

# ONLINE APPENDIX

## Why Are Indian Children So Short? The Role of Birth Order and Son Preference

Seema Jayachandran  
Northwestern University

Rohini Pande  
Harvard University

April 27, 2017

Online Appendix Table 1: Summary statistics for within-India analysis

	India	Kerala & Northeast	Below-median child sex ratio	Muslims
<b>NFHS 1-3</b>				
Child's age (months)	24.74 [15.33]	25.42 [15.35]	25.02 [15.37]	25.07 [15.50]
Child's birth order	2.74 [1.87]	2.76 [1.93]	2.69 [1.84]	3.24 [2.19]
Child's HFA z-score	-1.65 [1.91]	-1.39 [1.93]	-1.56 [1.90]	-1.73 [1.94]
Child's WFA z-score	-1.56 [1.41]	-1.20 [1.41]	-1.49 [1.41]	-1.60 [1.39]
Mother's age at birth (years)	24.70 [5.32]	25.72 [5.72]	24.77 [5.52]	24.97 [5.68]
Mother's total children born	2.97 [1.89]	2.99 [1.96]	2.90 [1.87]	3.50 [2.20]
Mother is literate	0.51 [0.50]	0.68 [0.46]	0.54 [0.49]	0.43 [0.49]
Average pooled inputs	0.40 [0.33]	0.39 [0.33]	0.41 [0.33]	0.37 [0.33]
Child sex ratio (boys/girls)	1.08 [0.05]	1.04 [0.01]	1.04 [0.01]	1.08 [0.04]
Main sample of children <60 months (N)	95,220	17,899	42,424	14,053
<b>IHDS</b>				
Child's age (months)	33.43 [16.00]	34.52 [15.22]	33.31 [16.10]	33.59 [16.43]
Child's birth order	2.77 [1.71]	2.13 [1.03]	2.44 [1.37]	3.81 [2.41]
Child's HFA z-score	-1.55 [2.30]	-1.41 [2.51]	-1.49 [2.37]	-1.23 [2.44]
Child's WFA z-score	-1.16 [1.58]	-0.69 [1.89]	-1.10 [1.66]	-0.93 [1.64]
Mother's age at birth (years)	25.93 [5.31]	27.17 [4.77]	25.77 [5.44]	27.23 [6.74]
Mother's total children born	3.11 [1.71]	2.42 [0.99]	2.74 [1.37]	4.21 [2.36]
Mother is literate	0.64 [0.48]	0.91 [0.29]	0.68 [0.47]	0.57 [0.50]
Child sex ratio (boys/girls)	1.09 [0.05]	1.04 [0.01]	1.05 [0.01]	1.08 [0.04]
Main sample of children <60 months (N)	3,615	189	1,266	351

Notes: The means of the specified variables are calculated separately for each subsample. Standard deviations appear in brackets. The following variables are summarized at the mother level: total children born, mother is literate, and mother completed her fertility. Variables summarized at the child level include: all child variables, mother's age at birth, average pooled inputs, and child sex ratio. Except for the first variable, all IHDS variables are summarized for the completed fertility sample.

Online Appendix Table 2: Accounting exercise

	HFA z-score (1)	Accounting Exercise 1	Accounting Exercise 2
		HFA z-score (2)	HFA z-score (3)
India	-0.162 [0.017]		
Gradient proxy		0.400 [0.070]	0.688 [0.132]
Mean of gradient proxy			
India		-0.331	-0.234
Africa		-0.066	-0.037
Kerala & Northeast		-0.229	-0.155
Rest of India		-0.358	-0.257
Log GDP per capita	Yes	Yes	Yes
Sample	Full	Africa	Africa
Observations	168,108	126,039	126,039

Notes: Standard errors are clustered by PSU and appear in brackets. In column (1), the sample consists of children aged 5 or less in NFHS-3 and 27 African countries. In columns (2)-(3), the sample is restricted to African countries. In column (2), *Gradient proxy* is defined as the weighted average of the height gap between second borns and first borns and the height gap between third- and higher-borns and first borns. In column (3), *Gradient proxy* is defined as the regression coefficient that is obtained by regressing, separately by country-wave, HFA z-scores on a linear birth order variable that is top-coded at 3+. All columns include child age dummies and log GDP per capita in the birth year. See Online Data Appendix for further details.

	Accounting Exercise 1	Accounting Exercise 2
Calculation 1: Share explained by birth order gradient		
– India-Africa gap in birth order gradient	-0.265	-0.197
– India-Africa gap in birth order gradient $\times$ Gradient coeff	-0.106	-0.136
– Explained share of India-Africa level gap in height (shown in column 1)	65 %	84 %
Calculation 2: Share explained by birth order gradient rooted in eldest son preference		
– Rest of India - Kerala & NE gap in birth order gradient	-0.129	-0.102
– Rest of India - Kerala & NE gap in birth order gradient $\times$ Gradient coeff	-0.052	-0.070
– Explained share of India-Africa level gap in height (shown in column 1)	32 %	43 %

Online Appendix Table 3: Birth order gradient in the India height gap: Robustness checks

	HFA z-score (1)	HFA z-score (2)	HFA z-score (3)	HFA z-score (4)	HFA z-score (5)	HFA z-score (6)	HFA z-score (7)	Height in cm (8)	Height in cm (9)
India × 2nd child	-0.151 [0.027]	-0.144 [0.027]	-0.088 [0.054]	-0.154 [0.027]	-0.294 [0.059]	-0.140 [0.026]	-0.260 [0.049]	-0.521 [0.087]	-1.189 [0.156]
India × 3rd+ child	-0.201 [0.033]	-0.219 [0.033]	-0.223 [0.092]	-0.215 [0.034]	-0.541 [0.109]	-0.218 [0.032]	-0.429 [0.087]	-0.670 [0.106]	-1.823 [0.275]
2nd child	-0.014 [0.017]	-0.003 [0.017]	-0.039 [0.028]	-0.033 [0.017]	-0.311 [0.034]	-0.041 [0.016]	-0.199 [0.027]	-0.105 [0.053]	-0.752 [0.088]
3rd+ child	-0.123 [0.020]	-0.081 [0.020]	-0.120 [0.045]	-0.118 [0.020]	-0.629 [0.061]	-0.124 [0.019]	-0.391 [0.044]	-0.502 [0.063]	-1.491 [0.141]
Africa mean of outcome	-1.351	-1.351	-1.351	-1.316	-1.316	-1.351	-1.351	81.476	81.476
Age & other controls	Yes	Yes	No	Yes	No	Yes	No	Yes	No
Mother FEs	No	No	Yes	No	Yes	No	Yes	No	Yes
Sample	All	All	All	Birth order $\leq 4$	Birth order $\leq 4$	All	All	All	All
Additional controls	Desired fertility	Birth spacing	Birth spacing	No	No	Birth order among living siblings	Birth order among living siblings	No	No
Observations	167,737	167,737	83,228	121,221	53,133	167,737	83,228	167,737	83,228

Notes: Standard errors are clustered by mother and appear in brackets. All regressions include child age dummies, and child age dummies interacted with *India*. Columns (1), (2), (4), (6), and (8) additionally include maternal age, mother's literacy, PSU fixed effects, and maternal age and mother's literacy interacted with *India*. Column (1) includes dummies for mother's desired fertility, top-coded at 6 children, and dummies for mother's desired fertility interacted with *India*. This regression does not have a parallel mother fixed effects specification, since desired fertility is absorbed by mother fixed effects. Columns (2)-(3) include birth spacing in months and birth spacing interacted with *India*. Birth spacing for firstborns is imputed with the sample average for the rest of the sample. In columns (4)-(5), the sample is restricted to children of birth order 4 or less. In columns (6)-(7), birth order is redefined as the birth order among currently living siblings. See Online Data Appendix for further details.

Online Appendix Table 4: Birth order gradients compared to other regions

<i>Comparison sample:</i>	<i>Countries with similar GDP to India</i>			<i>Europe, Central &amp; West Asia</i>			<i>Bangladesh &amp; Pakistan</i>		
	HFA z-score (1)	HFA z-score (2)	HFA z-score (3)	HFA z-score (4)	HFA z-score (5)	HFA z-score (6)	HFA z-score (7)	HFA z-score (8)	HFA z-score (9)
India	-0.034 [0.018]			-0.785 [0.020]			0.221 [0.020]		
India × 2nd child	-0.095 [0.025]	-0.121 [0.026]	-0.259 [0.048]	-0.064 [0.028]	-0.085 [0.030]	-0.200 [0.052]	-0.111 [0.028]	-0.057 [0.030]	-0.182 [0.062]
India × 3rd+ child	-0.284 [0.024]	-0.191 [0.032]	-0.514 [0.085]	-0.298 [0.028]	-0.153 [0.037]	-0.459 [0.097]	-0.192 [0.028]	-0.059 [0.038]	-0.297 [0.114]
2nd child	-0.025 [0.015]	-0.052 [0.016]	-0.152 [0.027]	-0.058 [0.020]	-0.088 [0.021]	-0.211 [0.034]	-0.011 [0.019]	-0.116 [0.021]	-0.229 [0.048]
3rd+ child	-0.159 [0.013]	-0.155 [0.019]	-0.251 [0.044]	-0.147 [0.019]	-0.193 [0.026]	-0.306 [0.064]	-0.251 [0.019]	-0.287 [0.027]	-0.468 [0.088]
Comparison group mean of outcome	-1.303	-1.303	-1.303	-0.560	-0.560	-0.560	-1.610	-1.610	-1.610
Age & other controls	No	Yes	No	No	Yes	No	No	Yes	No
Mother FEs	No	No	Yes	No	No	Yes	No	No	Yes
Observations	166,709	166,281	81,742	83,998	83,461	39,463	75,535	75,435	30,357

Notes: Standard errors are clustered by mother and appear in brackets. Child age dummies are included in all regressions. Columns (2), (5) and (8) include mother's literacy, maternal age, PSU fixed effects, and child age dummies, mother's literacy, and maternal age interacted with *India*. In columns (2), (5) and (8), the main effect *India* is absorbed by PSU fixed effects. In column (3), (6) and (9), the main effect *India* is absorbed by mother fixed effects. In columns (1)-(3), the omitted category includes 23 DHS's (2004-2010) of countries with height data that had a log GDP per capita within in a 50 percent upper and lower bound of India's 2005-6 log GDP per capita in its survey year. In columns (4)-(6), the omitted category includes 16 DHS's (1995-2012) of European, Central and West Asian countries with height data available. In columns (7)-(9), the omitted category includes Bangladesh 2004, Bangladesh 2007, Bangladesh 2011 and Pakistan 2012-13. See Online Data Appendix for further details.

Online Appendix Table 5: Cultural norms and child outcomes: Additional within-India evidence

<i>Low son preference proxy:</i>	<i>Kerala &amp; Northeast</i>		<i>Below-median child sex ratio</i>		<i>Muslims</i>	
	Average pooled inputs (1)	WFA z-score (2)	Average pooled inputs (3)	WFA z-score (4)	Average pooled inputs (5)	WFA z-score (6)
Low son pref proxy × 2nd child	0.012 [0.006]	0.544 [0.384]	0.010 [0.005]	0.145 [0.151]	0.019 [0.007]	0.346 [0.268]
Low son pref proxy × 3rd+ child	-0.016 [0.007]	0.663 [0.688]	-0.011 [0.006]	0.233 [0.255]	0.061 [0.009]	0.055 [0.361]
2nd child	-0.089 [0.003]	-0.381 [0.074]	-0.092 [0.003]	-0.421 [0.093]	-0.089 [0.003]	-0.385 [0.080]
3rd+ child	-0.170 [0.003]	-0.404 [0.120]	-0.168 [0.004]	-0.457 [0.142]	-0.177 [0.003]	-0.364 [0.131]
Low son pref group mean of outcome	0.392	-0.690	0.411	-1.098	0.371	-0.932
High son pref group mean of outcome	0.399	-1.185	0.386	-1.192	0.402	-1.211
Sample	NFHS 1-3	IHDS 1	NFHS 1-3	IHDS 1	NFHS 1-3	IHDS 1
Age & other controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	95,089	3,615	95,089	3,615	82,054	3,405

Notes: Standard errors are clustered by mother and appear in brackets. In columns (1), (3), and (5), the sample uses NFHS 1-3. NFHS-1 only has data for children aged 4 years and younger, and NFHS-2 only has data for children aged 3 years and younger. In columns (2), (4), and (6), the sample uses IHDS 1 and consists of children aged 1-59 months in IHDS 1 whose mothers (likely) had completed their fertility, i.e. their mothers did not give birth between wave 1 and wave 2. NFHS-1 and NFHS-2 do not have data on whether the child is taking iron pills, whether the child had a postnatal check within 2 months of birth, and whether the child received one of the four doses of the polio vaccine. *Average pooled inputs* is thus restricted to 4 prenatal and 1 postnatal inputs. All columns include child age dummies, maternal age, mother's literacy, and child age dummies, maternal age, and mother's literacy interacted with *Low son pref proxy*. Columns (1), (3), and (5), include survey and PSU fixed effects, and survey and PSU fixed effects interacted with *Low son pref proxy*. Columns (2), (4), and (6) include total fertility dummies, top-coded at 6 children, and total fertility dummies interacted with *Low son pref proxy*. In columns (1), (3), and (5), the main effect *Low son pref proxy* is absorbed by PSU fixed effects. In columns (2), (4), and (6), the main effect *Low son pref proxy* is included but not shown. Matrilineal states include Arunachal Pradesh, Assam, Kerala, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura. *Child sex ratio* is defined as the number of boys aged 0-6 years over the number of girls aged 0-6 years in the respondents state-by-rural cell. In columns (5)-(6), the sample is restricted to Muslims and Hindus. See Online Data Appendix for further details.

Online Appendix Table 6: Birth order gradient and gender gap in Kerala & NE compared to Africa

	HFA z-score (1)	WFA z-score (2)	HFA z-score (3)	WFA z-score (4)
Kerala & Northeast × 2nd child	-0.139 [0.051]	-0.155 [0.037]		
Kerala & Northeast × 3rd+ child	-0.201 [0.059]	-0.204 [0.043]		
Kerala & Northeast × Girl			-0.079 [0.038]	-0.098 [0.027]
Africa mean of outcome	-1.351	-0.877	-1.351	-0.877
Age & other controls	Yes	Yes	Yes	Yes
Observations	134,584	134,584	134,584	134,584

Notes: Standard errors are clustered by mother and appear in brackets. All columns include child age dummies, mother's literacy, maternal age, PSU fixed effects, and child age dummies, mother's literacy, and maternal age interacted with *Kerala & NE*. The main effect *Kerala & NE* is absorbed by PSU fixed effects. In columns (1)-(2), the main effects *2nd child* and *3rd child+* are included but not shown. In columns (3)-(4), the main effect *Girl* is included but not shown. See Online Data Appendix for further details.

Online Appendix Table 7: Birth order gradient in the India height gap using NFHS-1

	HFA z-score (1)	HFA z-score (2)	HFA z-score (3)	HFA z-score (4)	WFA z-score (5)
India	-0.306 [0.013]	-0.098 [0.025]			
India × 2nd child		-0.162 [0.035]	-0.167 [0.037]	-0.251 [0.074]	-0.184 [0.029]
India × 3rd+ child		-0.357 [0.031]	-0.280 [0.042]	-0.427 [0.130]	-0.258 [0.033]
2nd child		0.085 [0.020]	0.013 [0.021]	-0.088 [0.040]	0.021 [0.018]
3rd+ child		0.030 [0.016]	-0.135 [0.023]	-0.193 [0.066]	-0.093 [0.020]
Africa mean of outcome	-1.476	-1.476	-1.476	-1.476	-1.033
Sample	Children ≤ 48 mths				
Age & other controls	No	No	Yes	No	Yes
Mother FEs	No	No	No	Yes	No
Observations	92,726	92,726	92,644	32,797	92,644

Notes: Standard errors are clustered by mother and appear in brackets. Sample for India uses NFHS-1 rather than NFHS-3; NFHS-1 only has data for children age 4 years and younger. The comparison sample includes African DHS's between 1991 and 1997. Four of these surveys (Ghana 1993, Madagascar 1997, Mali 1995-96 and Zimbabwe 1994) only have data for children age 3 years and younger. The other surveys have data for children aged 4 years and younger. Child age dummies are included in all regressions. Columns (3) and (5) include mother's literacy, maternal age, PSU fixed effects, and child age dummies, mother's literacy, and maternal age interacted with *India*. In columns (3) and (5), the main effect *India* is absorbed by PSU fixed effects. In column (4), the main effect *India* is absorbed by mother fixed effects. See Online Data Appendix for further details.

Online Appendix Table 8: The role of gender using NFHS-1

	HFA z-score (1)	HFA z-score (2)	HFA z-score (3)	WFA z-score (4)	WFA z-score (5)	WFA z-score (6)
India × Girl	-0.140 [0.025]			-0.047 [0.020]		
India × 2nd child		-0.132 [0.055]	-0.106 [0.061]		-0.133 [0.043]	-0.119 [0.048]
India × 3rd+ child		-0.257 [0.061]	-0.222 [0.074]		-0.228 [0.048]	-0.208 [0.059]
India × 2nd child × Girl		-0.071 [0.078]	-0.103 [0.086]		-0.111 [0.062]	-0.149 [0.068]
India × 3rd+ child × Girl		-0.053 [0.084]	-0.100 [0.102]		-0.069 [0.067]	-0.127 [0.082]
India × No elder brother			0.046 [0.053]			0.026 [0.042]
India × Girl × No elder brother			-0.061 [0.074]			-0.073 [0.059]
Africa mean of outcome	-1.476	-1.476	-1.476	-1.033	-1.033	-1.033
Age & other controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	92,644	91,984	91,984	92,644	91,984	91,984

Notes: Standard errors are clustered by mother and appear in brackets. All columns include child age dummies, mother's literacy, mother's age, and PSU fixed effects. In columns (2)-(3) and (5)-(6), child age dummies, maternal age, and mother's literacy are interacted with *Girl*, *India*, and *India* × *Girl* and PSU fixed effects are interacted with *Girl*. The main effect *Girl* is included in all regressions but not shown. In columns (2) and (5), the main effects *2nd child*, *3rd+ child*, *Girl*, *2nd child* × *Girl* and *3rd+ child* × *Girl* are included but not shown. In columns (3) and (6), all other main effects (*2nd child*, *3rd+ child*, *Girl*, *2nd child* × *Girl*, *3rd+ child* × *Girl*, *No elder brother*, *Girl* × *No elder brother*) are included but not shown. In all columns, the main effect *India* absorbed by PSU fixed effects. In columns (2)-(3) and (5)-(6), the main effect *India* × *Girl* absorbed by PSU fixed effects and their interactions with *Girl*. See Online Data Appendix for further details.

Online Appendix Table 9: Fertility continuation in NFHS-1 and NFHS-3

	<i>NFHS-1</i>		<i>NFHS-3</i>	
	Beyond desired fertility (1)	Wants more children (2)	Beyond desired fertility (3)	Wants more children (4)
India × No son within desired fertility	0.068 [0.008]		0.067 [0.007]	
No son within desired fertility	-0.114 [0.004]		-0.047 [0.003]	
India × At or beyond desired fertility × No son yet		0.300 [0.027]		0.189 [0.017]
India × At or beyond desired fertility		-0.133 [0.008]		-0.130 [0.007]
India × No son yet		0.002 [0.008]		0.022 [0.009]
At or beyond desired fertility × No son yet		0.049 [0.021]		0.025 [0.011]
Africa mean of outcome	0.179	0.703	0.172	0.667
India mean of outcome	0.284	0.484	0.326	0.339
PSU FEs & literacy controls	Yes	Yes	Yes	Yes
Observations	70,160	65,131	115,768	114,362

Notes: Standard errors are clustered by mother and appear in brackets. All columns include mother's literacy, PSU fixed effects, and mother's literacy interacted with *India*. The main effect *India* is absorbed by PSU fixed effects. In columns (2) and (4), the main effects *At or beyond desired fertility* and *No son yet* are included but not shown. See Online Data Appendix for further details.

Online Appendix Table 10: Alternative explanations for the Indian birth order gradient

	HFA z-score (1)	Diarrhea in last 2 weeks (2)	HFA z-score (3)	HFA z-score (4)	HFA z-score (5)	HFA z-score (6)
India × 2nd child	-0.156 [0.031]	-0.001 [0.005]	-0.165 [0.028]	-0.142 [0.030]	-0.154 [0.029]	-0.153 [0.046]
India × 3rd+ child	-0.185 [0.038]	0.012 [0.006]	-0.217 [0.035]	-0.215 [0.036]	-0.212 [0.035]	-0.211 [0.054]
2nd child	0.244 [0.357]	-0.001 [0.003]	-0.024 [0.019]	-0.036 [0.025]	-0.061 [0.027]	-0.003 [0.041]
3rd+ child	-0.404 [0.426]	0.001 [0.004]	-0.138 [0.023]	-0.133 [0.028]	-0.199 [0.033]	-0.100 [0.047]
2nd child × Mother's height	-0.161 [0.226]					
3rd+ child × Mother's height	0.183 [0.269]					
2nd child × Open defecation			0.035 [0.030]			
3rd+ child × Open defecation			0.055 [0.035]			
2nd child × Percent non-resident among children				0.251 [0.178]		
3rd+ child × Percent non-resident among children				0.176 [0.204]		
2nd child × Nr. of adult females in hh					0.022 [0.012]	
3rd+ child × Nr. of adult females in hh					0.044 [0.015]	
2nd child × Land scarcity						-0.003 [0.015]
3rd+ child × Land scarcity						-0.007 [0.017]
Africa mean of outcome	-1.351	0.156	-1.351	-1.351	-1.351	-1.351
Age & other controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	166,292	167,737	162,503	167,737	167,737	167,737

Notes: Standard errors are clustered by mother and appear in brackets. Control variables included are child age dummies, mother's literacy, maternal age, PSU fixed effects, and child age dummies, mother's literacy, and maternal age interacted with *India*. The main effect *India* is absorbed by PSU fixed effects. Column (1) additionally controls for child's age, maternal age, mother's literacy, and PSU fixed effects interacted with *Mother's height*. Column (3) additionally controls for child's age, maternal age, mother's literacy, and PSU fixed effects interacted with *Open defecation*. *Open defecation* is a dummy variable that equals 1 if the mother reports that the household has no toilet facility. Column (4) also controls for child's age, maternal age, and mother's literacy interacted with *Percent non-resident among children*. Column (5) controls for child's age, maternal age, mother's literacy, and PSU fixed effects interacted with *Nr. of adult females in hh*. *Nr. of adult females in hh* is defined as the total number of interviewed females aged 18 or above in the household. Column (6) controls for child's age, maternal age, and mother's literacy interacted with *Land scarcity*. *Land scarcity* is defined as the log of the respondent's country's total population in 1961 divided by its land area in square km in 1961. See Online Data Appendix for further details.

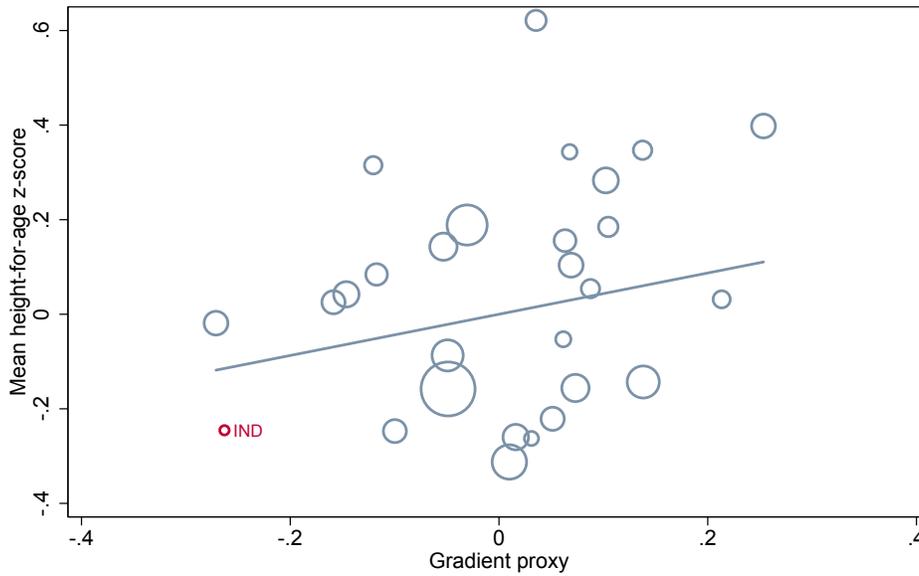
Online Appendix Table 11: Diminishing returns to inputs

<i>Input proxy:</i>	<i>Income per capita</i>		<i>Consumption per capita</i>	
	HFA z-score (1)	WFA z-score (2)	HFA z-score (3)	WFA z-score (4)
Input proxy	0.415 [0.101]	0.492 [0.069]	0.638 [0.147]	0.810 [0.095]
Input proxy (quadratic)	-0.046 [0.027]	-0.063 [0.018]	-0.104 [0.048]	-0.163 [0.028]
Indian mean of outcome	-1.607	-1.190	-1.607	-1.190
Observations	6,351	6,351	6,424	6,424

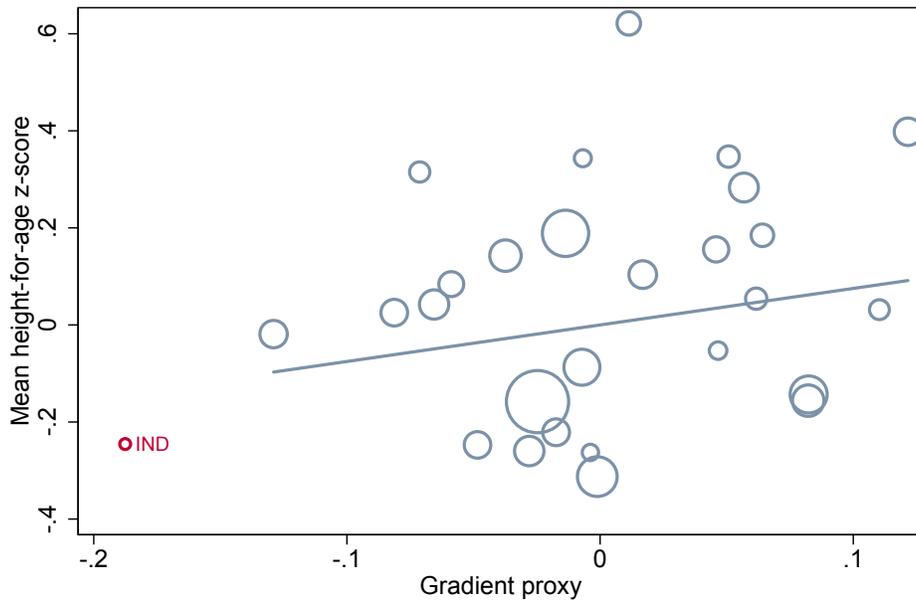
Notes: Standard errors are clustered by mother and appear in brackets. Sample uses children aged 1-59 months in IHDS 1. Control variables included in all columns are child age dummies and *Girl*. In columns (1)-(2), the *Input proxy* is monthly household income per capita, trimmed at 1 percent. In columns (3)-(4), the *Input proxy* is monthly household consumption per capita, trimmed at 1 percent. One unit of the input proxy is equal to 1,000 rupees. The negative coefficient for the quadratic term is suggestive for diminishing marginal returns. See Online Data Appendix for further details.

Online Appendix Figure 1: Child height versus gradient proxy (conditional on GDP)

Accounting Exercise 1



Accounting Exercise 2



Notes: The graphs plot the residuals obtained by regressing the mean HFA z-score in each country on log GDP per capita in birth year against the residuals obtained by regressing the gradient proxy (used in the accounting exercises in section III.F) on log GDP per capita in birth year. In the upper graph, *Gradient proxy* is defined as the weighted average of the height gap between second borns and first borns and the height gap between third- and higher-borns and first borns. In the lower graph, *Gradient proxy* is defined as the regression coefficient that is obtained by regressing, separately by country-wave, HFA z-scores on a linear birth order variable that is top-coded at 3+. GDP data are based on the Penn World Table 9.0 (Feenstra, Inklaar and Timmer, 2015).

# Online Data Appendix

## DHS surveys used in main analysis

The data sets from Sub-Saharan Africa in our main analysis are Democratic Republic of the Congo 2007 (V), Republic of the Congo (Brazzaville) 2005 (V), Cameroon 2004 (IV), Chad 2004 (IV), Ethiopia 2005 (V), Ghana 2008 (V), Guinea 2005 (V), Kenya 2008-9 (V), Lesotho 2004 (IV), Lesotho 2009 (VI), Liberia 2007 (V), Madagascar 2003-4 (IV), Malawi 2004 (IV), Mali 2006 (V), Namibia 2006-7 (V), Niger 2006 (V), Nigeria 2008 (V), Rwanda 2005 (V), Sao Tome and Principe 2008-9 (V), Senegal 2005 (IV), Sierra Leone 2008 (V), Swaziland 2006-7 (V), Tanzania 2004-5 (IV), Tanzania 2010 (VI), Uganda 2006 (V), Zambia 2007 (V), and Zimbabwe 2005-6 (V). The DHS questionnaire version (IV, V, or VI) is given in parentheses. The data set for India is India 2005-6 (NFHS-3).

## Indian Human Development Survey

The Indian Human Development Survey (IHDS) is a nationally representative two-wave panel conducted in 2005 and 2012. It contains anthropometric measures as well as data on household income and consumption. We use child anthropometric measurements from wave 1. Because birth date data are often missing in wave 1, we restrict our sample to children who were surveyed in both waves and use birth date data from wave 2. All other variables (e.g., control variables such as mother’s literacy) are from wave 1. The other way we make use of the second wave is to infer whether fertility was complete by wave 1; we construct our completed-fertility sample as women who are not pregnant in wave 2 and who did not give birth to a child between wave 1 and wave 2.

## Height-for-age z-score

For comparing height across children of different gender and age, we create normalized variables using the World Health Organization (WHO) method (WHO Multicentre Growth Reference Study Group, 2006). The WHO provides the distribution of height separately for boys and girls, by age in months from a reference population of children from Brazil, Ghana, India, Norway, Oman and the United States. Because child height has a skewed distribution, the WHO recommends a restricted application of the LMS method using a Box-Cox normal distribution. The formula used is as follows:

$$\text{z-score} = \frac{(\text{observed value}/M)^L - 1}{L \times S}$$

The WHO provides the values of  $M$ ,  $L$  and  $S$  for each reference population by gender and age.  $M$  is the reference median value for estimating the population mean,  $L$  is the power used to transform the data to remove skewness, and  $S$  is the coefficient of variation. We follow the WHO guidelines and exclude observations with HFA z-score  $> 6$  or  $< -6$ , as these are likely to be erroneous data. For similar reasons, we also exclude observations with WFA z-score  $> 5$  or  $< -6$ .

## Birth order

Birth order is defined as birth order among children ever born to one’s mother. Multiple births, such as twins, are assigned the same birth order. For a child born subsequent to a multiple birth, birth order is incremented by the size of the multiple birth, e.g., the next child born after firstborn twins is birth order 3.

## Child’s age

For all children whose anthropometric data are recorded, the DHS also provides measurement date. Our child age variable is in months, and is constructed by calculating the number of days elapsed between child’s birth and measurement date, and then converting this age into months (rounding down if the value is a non-integer). When we refer to a child as  $n$  months old, we mean the child is in its  $n^{th}$  month of life such that a child who is one week old is in its 1st month of life, hence 1 month old.

## Prenatal inputs

*Total prenatal visits* is collected for the most recent birth in the past 5 years. Hence, our sample is restricted to youngest living child from each family for this variable. It is available in all 27 African DHS's and NFHS-3. It is the mother's self-report of the total number of prenatal visits during the pregnancy. It is 0 if the mother never went for a prenatal visit. We top-code the maximum number of visits at 20.

*Mother took iron supplements* is collected for the most recent birth in the past 5 years. It is available in all 27 African DHS's and NFHS-3. It is the mother's self-report of whether she took iron supplements during the pregnancy of her youngest living child.

*Mother's total tetanus shots* is collected for the most recent birth in the past 5 years. The exception is the Democratic Republic of the Congo (2007), which collected it for all births in the past 5 years; we restrict the sample to the most recent birth for consistency. It is available in all 27 African DHS's and NFHS-3. It is the mother's self-report of the number of tetanus toxoid injections given during the pregnancy to avoid convulsions after birth. The DHS recorded having more than 7 injections as 7.

*Delivery at health facility* is collected for all births in the past 5 years. It is available in all 27 African DHS's and NFHS-3. It is calculated based on the mother's self-report of where child was delivered. Delivery at a home is defined as a delivery at any home, including the respondent's home, her parents' home, traditional birth attendant's home or some other home. Any delivery that did not occur at a home is considered a delivery at health facility.

## Postnatal inputs

*Postnatal check within 2 months* is collected for the most recent birth in the past 5 years. It is available in 13 African DHS's (Ghana 2008, Kenya 2008-9, Lesotho 2009, Liberia 2007, Namibia 2006-7, Nigeria 2008, Sao Tome and Principe 2008-9, Sierra Leone 2008, Swaziland 2006-7, Tanzania 2010, Uganda 2006, Zambia 2007, and Zimbabwe 2005-6) as well as NFHS-3. It is the mother's self-report of whether the child received a postnatal check within 2 months after it was born.

*Child taking iron pills* is collected for all births in the past 5 years. It is available in 10 African DHS's (Ghana 2008, Kenya 2008-9, Liberia 2007, Namibia 2006-7, Nigeria 2008, Sao Tome and Principe 2008-9, Sierra Leone 2008, Swaziland 2006-7, Tanzania 2010, and Uganda 2006) as well as NFHS-3. It is the mother's self-report of whether the child is currently taking iron pills.

*Child's total vaccinations* is collected for all births in the past 5 years. It is available in all 27 African DHS's and NFHS-3. It is the mother's self-report of the total number of vaccinations the child has received to date from among those that the DHS collects data on: BCG, 3 doses of DPT, 4 doses of polio, and measles. Thus the value of child's total vaccinations is 9 if the child received all vaccines. The sample is restricted to children who should have completed their course of vaccinations, specifically those age 13-59 months, as the recommended age for the vaccinations is up to age 12 months.

## Other child outcomes

*Pooled inputs.* The pooled input regressions (both input-level regressions in Table 6 and average pooled input regressions in Table 3) are based on binary versions of the seven input variables. Two of the four prenatal inputs (Mother took iron supplements and delivery at health facility) are dummy variables as are two of the three postnatal inputs (Postnatal check within 2 months and Child taking iron pills). For the pooled input regressions, we convert the remaining four input variables into dummy variables that equal 1 if the original variable for a respondent is greater than the sample median. Specifically, the dummy variables we create are 1) total prenatal visits  $>3$ ; 2) mother's total tetanus shots  $>2$ ; 3) total vaccinations  $>8$ . The average pooled inputs is then the average value across the seven health-input dummy variables.

*Child's Hb level* is the child's hemoglobin level in g/dl adjusted by altitude. It is defined for children 6 months or older and is not available for 6 surveys: Chad 2004, Kenya 2008-9, Liberia 2007, Namibia 2006-7, Nigeria 2008, and Zambia 2007.

*Infant mortality* is an indicator for whether the child is deceased. It is available in all 27 African DHS's and NFHS-3. It is the mother's self-report of whether the child is deceased. The sample is restricted to children age 13-59 months because whether they died in infancy is censored for children under age 1 year.

*Diarrhea in last 2 weeks* is collected for all births in the past 5 years. It is available in all 27 African DHS's and NFHS-3. It is the mother's self-report of whether the child had diarrhea in the 2 weeks before

the survey.

## Maternal outcomes

*Wants more children* is created based on the question, “Would you like to have another child, or would you prefer not to have more children?” It is coded as 0.5 if the mother said she is undecided whether she wants to have more children and 0 if she wants no more children or has been sterilized. This variable is missing if the woman is infecund, not currently married, or indicated that she has never had sex.

## Variables used in heterogeneity analyses

*Child sex ratio* is calculated as the number of boys aged 0-6 years old over the number of girls aged 0-6 years old in the respondent’s state-by-region (either urban or rural) and comes from the 2001 Indian census. Higher values indicate greater gender imbalance favoring boys.

*Mother’s height* is measured for mothers of children born in the 5 years preceding the survey. It is available in all 27 African DHS’s and NFHS-3. Mother’s height is converted to meters and is coded as missing if the height is less than 1.25 meters.

*Open defecation* is available for all births in the past 5 years in the full sample of 27 African DHS’s and NFHS-3. It is the mother’s self-report of whether the household has no toilet facility.

*Percent non-resident among children* is calculated as the percentage of children aged 10 years or lower who are living outside of the household, calculated at the PSU-level. Children’s age and whether they are living in the household are available in the full sample of 27 African DHS’s and NFHS-3. Each mother’s total number of living children 10 years old or younger are calculated, and summed at the PSU level. Then, the percentage of such children living outside of the household is calculated.

*Number of adult females in household* is calculated as the number of females aged 18 or above who live in the same household.

*Land scarcity* is calculated as the log of each country’s total population in 1961 over its land area in square km in 1961 and comes from the Food and Agriculture Organization of the United Nations (FAO).

## Other variables

*Mother is literate* is available for the full sample of 27 African DHS’s and NFHS-3. It is based on a literacy test that requires her to read at least part of a sentence in her language. If literacy data are missing, we impute it using the country-specific literacy rate at the number of school years she completed. If we lack information on number of school years, we include dummies for missing values of literacy in the regressions.

*Primary sampling units (PSUs)* are the smallest geographic unit used in DHS’s multi-stage sampling procedure. A PSU is typically an enumeration area or part of an enumeration area used for the national population census.

*Birth spacing* is the number of months of spacing between a child and his or her older sibling. It is calculated using the children’s age and is top-coded at 120 months. Because a firstborn child does not have an older sibling, in regressions where we control for birth spacing, for firstborns, we set the variable equal to the sample mean. The regressions also *de facto* include a flag for the observations with the imputed value because we have birth order dummies.

*Desired fertility* is available for the full sample of 27 African DHS’s and NFHS-3 and is based on the mother’s answer to a hypothetical question: if she could go back to the start of her childbearing, what is the ideal number of children she would want. We top-code the variable at 6 children.

*GDP per capita in birth year* is defined as the national expenditure-side real gross domestic product at chained purchasing power parities (in million 2011 USD) in the child’s birth year and is based on the Penn World Table 9.0 (Feenstra, Inklaar and Timmer, 2015).

## Alternative samples used

The main sample includes children age 1-59 months who have anthropometric data. (Note that we are using the convention where a child in the first month of life is 1 (rather than 0) months old.) There is a high rate of missing data for children in their 60th month of life, and hence we limit the sample to children who are 59 months old or younger. In Table 2, column (4), *Completed fertility* is the sample restricted to children whose mothers do not want to have more children, are sterilized, or are infecund. In column (9), the

sample consists of ever-born children aged 13-59 months. We exclude younger children, as infant mortality is censored for children less than 1 year old.

*Kerala & Northeast* include the Indian states Arunachal Pradesh, Assam, Kerala, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura.

*Countries with similar GDP to India* include 25 DHS's administered between 2004-2010 from countries that have height data available and that had a log GDP per capita within a 50 percent upper and lower bound of India's 2005-6 log GDP per capita. These countries are: Benin 2006 (V), Bolivia 2008 (V), Burkina Faso 2010 (VI), Cambodia 2005 (V), Cambodia 2010 (VI), Cameroon 2004 (IV), Chad 2004 (IV), Egypt 2005 (V), Ghana 2008 (V), Haiti 2005-6 (V), Honduras 2005-6 (V), Kenya 2008-9 (V), Lesotho 2004 (IV), Lesotho 2009 (VI), Mali 2006 (V), Moldova 2005 (V), Nigeria 2008 (V), Sao Tome and Principe 2008-9 (V), Senegal 2005 (IV), Senegal 2010-11 (VI), Tanzania 2010 (VI), Timor-Leste 2009-10 (VI), Zambia 2007 (V), Zimbabwe 2005-6 (V), and Zimbabwe 2010-11 (VI).

*Europe, Central & West Asia* includes 16 DHS's spanning 1995-2012 for European, Central and West Asian countries with height data available: Albania 2008-2009 (V), Armenia 2005 (V), Armenia 2010 (VI), Azerbaijan 2006 (V), Jordan 2002 (IV), Jordan 2007 (V), Jordan 2012 (VI), Kazakhstan 1995 (III), Kazakhstan 1999 (IV), Kyrgyz Republic 1997 (III), Kyrgyz Republic 2012 (VI), Moldova 2005 (V), Tajikistan 2012 (VI), Turkey 1998 (IV), Turkey 2003 (IV), and Uzbekistan 1996 (III). Because of the relative paucity of surveys in this region, we expand the time period to cover 1995 to 2012 rather than just 2004 to 2010.

*Bangladesh & Pakistan* includes 4 DHS's: Bangladesh 2004 (IV), Bangladesh 2007 (V), Bangladesh 2011 (VI), and Pakistan 2012-13 (VI). We expand the time period beyond the 2004 to 2010 time period used in our main analysis because earlier DHS's in Pakistan do not have data on child's height.

*NFHS-1 and NFHS-2* are the first two rounds of the National Family Health Survey; our main sample for India is the most recent round, NFHS-3. NFHS-1 (1992-3) collects height data for children up to age 4, while NFHS-2 (1998-9) does so for children up to age 3. Due to data availability, pooled inputs used in Online Appendix Table 5 include 5 of the 7 inputs used in NFHS-3, and total vaccinations is based on 8 of the 9 vaccines used in NFHS-3. Specifically, NFHS-1 and NFHS-2 do not have data on whether the child is taking iron pills, whether the child had a postnatal check within 2 months of birth, and whether the child received one of the four doses of the polio vaccine. In Online Appendix Tables 7-9, the comparison sample for NFHS-1 includes African DHS's between 1991 and 1997. These are Cameroon 1991 (II), Ghana 1993 (II), Kenya 1993 (II), Namibia 1992 (II), Madagascar 1992 (II), Madagascar 1997 (III), Mali 1995-96 (III), Malawi 1992 (II), Niger 1992 (II), Rwanda 1992 (II), Senegal 1992-93 (II), Chad 1996-97 (III), Tanzania 1991-92 (II), Tanzania 1996 (III), Uganda 1995 (III), Zambia 1992 (II), Zambia 1996 (III) and Zimbabwe 1994 (III). Ghana 1993, Madagascar 1997, Mali 1995-96 and Zimbabwe 1994 only have data for children aged 3 years and younger. The other DHS's have data for children aged 4 years and younger.

## Accounting exercise

For the first accounting exercise in Online Appendix Table 2, the *Gradient proxy* is defined as the weighted average of the height gap between first borns and second borns and the height gap between first borns and third- and higher-borns in each country-wave. The weight for the height gap between first borns and second borns is equal to the total number of ever-born second borns divided by the total number of ever-born second- and higher-borns in the completed fertility sample in that country-wave. Analogously, the weight for the height gap between first borns and third- and higher-borns is equal to the total number of ever-born third- and higher-borns divided by the total number of ever-born second- and higher-borns in the completed fertility sample for that country-wave.

For the second accounting exercise in Online Appendix Table 2, the *Gradient proxy* is defined as the regression coefficient that is obtained by regressing, separately by country-wave, HFA z-scores on a linear birth order variable that is top-coded at 3+.