth order, as shown in Appendix Figure A4.

selection, I see that while male HAZ scores have a significant negative effect of selection, no such effect is found for the female group. This would be the case if parents especially put more effort in getting even a weaker male child to survive but do not put such an effort in the case of a girl child. Due to the patriarchy structure in India with explicit son preference, the results seem plausible.

Parents and families face different financial constraints over time and hence may devote less or more resources to latter-born children (Jayachandran and Pande 2017). If the family is more constrained, the outcome for health of latter born child will be negatively affected than their older counterparts. Mothers giving birth at an older age way back in time would face higher probability of a latter-born child dying than mother giving late birth in a new era when technology is able to save her children, without changing family's health behavior much. This is represented in the data in Appendix Figure A3. For any given birth order, the infant mortality rate is lower in 2005 than it was in 1993.

Results in Table 4, columns (2) and (4), corroborate this hypothesis. High birth order children display a negative selection bias implying that technology may have been able to save the inherently weak children lowering the HAZ and WAZ scores. On the other hand, low birth order children have positive selection bias implying those who are surviving are in fact healthier than the population in general.

Table 4: Heterogeneity by gender and birth order

	Infant Mortality			Neonatal Mortality		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent Variable	**HAZ**	**HAZ**	**HAZ**	**HAZ**	**HAZ**	**HAZ**
Female	-0.266 (0.739)			1.64 (1.13)		
N	32586			32417		
Number of Cohorts	231			221		
Male	-2.43*** (0.73)			-1.44** (0.657)		
N	35300			35300		
Number of Cohorts	233			233		
Birth order<=2		1.96*** (0.56)			5.62*** (0.853)	
N		38190			38138	
Number of Cohorts		217			212	
Birth Order>2		-2.94*** (0.438)			-3.03*** (0.673)	
N		29696			29394	
Number of Cohorts		219			209	
Birth order<=2 & Male			2.02*** (0.691)			4.52*** (0.95)
N			19726			19698
Number of Cohorts			205			200
Birth order>2 & Male			-3.29*** (0.663)			-3.29*** (0.88)
N			15574			15417
Number of Cohorts			213			203
Birth order<=2 & Female			2.13** (1.06)			7.61*** (1.40)
N			18464			18440
Number of Cohorts			207			203
Birth order>2 & Female			-2.64*** (0.59)			-2.87*** (1.00)
N			14122			13977
Number of Cohorts			212			202

Note: All the cells represent different regressions on a pooled sample of mothers in multiple surveys in India, according to different criterions. The sample is restricted to females or males in (1) and (4), lower birth order (<=2) or higher birth order (>2) in (2) and (5), and interaction between gender and birth order in (3) and (6). The control variables are sex of child, whether born in multiple birth, birth order, birth month, mother cohort fixed effects, interview-time dummies, mother's age at birth, caste, religion, piped water availability, access to toilet, and whether house has electricity. Standard errors clustered at mother cohort level are reported in brackets.

*** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

The distinction between birth orders is also apparent when females are higher birth order children. Being a female at higher birth order acts as a double disadvantage for female children as in a resource constrained environment, females receive fewer resources than their male counterparts and if the family size is bigger and they are later born, this distinction should be sharper. Table 4, columns (3) and (6), test for the differences in selection by birth order and gender. Both male and female children with high birth order display negative selection, consistent with columns (2) and (4). Male children see a higher negative selection (-3.29) than females (-2.64). It reinforces the fact that higher order male children are more likely to be survived by advanced technology than female children at higher birth orders.

I also expect to see spatial heterogeneity in selection effect. Urban areas have access to better healthcare facilities and therefore, better access to skilled delivery. We should expect to see a strong negative selection in urban areas while not for rural areas. The results are presented in Table 5, columns (2) and (5). As expected, we see evidence of greater negative selection in urban areas for both infant and neonatal mortality. We do not see any statistically significant selection effect of skilled delivery in rural areas.[13]

Based on regions displaying different culture of son preference, heterogeneity by states is also expected. Kerala has distinctively better male-to-female child sex ratio (1.04) than the average in India (1.08), according to 2001 census. Since most of the negative mortality selection is observed in male child, with parity, I should observe no selection effect for this state. Table 5, column (1) and (4) displays the results. There is no statistically significant evidence of mortality selection in Kerala. The magnitude of the coefficient is also small and positive. According to Census 2001,

[13] Since mother cohorts are defined by the place of residence, heterogeneity for urban and rural areas cannot be carried out at the mother cohort level. The analysis for the heterogeneity has been performed with state fixed effects.

while Kerala has a high child sex ratio, there are six states which perform poorly and below the national average – Gujarat, Punjab, Haryana, New Delhi, and Himachal Pradesh. With a strong preference for boys documented for these states, the effect of negative mortality selection should be more pronounced. Table 5 columns (3) and (6) show that the coefficient for mortality selection is negative, high in magnitude, and statistically significant.

Table 5: Spatial Heterogeneity

	Infant Mortality			**Neonatal Mortality**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent Variable	**HAZ**	**HAZ**	**HAZ**	**HAZ**	**HAZ**	**HAZ**
Kerala	0.649 (7.79)			2.86 (4.36)		
N	596			592		
Number of Cohorts	81			79		
Urban Areas		-4.08*** (0.88)			-2.66** (1.15)	
N		22446			22446	
Number of Cohorts		29			29	
Rural Areas		-1.49 (0.98)			-1.45 (1.31)	
N		46260			46260	
Number of Cohorts		29			29	
Worse sex ratio states			-7.2*** (0.82)			-9.63*** (1.20)
N			15133			15060
Number of Cohorts			227			217

Note: All the cells represent different regressions on a pooled sample of mothers in multiple surveys in India, according to different criterions. The sample is restricted to state of Kerala in (1) and (4), urban and rural areas in (2) and (5), and worse sex ratio states – Gujarat, Punjab, Haryana, New Delhi, and Himachal Pradesh in (3) and (6). The control variables are sex of child, whether born in multiple birth, birth order, birth month, mother cohort fixed effects, interview-time dummies, mother's age at birth, caste, religion, piped water availability, access to toilet, and whether house has electricity. Standard errors clustered at mother cohort/state level are reported in brackets.

*** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

6.3 Robustness Checks

It could be argued that the excluded variable of skilled birth delivery may not satisfy the exclusion restriction since it can be taken as a proxy for accessibility to health center which also affects HAZ. Even though this exclusion conditional on other covariates, cannot be tested econometrically, I provide evidence that it does not belong in the second stage HAZ regression. To account for accessibility to health inputs, another variable detailing the number of antenatal visits made by mother to the hospital is included in both the first and second stage regressions. For infant mortality selection, as presented in Table 6 (5), the inverse Mills ratio is negative and statistically significant.

Similarly, it could be argued that a healthy mother would be more probable to experience a lesser incidence of infant mortality and at the same time be better able to raise a healthy child. Since this is an individual mother attribute which varies within a cohort, this is not taken care of by the mother cohort fixed effect, which accounts for mothers being poor, in rural areas, in a certain age group, and education. NFHS collects data on weight of the woman in the women's questionnaire for mothers of children born in the three/five years preceding the survey. To control for health of the mother, I include weight of the mother in both the first and second stage regressions. I find the negative and significant effect of infant mortality selection prevalent in this specification as well, as seen in Table 6 (7).

The results are also robust to changing the outcome variable from HAZ and WAZ to height in centimeters and weight in kilograms. As seen in columns (3) and (4) of Table 6, I still find effects of negative mortality selection and it is significant at the conventional levels. The model is also robust to inclusion of state time trends, which take into account differential levels of development of the states implying that this selection is observable within mothers of similar characteristics within state as well. Since the height and weight data for children in different rounds of NFHS differs by age of child, I restrict the sample to children within 36 months of interview year across all survey years to maintain comparability. The estimates are given in Table 6, column (6). The coefficient on inverse Mills ratio is virtually unchanged from the previous regressions. Lastly, since there are three pooled surveys which have been carried out in different years and at different times of the years, to control more flexibly for survey time; a linear, quadratic and cubic term of the month-year survey time is added to the specification along with age of child and year of birth. This does not significantly affect the results and the coefficient is similar as before.

Table 6: Robustness - Different specifications for mortality selection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent Variable	**HAZ**	**HAZ**	**Height in CM**	**Weight in KG**	**HAZ**	**HAZ**	**HAZ**
Sample	Including state time trends	Including year of birth, square and cubic terms of survey year	Replace the outcome variable, HAZ	Replace the outcome variable, WAZ	Include Antenatal Visits (infant survival)	<3years from interview date	Include weight of the mother (infant survival)
Inverse Mills Ratio	-3.3***	-1.51***	-15.55***	-1.27***	-2.29***	-1.87***	-1.46**
	(0.37)	(0.46)	(1.33)	(0.37)	(0.324)	(0.53)	(0.59)
Explanatory Variables	YES	YES	YES	YES	YES	YES	YES
Cohort FE	YES	YES	YES	YES	YES	YES	YES
Interview Dummy	YES	YES	YES	YES	YES	YES	YES
Number of groups	233	233	233	233	229	226	216
Observation	67886	67886	70480	77125	57360	57025	47214

Note: All the cells represent different regressions on a pooled sample of mothers in multiple surveys in India, according to different criterions. The regression controls for state-time trends in addition to state and year of survey dummies in (1), (2) includes the year of birth along with age at birth and linear, quadratic and cubic controls for survey month-year, (3) replaces the HAZ outcome variable to height measured in centimeters and (4) measures the outcome in weight measured in kilograms instead of WAZ, (5) includes antenatal visits as a control in both first and second stage regressions, (6) includes restricting sample to births within 36 months, and (7) includes mother's weight as a control in both first and second stage regressions. The control variables are sex of child, whether born in multiple birth, birth order, birth month, mother cohort fixed effects, interview-time dummies, mother's age at birth, caste, religion, piped water availability, access to toilet, and house has electricity. Standard errors clustered at mother cohort level are reported in brackets.
*** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table 7 approaches the problem of selection by another empirical method and creates a counterfactual. I can calculate the predicted HAZ score if deceased children were included in the sample. In NFHS, I have fertility history of mother and many women give birth to more than one child. Assuming that mothers have similar abilities in raising all their children and siblings who died will be similar to the ones who survived, after controlling for child characteristics, I develop a predicted HAZ score for the whole sample, including the children who died. This predicted HAZ is better in terms of being able to control for mother unobserved characteristics than random matching based on observable covariates. Since over the years medical technology has improved with better neonatal care, HAZ scores in later survey years will be an underestimate in the sample. I expect that in the predicted sample, the coefficient on 2005 survey year is less positive and statistically significant in increasing HAZ scores.

HAZ has been increasing over the survey years.[14] But some of this increase is explained by changing child demographics and women demographics. Table 7, column (1) provides the results of regression of HAZ on child covariates and survey years. The coefficient on 2005 survey year is positive and statistically significant. Further, I control for mother characteristics in (2). The coefficient on 2005 dummy is still positive, statistically significant, but a little lesser in magnitude. Column (3) now uses the predicted HAZ as the outcome variable. Predicted HAZ is obtained by regressing HAZ on child covariates and family fixed effects and getting a linear prediction. This now includes the predicted HAZ scores for children who have died, creating a counterfactual of survival. The coefficient on the 2005 interview year dummy is significant but almost similar in

[14] Graph in Appendix Figure A1.

magnitude. Similarly, in (4)-(6), the WAZ scores are positive and significant, but not very different in magnitudes.

Table 7: Including deceased children in HAZ and WAZ

	HAZ			WAZ		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent Variable	**HAZ**	**HAZ**	**HAZ-hat**	**WAZ**	**WAZ**	**WAZ-hat**
Child's Age	-0.004***	-0.009***	-0.005***	-0.002***	-0.006***	-0.0035***
	(0.0007)	(0.0015)	(0.006)	(0.0005)	(0.001)	(0.0003)
Female	0.294***	0.30***	-0.0065	0.27***	0.28***	-0.28
	(0.037)	(0.037)	(0.033)	(0.028)	(0.028)	(0.019)
Multiple	-0.149**	-0.216**	-0.342***	-0.186***	-0.25***	-0.502***
	(0.073)	(0.070)	(0.074)	(0.061)	(0.058)	(0.044)
Interview Year						
1998	-0.021	-0.041**	0.180	0.103***	0.064***	0.0107
	(0.018)	(0.018)	(0.016)	(0.013)	(0.013)	(0.009)
2005	0.496***	0.374***	0.340***	0.279***	0.134***	0.203***
	(0.016)	(0.016)	(0.014)	(0.012)	(0.012)	(0.008)
Mother and House Controls	NO	YES	NO	NO	YES	NO
Observations	72116	70271	90818	81018	79141	90818

Note: HAZ and WAZ represent the available anthropometric scores of all the living children. HAZ-hat and WAZ-hat are the predicted HAZ and WAZ scores for the full sample including the deceased children. The predicted values are calculated by a regression of HAZ and WAZ on child characteristics and controlling for mother fixed effects. The other child control variables included in the specifications are birth order, female specific with birth order, and birth month. Mother and house characteristics include mother's age at birth, caste, religion, rural or urban residence, education, being poor and mother's age. Standard errors clustered at the mother level are reported in brackets.
*** Significant at 1% level, ** significant at 5% level, * significant at 10% level.

7 Conclusion

This paper analyzes the paradox of decreasing infant mortality but not a corresponding increase in health anthropometric scores in India over time. Using three rounds of health surveys in India, I find evidence of negative mortality selection. This negative selection is observed when the anthropometric scores are above -2 standard deviations. With improved technology, weaker children are surviving pulling down the sample HAZ and WAZ scores than what the scores would have been otherwise. With a lot of emphasis on reduction in mortality and improving child nutrition as part of Sustainable Development Goals, this result should be taken into consideration. India has recently embarked on establishment of Special Newborn Care Units (SNCUs) in 2013 at district hospitals and sub-district hospitals to provide care for sick newborns. The improved efforts to decrease infant and neonatal deaths by providing one shot interventions like these should be followed by health and nutrition interventions to keep malnourishment away. The maternal and child development literature has focused on interventions like promotion of breastfeeding, micronutrient interventions for children like zinc, iodine, iron, and Vitamin A supplementation as well as general community education strategies like promotion of handwashing to reduce stunting and suboptimal development. If weaker children survive due to successful neonatal interventions, without provision of appropriate care afterwards, it would worsen the case of malnutrition.

I also find evidence of heterogeneous effects based on gender and birth order of the child. Given the strong son preference in India this is not surprising. Male children are more likely to experience negative selection in mean HAZ and WAZ scores with no such effects for the female child. With a patrilineal structure in India, all efforts to save a male child are expected, which makes the probability of finding a weaker male child in the distribution higher; pulling down the overall anthropometric scores. In terms of birth order, I observe negative selection effect in higher birth order children, for both males and females. Spatial heterogeneity is also observed with states historically better off in child sex ratio composition like Kerala displaying no evidence of selection while states with skewed sex ratios like Punjab, Haryana, Gujarat, New Delhi and Himachal Pradesh, display a strong negative selection indicating the presence of son preference. These patterns of mortality selection may be depressing the child anthropometry scores, hiding the fact that India maybe doing better in terms of number of malnourished children in the population overtime than the raw data suggests.

Acknowledgements

I am grateful to Anil Deolalikar, Mindy Marks, Joseph Cummins, and Aman Ullah for their comments. This paper has benefitted from presentation at the South Asian University Seminar series, Population Association of America (PAA) Annual Meeting in 2016, and CSWEP workshop in WEAI meetings. This paper was developed in University of California, Riverside and I am grateful for the funding provided by the graduate division. All remaining errors are my own.

References

Akachi, Yoko and David Canning. 2010. "Health Trends in Sub-Saharan Africa: Conflicting Evidence from Infant Mortality Rates and Adult Heights." *Economics and Human Biology* 8(2):273–88.

Alderman, Harold, Michael Lokshin, and Sergiy Radyakin. 2011. "Tall Claims: Mortality Selection and the Height of Children in India." *Economics and Human Biology* 9(4):393–406.

Behrman, J. R. 1988. "Nutrition, Health, Birth Order and Seasonality: Intrahousehold Allocation among Children in Rural India." *Journal of Development Economics* 28(1):43–62.

Behrman, Jere R. and Paul Taubman. 1986. "Birth Order , Schooling , and Earnings." *Journal of Labor Economics* 4(3):S121–45.

Belli, Pedro. 1971. "The Economic Implications of Malnutrition: The Dismal Science Revisited." *Economic Development and Cultural Change* 20(1):1.

Boerma, J. Ties, A. Elisabeth Sommerfelt, and George T. Bicego. 1992. "Child Anthropometry in Cross-Sectional Surveys in Developing Countries: An Assessment of the Survivor Bias." *American Journal of Epidemiology* 135(4):438–49.

Borjas, George J. and Glenn T. Sueyoshi. 1994. "A Two-Stage Estimator for Probit Models with Structural Group Effects." *Journal of Econometrics* 64(1–2):165–82.

Bozzoli, Carlos, Angus Deaton, and Climent Quintana-domeque. 2009. "Adult Height and Childhood Disease." *Demography* 46(4):647–69.

Bushway, Shawn, Brian D. Johnson, and Lee Ann Slocum. 2007. "Is the Magic Still There? The Use of the Heckman Two-Step Correction for Selection Bias in Criminology." *Journal of Quantitative Criminology* 23(2):151–78.

Caulfield, Laura E., Mercedes de Onis, Monika Blössner, and Robert E. Black. 2004. "Undernutrition as an Underlying Cause of Child Deaths Associated with Diarrhea, Pneumonia, Malaria, and Measles." *The American Journal of Clinical Nutrition* 80(1):193–98.

Dancer, Diane, a N. U. Rammohan, and Murray D. Smith. 2008. "Infant Mortality and Child Nutrition." 1035(July):1015–35.

Deaton, Angus. 1985. "Deaton_Panel_Data_from_Time_Series_of_Cross_Sections.Pdf." *Journal of Econometrics* 30:109–26.

Deaton, Angus. 2007. "Height, Health, and Development." *Proceedings of the National Academy of Sciences of the United States of America* 104(33):13232–37.

Horton, Sue. 1988. "Birth Order and Child Nutritional Status: Evidence from the Philippines." *Economic Development and Cultural Change* 36(2):341–54.

Jayachandran, Seema and Rohini Pande. 2017. "Why Are Indian Children so Short? The Role of Birth Order and Son Preference." *American Economic Review* 107(9):2600–2629.

Little, Roderick J. A. 1985. "A Note about Models for Selectivity Bias." *Econometrica* 53(6):1469–74.

Moradi, Alexander. 2010. "Selective Mortality or Growth after Childhood? What Really Is Key to Understand the Puzzlingly Tall Adult Heights in Sub-Saharan Africa." *Working Paper* 1–23.

Panagariya, Arvind. 2013. "Does India Really Suffer from Worse Child Malnutrition Than Sub-Saharan Africa?" *Economic and Political Weekly* 48(18):98–111.

Pelletier, D. L., E. A. Frongillo, D. G. Schroeder, and J. P. Habicht. 1995. "The Effects of Malnutrition on Child Mortality in Developing Countries." *Bulletin of the World Health Organization* 73(4):443–48.

Pinchinat, Sybil, Catherine Enel, Gilles Pison, Géraldine Duthé, Emmanuel Lagarde, François Simondon, and Kirsten B. Simondon. 2004. "No Improvement in Weight-for-Age of Young Children in Southern Senegal, 1969-1992, despite a Drastic Reduction in Mortality. Evidence from a Growth Monitoring Programme." *International Journal of Epidemiology* 33(6):1202–8.

Pitt, Mark M. 1997. "Estimating the Determinants of Child Health When Fertility and Mortality Are Selecetive." *The Journal of Human Resources* 32(1 (Winter)):129–58.

Rice, Amy L., Lisa Sacco, Adnan Hyder, and Robert E. Black. 2000. "Malnutrition as an Underlying Cause of Childhood Deaths Associated with Infectious Diseases in Developing Countries." *Bulletin of the World Health Organization* 78(10):1207–21.

Rodríguez, Jhon James Mora and Juan Muro. 2014. "Consistent Estimation in Pseudo Panels in the Presence of Selection Bias." *Economics: The Open-Access, Open-Assessment E-Journal* 8(2015):1–26.

Savage, Tim, José G. B. Derraik, Harriet L. Miles, Fran Mouat, Paul L. Hofman, and Wayne S. Cutfield. 2013. “Increasing Maternal Age Is Associated with Taller Stature and Reduced Abdominal Fat in Their Children.” *PloS One* 8(3):e58869.

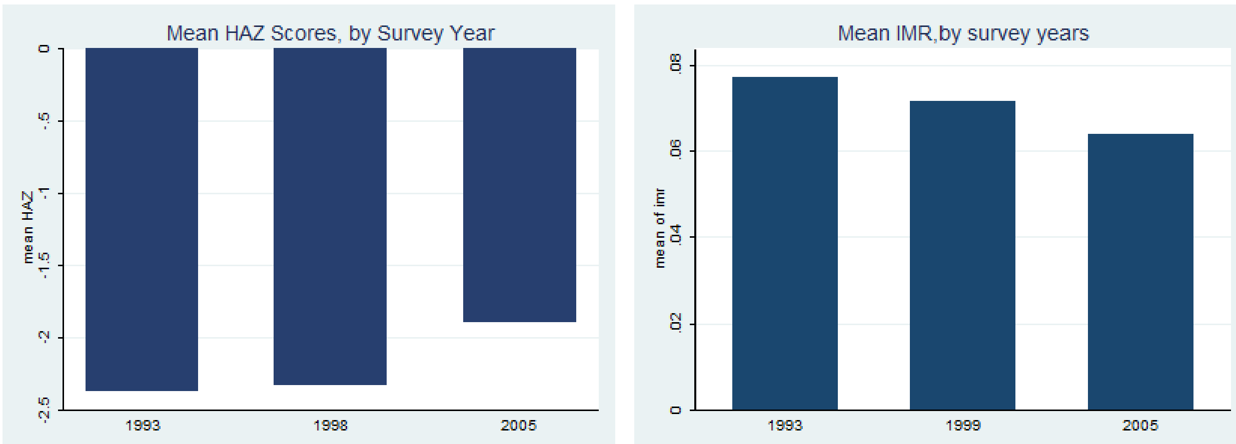
Spears, Dean, Arabinda Ghosh, and Oliver Cumming. 2013. “Open Defecation and Childhood Stunting in India: An Ecological Analysis of New Data from 112 Districts.” *PLoS ONE* 8(9):1–9.

Vella, Francis. 1998. “Estimating Models with Sample Selection Bias: A Survey.” *The Journal of Human Resources* 33(1):127–69.

World Health Organization. 2011. *World Health Statistics Report, 2011*.

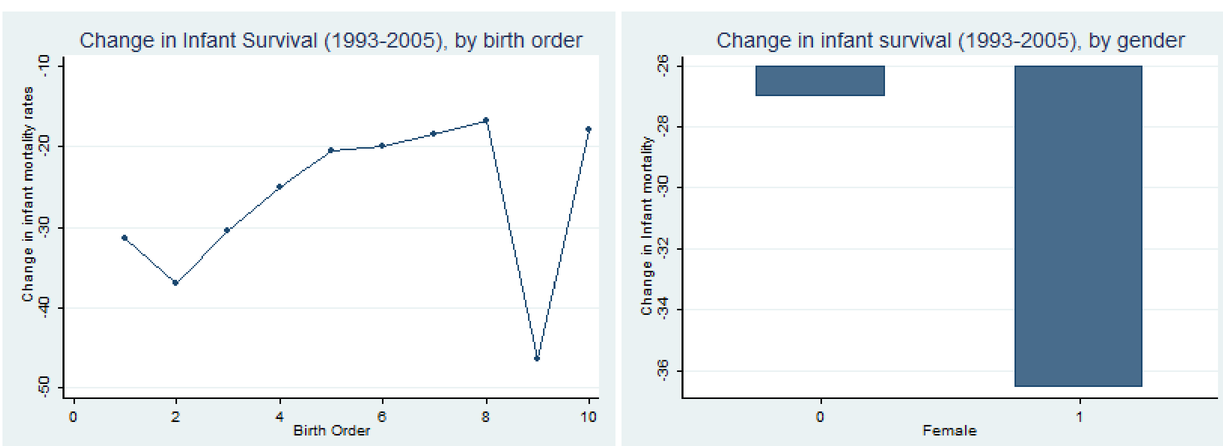
Appendix

Figure A1: Mean HAZ Scores and IMR, by Survey Years



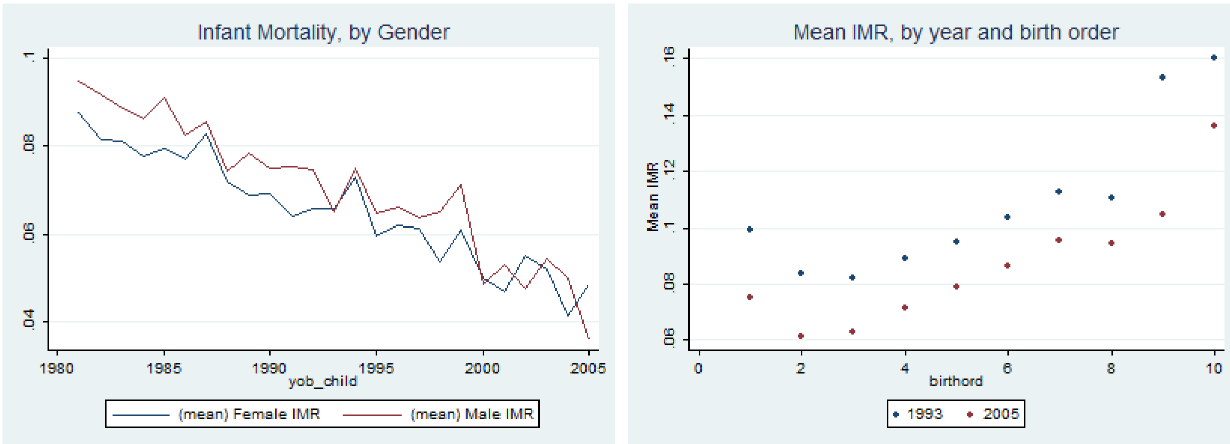
Note: These bar graphs plot the mean HAZ Scores and mean infant deaths by the three survey years – 1993, 1998 and 2005 in our sample.

Figure A2: Infant Mortality Change, by Child Characteristics



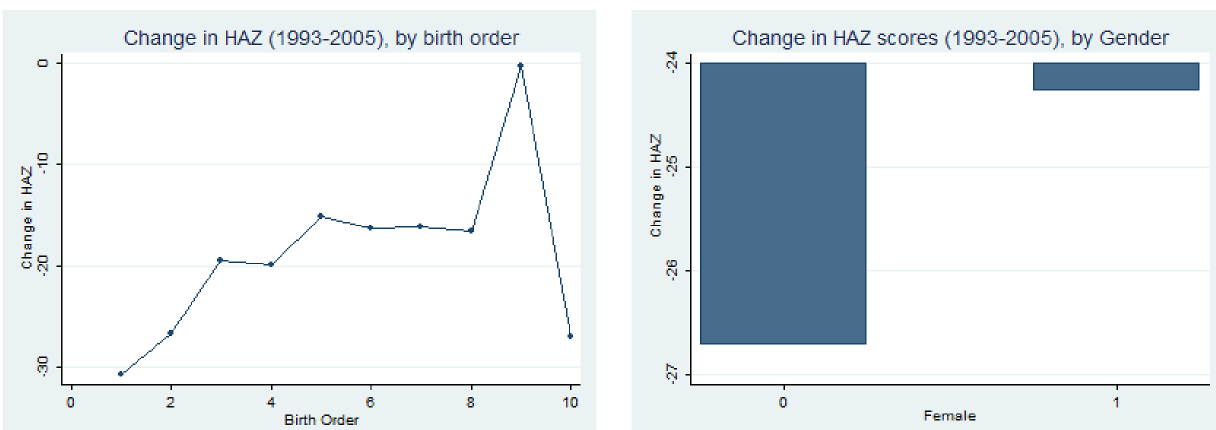
Note: The first graph shows how change in infant mortality varies by birth order. The changes in infant mortality are calculated by changes in mean infant deaths by birth order over 1993 to 2005, divided by mean infant mortality over that cell in 1993 and multiplied by 100. The second graph plots a bar graph of infant mortality changes for males and females where male is denoted by female=0.

Figure A3: Infant Mortality Trends, by Gender and Birth Order



Note: This graph shows the decline in infant deaths overtime (1980-2005) by gender. Panel 2 plots the mean infant deaths by birth order for the two survey years, 1993 and 2005.

Figure A4: Changes in HAZ (1993-2005), by Gender and Birth Order



Note: The first graph shows how change in HAZ varies by birth order, where change is defined as the difference between mean HAZ by birth order between 1993 and 2005, divided by HAZ in 1993 and multiplied by 100. The second graph plots a bar graph of HAZ changes for males and females where male is denoted by female=0.

Table A1: First-Stage Probit Estimates (Marginal Effects)

	(1)	(2)	(3)
Dependent Variable	**Survival till 1 year**	**Survival till 1 month**	**Survival till 1 year**
Delivery	0.0076*** (0.002)	0.0038** (0.0019)	0.0104*** (0.002)
Female	0.010*** (0.003)	0.007*** (0.002)	0.010*** (0.003)
Multiple	-0.24*** (0.012)	-0.195*** (0.013)	-0.24*** (0.012)
Piped water	0.007 (0.004)	0.004 (0.023)	0.009** (0.004)
No toilet	-0.006*** (0.002)	-0.006*** (0.016)	-0.008*** (0.003)
Electricity	0.007*** (0.002)	0.007*** (0.002)	0.01*** (0.002)
Mother's Age at Birth	0.0014** (0.0005)	0.0011*** (0.0002)	0.0022** (0.0002)
State FE	YES	YES	YES
Cohort FE	YES	NO	NO
Observations	93999	94614	94614

Note: These are the first stage probit estimates of deriving the inverse Mills ratio. Other included variables are birth dummy, female interacted with birth dummy, survey year, month of birth, caste, religion, and state dummies. Mother cohorts are also included in (1). Standard errors are clustered at the state level.

Table A2: Correlation between inverse Mills Ratio and independent variables

Inverse Mills Ratio	**(1) 1-year survival**	**(2) 1-year survival**	**(3) 1-month survival**
Female	-0.0441	-0.0491	-0.1065
Multiple	0.5971	0.6511	0.7195
Interview Year	-0.1885	-0.2041	-0.1262
Month of birth	0.0212	0.0235	0.000
Caste	-0.0968	-0.1076	-0.0714
Religion	-0.1620	-0.1804	-0.1774
Mother's Age at Birth	-0.1681	-0.1902	-0.2243
Piped water	-0.2316	-0.2547	-0.1954
No toilet	0.3743	0.4117	0.3459
Electricity	-0.3574	-0.3925	-0.3240
State FE	YES	YES	YES
Cohort FE	YES	NO	NO

Note: These are the pairwise correlations of the inverse Mills ratio with other independent variables in the substantive equation.