ONLINE APPENDIX FOR: THE TWIN INSTRUMENT: FERTILITY AND HUMAN CAPITAL INVESTMENT

Sonia Bhalotra
The University of Essex and ISER

Damian Clarke
Universidad de Santiago de Chile

Contents

E-mail: srbhal@essex.ac.uk (Bhalotra); damian.clarke@usach.cl (Clarke)

Journal of the European Economic Association
Preprint prepared on 2 October 2019 using jeea.cls *v1.0.*

A. Appendix Figures and Tables

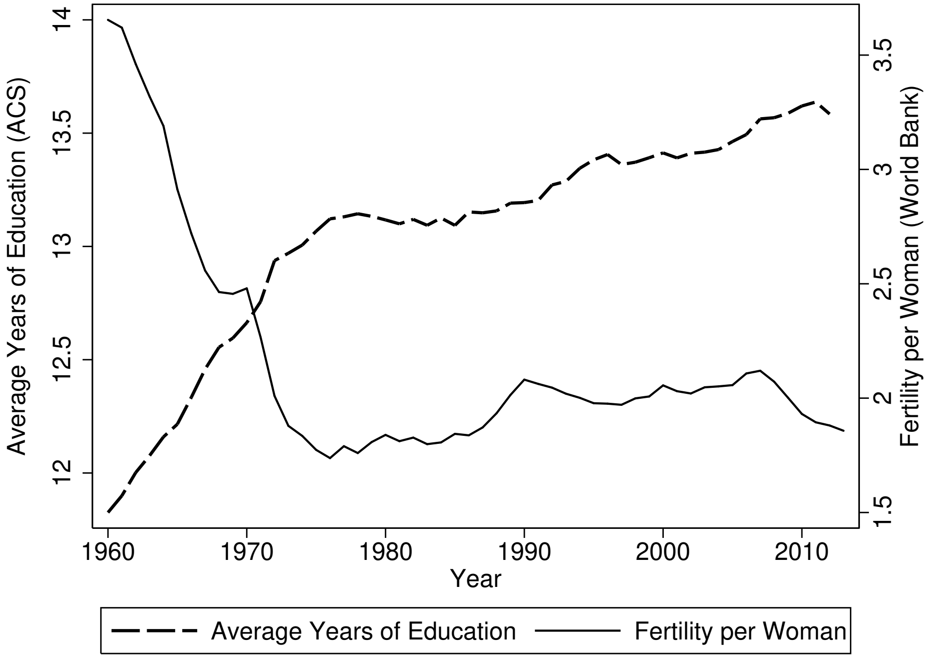
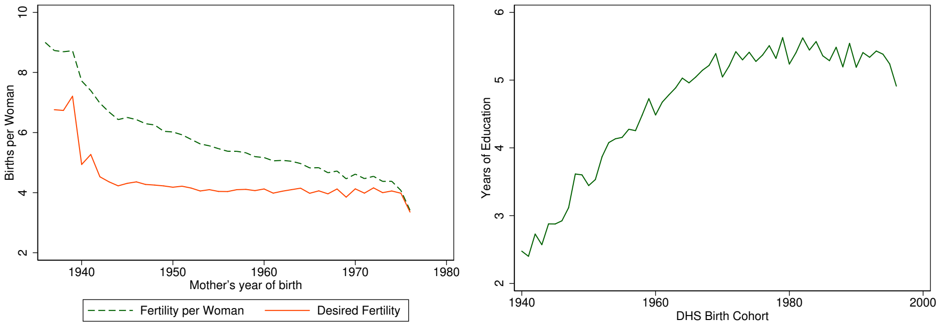


FIGURE A1. Education and Fertility Trends (USA). Notes: Trends in fertility and education are compiled from the World Bank databank and the American Community Surveys (ACS), respectively. Trends in fertility are directly reported by the World Bank as completed fertility per woman were she exposed to prevailing rates in a given year for her whole fertile life. Education is calculated using all women aged over 25 years in the ongoing ACS (2001-2013) collected by the United States Census Bureau. The figure presents average completed education for all women aged 25 in the year in question.



(A) Trends in Fertility

(B) Trend in Education

FIGURE A2. Education and Fertility (Developing Countries). Notes: Cohorts are made up of all individuals from the DHS who are aged over 35 years (for fertility), and over 15 years (for education). In each case the sample is restricted to those who have approximately completed fertility and education respectively. A full list of country and survey years are available in Table A1 and full summary statistics for these variables are provided in Table A2.

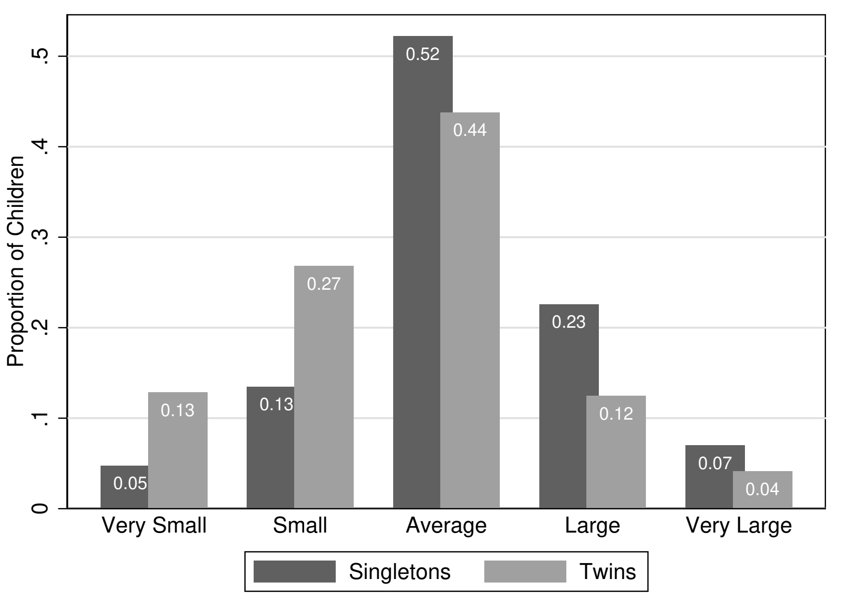


FIGURE A3. Birth Size of Twins versus Singletons (Developing Countries). Notes: Estimation sample consists of all surveyed births from DHS countries occurring within 5 years prior to the date of the survey. For each of these births, all mothers retrospectively report the (subjective) size of the baby at the time of birth.

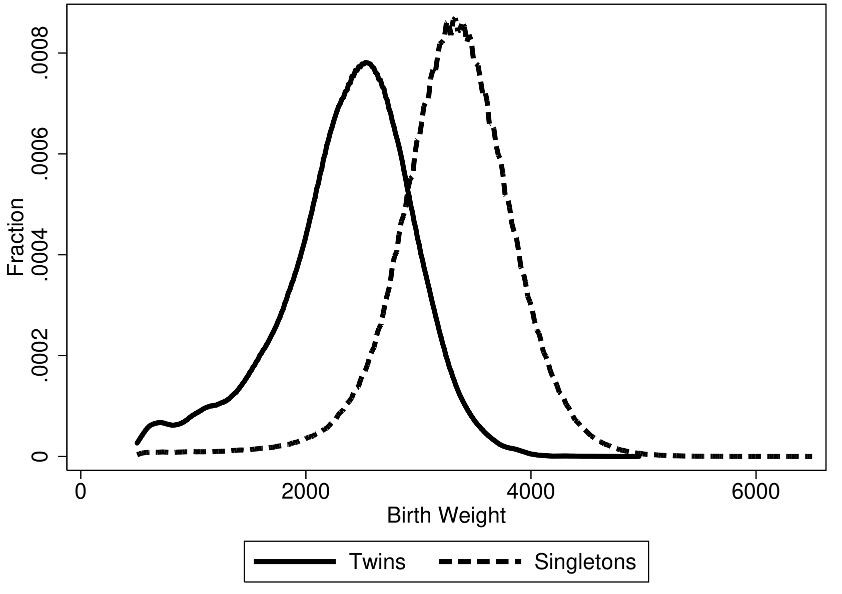
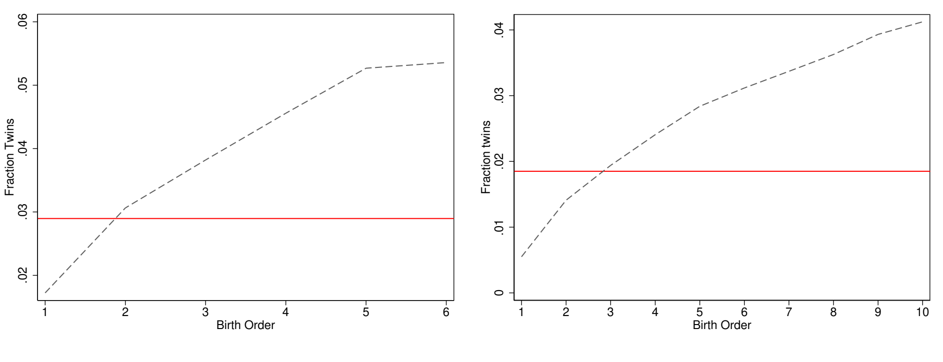


FIGURE A4. Birth Weight of Twins versus Singletons (USA). Notes: Estimation sample consists of all non-ART births from NVSS data between 2009 and 2013. Birthweights below 500 grams and above 6,500 grams are trimmed from the sample.



(A) United States

(B) Developing Countries

FIGURE A5. Proportion of Twins by Birth Order. Notes: The fraction of twin births is calculated from the full sample of non-ART users in NVSS data from 2009-2013 (panel A), and the full sample of DHS data (panel B). The solid line represents the average fraction of twins in the full sample (2.89% in US, 1.88% in DHS), while the dotted line presents twin frequency by birth order. The dotted line joins points at each birth order. Birth orders greater than 6 (USA) and 10 (DHS) are removed, as they account for less than 1% of all recorded births.

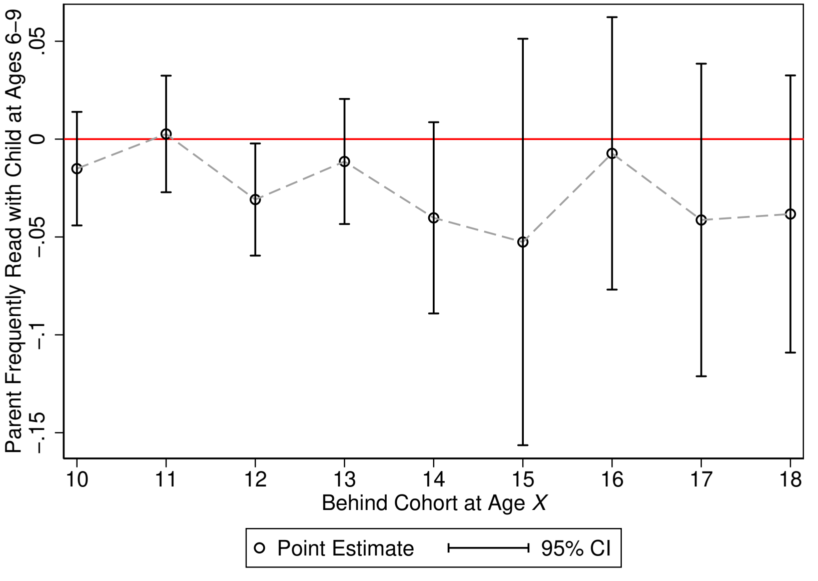
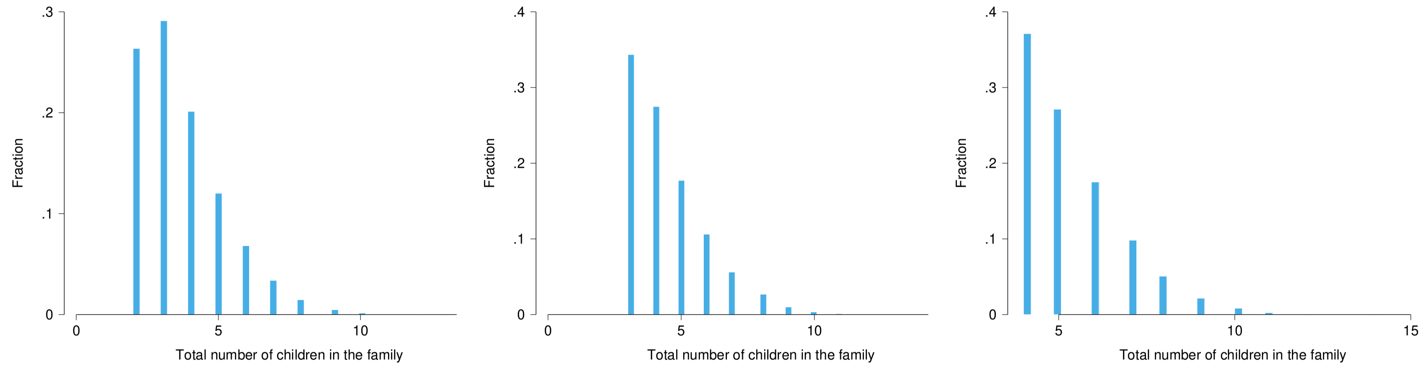


FIGURE A6. Reading with Children at Ages 6-9 and Future School Completion Rates. Notes: Each point estimate and 95% confidence interval displays the coefficient from a separate regression of whether an individual is behind their cohort at age $x \in \{10, \ldots, 18\}$ on whether the parent frequently read with the child between the ages of 6–9 years. All remaining details are identical to those in Figure 2.



(A) Two-Plus Group (B) Three-Plus Group (C) Four-Plus Group

FIGURE A7. Total Family Size in Analysis Samples. Notes: Histograms display the total family size of families meeting inclusion criteria for each estimation sample (two-plus, three-plus, and four-plus). By definition, the two-plus sample only includes families with at least two births, the three-plus sample only includes families with at least three births, and the four-plus sample only includes families with at least four births.

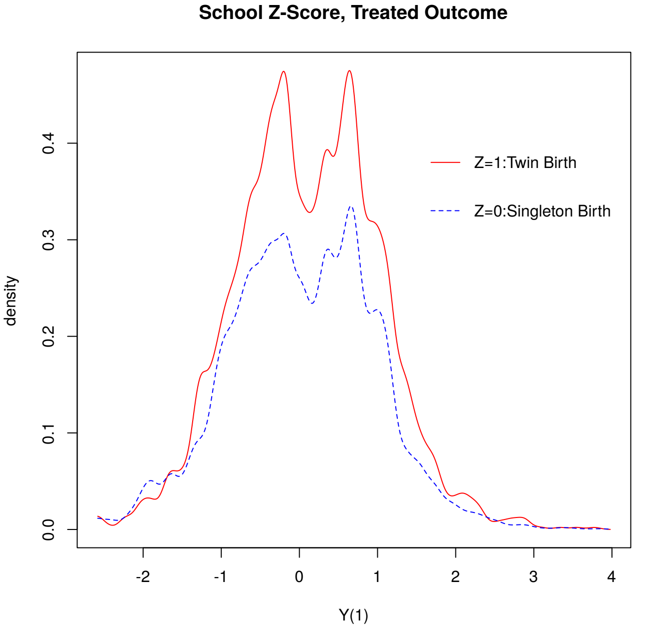
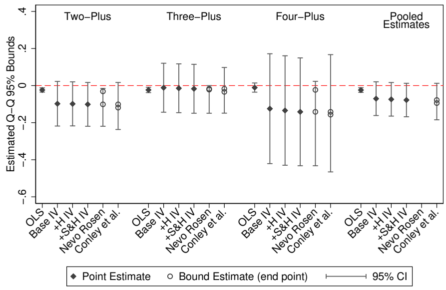
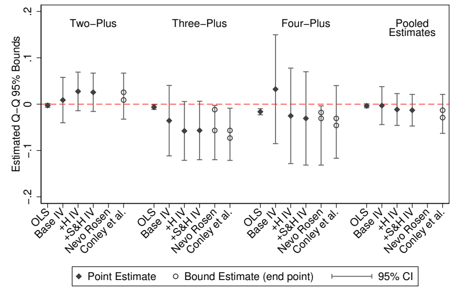


FIGURE A8. Density Test of Instrumental Validity from Kitagawa (2015). Notes: Kernel density plots document the sub-densities of the outcome variable of interest in IV regressions (school Z-score) for children preceding twins and for children not preceding twins in the 2+ sample. "Treated" refers to families with at least 3 children, and so both densities document frequencies only for this group. The Kitagawa (2015) test consists of determining whether the two densities intersect, with intersection being evidence of instrumental *in*validity. We follow Kitagawa in using a Gaussian kernel and bandwidth of 0.08. Outliers are suppressed from the graph to ease visualisation of the sub-densities. Results for the full version of the test including controls along with p-values associated with instrumental invalidity are presented in Table A12.

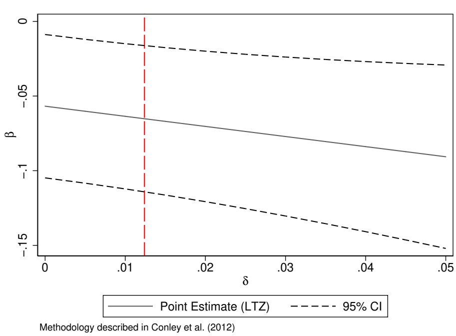


(A) Education Z-Score

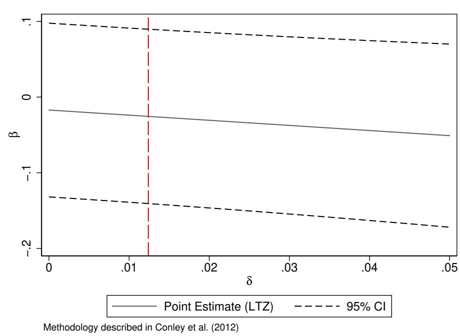


(B) Excellent Health

FIGURE A9. Parameter and Bound Estimates of the Q–Q Trade-off (USA). Notes: Refer to notes to Figure 5. Identical bounds are presented, but in this case based on NHIS data (with considerably fewer observations).

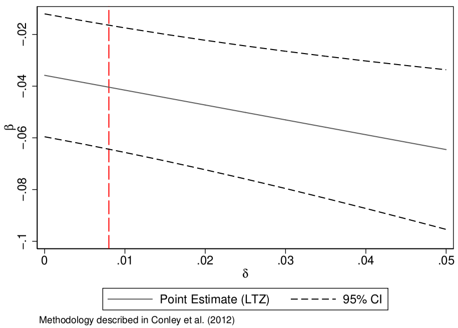


(A) Excellent Health

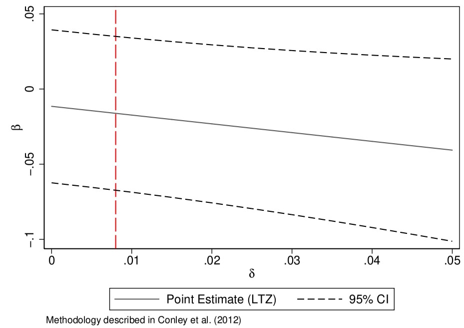


(B) Education Z-Score

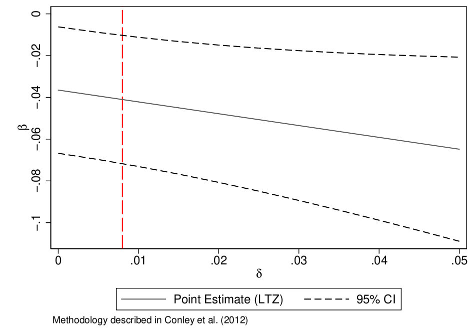
FIGURE A10. Plausibly Exogenous Bounds: (USA, 3+ Sample). Notes: See notes to Figure 6. An identical approach is employed, however now using USA (NHIS) data.



(A) Pooled Bounds

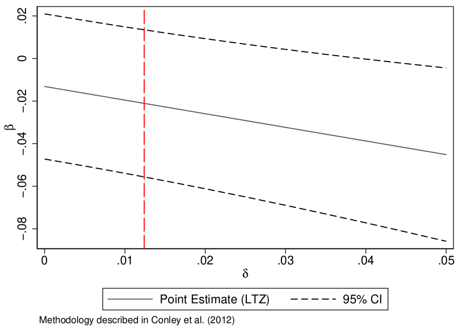


(B) Two Plus

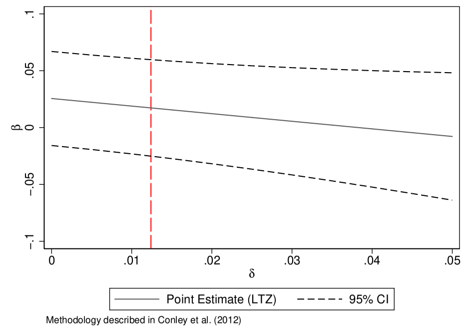


(C) Four Plus

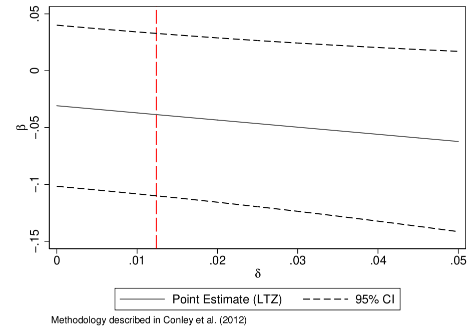
FIGURE A11. Plausibly Exogenous Bounds: School Z-Score (Developing Countries Alternative Parity Estimates). Notes: Refer to notes to Figure 6 of the main text.



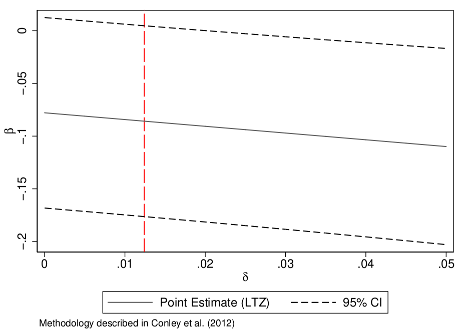
(A) Excellent Health (Pooled)



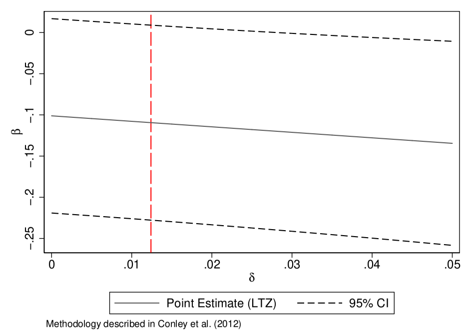
(B) Excellent Health (2+)



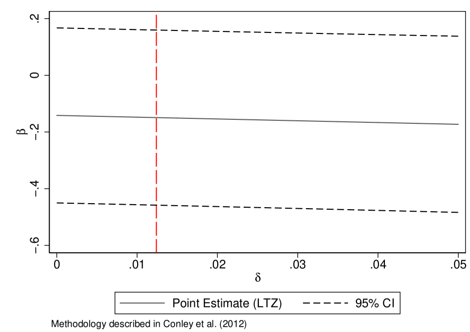
(C) Excellent Health (4+)



(D) Education Z-Score (Pooled)



(E) Education Z-Score (2+)



(F) Education Z-Score (4+)

FIGURE A12. Plausibly Exogenous Bounds: (USA Alternative Parity Estimates). Notes: Refer to notes to Figure 6 of the main text.

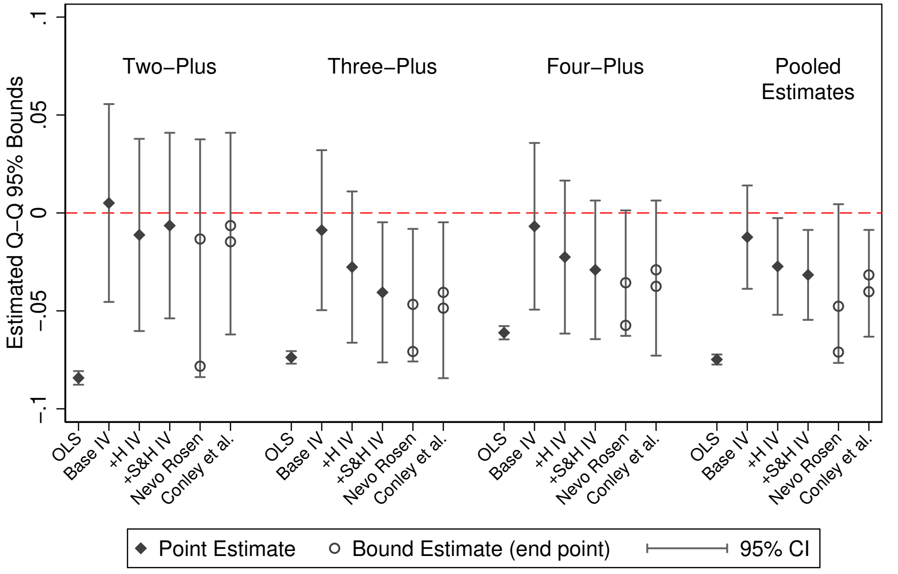
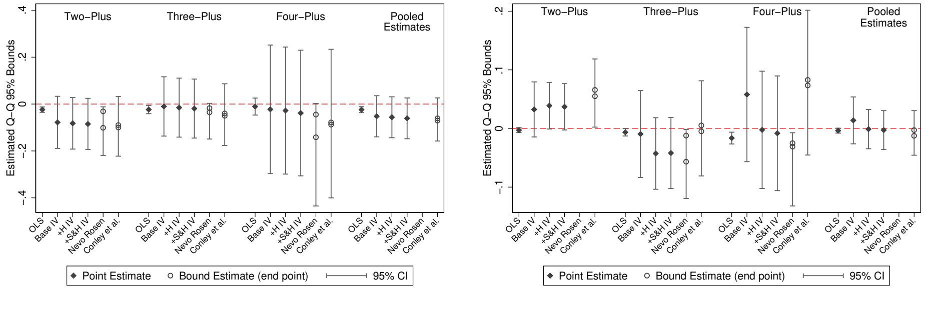


FIGURE A13. IV Bounds Adding the Sex Mix Instrument for Fertility (Developing Countries). Notes: Refer to notes to Figure 5. Identical bounds are estimated, however here fertility is instrumented by both twin birth, and the "same sex instrument" indicating whether the first N births are of an identical sex for each of the N+ samples. In bounding procedures, the same priors for twin birth are used, and priors for the same sex instrument assume that it is a valid instrument.



(A) Education Z-Score

(B) Excellent Health

FIGURE A14. IV Bounds Adding the Sex Mix Instrument for Fertility (USA). Notes: Refer to notes to Figure A9 and A13.

TABLE A1. Full Survey Countries and Years (DHS)

		Survey Year						
COUNTRY	INCOME	1	2	3	4	5	6	7
Albania	Middle	2008						
Armenia	Low	2000	2005	2010				
Azerbaijan	Middle	2006						
Bangladesh	Low	1994	1997	2000	2004	2007	2011	
Benin	Low	1996	2001	2006				
Bolivia	Middle	1994	1998	2003	2008			
Brazil	Middle	1991	1996					
Burkina Faso	Low	1993	1999	2003	2010			
Burundi	Low	2010						
Cambodia	Low	2000	2005	2010				
Cameroon	Middle	1991	1998	2004	2011			
Central African Republic	Low	1994						
Chad	Low	1997	2004					
Colombia	Middle	1990	1995	2000	2005	2010		
Comoros	Low	1996						
Congo Brazzaville	Middle	2005	2011					
Congo Democratic Republic	Low	2007						
Cote d Ivoire	Low	1994	1998	2005	2012			
Dominican Republic	Middle	1991	1996	1999	2002	2007		
Egypt	Low	1992	1995	2000	2005	2008		
Ethiopia	Low	2000	2005	2011				
Gabon	Middle	2000	2012					
Ghana	Low	1993	1998	2003	2008			
Guatemala	Middle	1995						
Guinea	Low	1999	2005					
Guyana	Middle	2005	2009					
Haiti	Low	1994	2000	2006	2012			
Honduras	Middle	2005	2011					
India	Low	1993	1999	2006				
Indonesia	Low	1991	1994	1997	2003	2007	2012	
Jordan	Middle	1990	1997	2002	2007			
Kazakhstan	Middle	1995	1999					
Kenya	Low	1993	1998	2003	2008			
Kyrgyz Republic	Low	1997						
Lesotho	Low	2004	2009					
Liberia	Low	2007						
Madagascar	Low	1992	1997	2004	2008			
Malawi	Low	1992	2000	2004	2010			
Maldives	Middle	2009						

Mali	Low	1996	2001	2006				
Moldova	Middle	2005						
Morocco	Middle	1992	2003					
Mozambique	Low	1997	2003	2011				
Namibia	Middle	1992	2000	2006				
Nepal	Low	1996	2001	2006	2011			
Nicaragua	Low	1998	2001					
Niger	Low	1992	1998	2006				
Nigeria	Low	1990	1999	2003	2008			
Pakistan	Low	1991	2006					
Paraguay	Middle	1990						
Peru	Middle	1992	1996	2000				
Philippines	Middle	1993	1998	2003	2008			
Rwanda	Low	1992	2000	2005	2010			
Sao Tome and Principe	Middle	2008						
Senegal	Middle	1993	1997	2005	2010			
Sierra Leone	Low	2008						
South Africa	Middle	1998						
Swaziland	Middle	2006						
Tanzania	Low	1992	1996	1999	2004	2007	2010	2012
Togo	Low	1998						
Turkey	Middle	1993	1998	2003				
Uganda	Low	1995	2000	2006	2011			
Ukraine	Middle	2007						
Uzbekistan	Middle	1996						
Vietnam	Low	1997	2002					
Yemen	Low	1991						
Zambia	Low	1992	1996	2002	2007			
Zimbabwe	Low	1994	1999	2005	2010			

Notes: Country income status is based upon World Bank classifications described at http://data.worldbank.org/about/country-classifications and available for download at http://siteresources.worldbank.org/DATASTATISTICS/Resources/OGHIST.xls (consulted 1 April, 2014). Income status varies by country and time. Where a country's status changed between DHS waves only the most recent status is listed above. Middle refers to both lower-middle and upper-middle income countries, while low refers just to those considered to be low-income economies.

TABLE A2. Summary Statistics

	Developing Countries			United States		
	Single	Twins	All	Single	Twins	All
Mother's Characteristics						
Fertility	4.515	6.998	4.646	1.925	3.094	1.955
	(2.324)	(2.583)	(2.403)	(1.002)	(1.185)	(1.024)
Age	35.03	37.15	35.15	36.16	37.24	36.19
	(6.801)	(6.542)	(6.804)	(8.423)	(8.069)	(8.415)
Education	4.366	3.280	4.309	12.57	12.74	12.58
	(4.642)	(4.162)	(4.624)	(2.310)	(2.220)	(2.308)
Height	155.5	157.3	155.6	-	-	-
	(7.090)	(7.055)	(7.101)	-	-	-
BMI	23.77	23.82	23.77	27.65	28.12	27.66
	(5.134)	(5.133)	(5.134)	(6.715)	(7.326)	(6.732)
Pr(BMI)<18.5	0.121	0.103	0.120	0.0197	0.0159	0.0196
	(0.327)	(0.303)	(0.325)	(0.139)	(0.125)	(0.139)
Excellent Health	-	-	-	0.318	0.324	0.318
	-	-	-	(0.465)	(0.468)	(0.465)
Children's Outcomes						
Age	10.91	11.10	10.92	11.19	10.77	11.18
	(3.564)	(3.548)	(3.563)	(3.891)	(3.901)	(3.891)
Education (Years)	3.192	2.868	3.170	5.151	4.650	5.139
	(3.162)	(3.011)	(3.154)	(3.851)	(3.769)	(3.850)
Education (Z-Score)	0.00191	-0.0958	-0.00449	0.00274	-0.110	0.0000
	(0.981)	(0.973)	(0.981)	(1.001)	(0.950)	(1.000)
Excellent Health	-	-	-	0.531	0.541	0.531
	-	-	-	(0.499)	(0.498)	(0.499)
Fraction Twin			0.0188			0.0257
			(0.136)			(0.158)
Birth Order Twin			4.383			2.196
			(2.386)			(1.064)
Observations	1,107,515	21,214	1,128,729	221,381	5,832	227,213

Notes: Summary statistics are presented for the full estimation sample consisting of all children aged 6–18 years born to the 553,496 mothers responding to any publicly available Demographic and Health Survey or aged 18 years and under born to the 88,178 mothers responding to the National Health Interview Survey from 2004 to 2014. Group means are presented with standard deviation below in parenthesis. Education is reported as total years attained, and Z-score presents educational attainment relative to birth and country cohort for DHS, and birth quarter cohort for NHIS (mean 0, std deviation 1). Maternal height is reported in centimetres, and BMI is weight in kilograms over height in metres squared. For a full list of DHS country and years of survey, see Appendix Table A1.

TABLE A3. The Quantity–Quality Trade-off and the Twin Instrument: Recent Studies Based on Linear IV Models

Author	Data, Period	Controls Included	Sample	Estimates OLS	Estimates IV
(1) Black et al. (2005)	Norway matched administrative files of individuals aged 16-74 during 1986-2000, (children $>$ 25 years). Outcome is completed years of education.	Age, parents' age, parents' education, sex.	Two Plus	-0.060 (0.003)	-0.038 (0.047)
			Three Plus	-0.076 (0.004)	-0.016 (0.044)
			Four Plus	-0.059 (0.006)	-0.024 (0.059)
(2) Cáceres-Delpiano (2006)	USA 1980 Census Five-Percent Public Use Micro Sample. Children aged 6-16 years. Outcome (reported here) is an indicator of whether the child is behind his or her cohort.	Age, state of residence, mother's education, race, mother's age, sex.	Two Plus	0.011 (0.000)	0.002 (0.003)
			Three Plus	0.017 (0.001)	0.010 (0.006)
(3) Angrist et al. (2010)	Israel 20% public-use microdata samples from 1995 and 1983 censuses, 18-60 year old respondents. Outcome (reported here) is highest grade completed.	Age, missing month of birth, mother's age, age at first birth and age at immigration, mother's and father's place of birth, and census year.	Two Plus	-0.145 (0.005)	0.174 (0.166)
			Three Plus	-0.143 (0.005)	0.167 (0.117)
(4) Li et al. (2008)	The 1 percent sample of the 1990 Chinese Population Census. Subjects are 6-17 year olds with mothers who are 35 years of age or younger. Outcome (reported here) is years of schooling.	Child age, gender, ethnic group, birth order, and place of residence. Parental age and educational level.	Two Plus	-0.031 (-29.6)†	0.002 (0.18)†
			Three Plus	-0.038 (-21.4)†	-0.024 (-1.70)†
(5) Fitzsimons and Malde (2014)	Mexican Survey data (ENCASEH) from 1996-1999. Subjects are 12-17 year olds. Outcome (reported here) is years of schooling.	Parent's age, parents' years of schooling and schooling dummies, birth spacing, household goods (rooms, land, water, etc).	Two Plus	-0.020 (0.001)	-0.019 (0.015)
			Three Plus	-0.020 (0.001)	0.007 (0.025)
			Four Plus	-0.018 (0.002)	-0.032 (0.036)

Author	Data, Period	Controls Included	Sample	Estimates	
				OLS	IV
(6) Rosenzweig and Zhang (2009)	The Chinese Child Twins Survey (CCTS), 2002-2003. Individuals selected from twins' (aged 7-18) and non-twin households. Outcome (reported here) is years of schooling	Mother's age at time of birth, child gender and age.	Reduced Form		-0.307 (1.92)†
			Reduced Form + Bwt		-0.225 (1.31)†
(7) Ponczek and Souza (2012)	1991 Brazilian Census microdata, 10 and 20% sample. Children of 10-15 years, and 18-20 years old. Outcome reported here is years of school completed.	Child's gender, age and race controls,; mother and family head's years of schooling, and age.	Two Plus (M)	-0.233 (0.010)	-0.137 (0.146)
			Two Plus (F)	-0.277 (0.015)	-0.372 (0.198)
			Three Plus (M)	-0.230 (0.010)	-0.060 (0.164)
			Three Plus (F)	-0.283 (0.015)	-0.634 (0.194)

Notes: Individual sources discussed further in the body of the text. Estimates reported in each study are presented along with their standard errors in parenthesis. Parentheses marked as † contain the t-statistic rather than the standard error. A summarised version of this table is presented as Table 1 in Bhalotra and Clarke (2018).

TABLE A4. OLS Estimates with and without Birth Order Controls (Pooled DHS Data)

	No Birth Order FEs			Birth Order FEs			
	(1) Base	(2) +S	(3) +S+H	(4) No Fertility	(5) Base	(6) +S	(7) +S+H
Fertility	-0.117***	-0.101***	-0.067***		-0.128***	-0.108***	-0.071***
	(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)
Birth Order 2				-0.178***	-0.059***	-0.063***	-0.040***
				(0.004)	(0.004)	(0.003)	(0.003)
Birth Order 3				-0.358***	-0.103***	-0.111***	-0.071***
				(0.005)	(0.005)	(0.005)	(0.005)
Birth Order 4				-0.499***	-0.103***	-0.120***	-0.078***
				(0.006)	(0.006)	(0.006)	(0.006)
Birth Order 5				-0.601***	-0.065***	-0.092***	-0.062***
				(0.007)	(0.007)	(0.007)	(0.007)
Birth Order 6				-0.693***	-0.021**	-0.061***	-0.050***
				(0.008)	(0.009)	(0.009)	(0.008)
Birth Order 7				-0.755***	0.047***	-0.010	-0.022**
				(0.009)	(0.010)	(0.010)	(0.010)
Birth Order 8				-0.786***	0.140***	0.064***	0.025**
				(0.010)	(0.012)	(0.012)	(0.011)
Birth Order 9				-0.844***	0.202***	0.105***	0.041***
				(0.012)	(0.014)	(0.014)	(0.014)
Birth Order $\geq$ 10				-0.865***	0.388***	0.256***	0.145***
				(0.014)	(0.016)	(0.016)	(0.015)
Observations	1,128,729	1,128,729	1,128,729	1,128,729	1,128,729	1,128,729	1,128,729

TABLE A5. OLS Estimates with and without Birth Order Controls (USA)

	No Birth Order FEs			Birth Order FEs			
	(1) Base	(2) +S	(3) +S+H	(4) No Fertility	(5) Base	(6) +S	(7) +S+H
Fertility	-0.026***	-0.027***	-0.023***		-0.023***	-0.024***	-0.020***
	(0.004)	(0.004)	(0.004)		(0.004)	(0.004)	(0.004)
Birth Order 2				-0.049***	-0.032***	-0.033***	-0.033***
				(0.008)	(0.008)	(0.008)	(0.008)
Birth Order 3				-0.103***	-0.061***	-0.060***	-0.059***
				(0.015)	(0.015)	(0.015)	(0.015)
Birth Order 4				-0.121***	-0.053**	-0.050**	-0.046*
				(0.025)	(0.025)	(0.025)	(0.025)
Birth Order 5				-0.095**	0.002	0.012	0.018
				(0.046)	(0.045)	(0.045)	(0.045)
Birth Order 6				-0.194**	-0.065	-0.057	-0.043
				(0.083)	(0.081)	(0.081)	(0.081)
Birth Order 7				-0.236	-0.079	-0.062	-0.047
				(0.157)	(0.157)	(0.156)	(0.158)
Birth Order 8				0.012	0.191	0.196	0.220
				(0.498)	(0.497)	(0.495)	(0.495)
Birth Order 9				-0.460***	-0.259**	-0.250**	-0.207
				(0.107)	(0.115)	(0.122)	(0.132)
Birth Order $\geq$ 10				-0.421***	-0.181***	-0.184***	-0.148**
				(0.054)	(0.056)	(0.054)	(0.068)
Observations	163,931	163,931	163,931	163,931	163,931	163,931	163,931

TABLE A6. Full Output on Health and Socioeconomic Controls from IV Estimates (Developing Countries)

Dependent Variable	2+		3+		4+	
School Z-Score	+H	+S&H	+H	+S&H	+H	+S&H
Fertility	-0.018	-0.012	-0.041**	-0.048**	-0.038*	-0.036*
	(0.027)	(0.026)	(0.021)	(0.020)	(0.021)	(0.019)
Mother's Height	0.057***	0.067***	0.040***	0.049***	0.028***	0.040***
	(0.008)	(0.013)	(0.008)	(0.011)	(0.007)	(0.009)
Mother's Height Squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mother's BMI	0.094***	0.027***	0.094***	0.039***	0.092***	0.045***
	(0.004)	(0.006)	(0.003)	(0.005)	(0.003)	(0.005)
Mother's BMI Squared	-0.001***	-0.000	-0.001***	-0.000***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Poorest Quintile		-0.268***		-0.257***		-0.241***
		(0.014)		(0.011)		(0.011)
Quintile 2		-0.110***		-0.110***		-0.086***
		(0.011)		(0.010)		(0.010)
Quintile 3		-0.036***		-0.029***		0.002
		(0.011)		(0.010)		(0.010)
Quintile 4		0.025**		0.056***		0.113***
		(0.010)		(0.010)		(0.010)
Richest Quintile		0.151***		0.227***		0.325***
		(0.011)		(0.011)		(0.012)
Birth Order 2			-0.115***	-0.079***	-0.090***	-0.067***
			(0.014)	(0.012)	(0.012)	(0.011)
Birth Order 3					-0.173***	-0.128***
					(0.025)	(0.022)
Observations	259,958	259,958	395,687	395,687	409,576	409,576
R-Squared	0.079	0.156	0.080	0.161	0.073	0.156
Joint Test Maternal Education (χ^2)		1484.456		1843.860		134.845
Joint Test Interactions (χ^2)		1411.989		1128.894		544.073

Notes: Full output is presented from IV regressions displayed in Table 6 on health and socioeconomic controls from models denoted "+H" (adding health controls) and "+S&H" (adding health and socioeconomic controls). Additionally, fixed effects for years of education of the mother are included in regressions though are not displayed in the interests of space. These fixed effects show a positive gradient with higher education associated with additional child education. Full notes are available in Table 6.

TABLE A7. Full Output on Health and Socioeconomic Controls from IV Estimates (USA Education)

Dependent Variable	2+		3+		4+	
School Z-Score	+H	+S&H	+H	+S&H	+H	+S&H
Fertility	-0.099	-0.101*	-0.014	-0.017	-0.134	-0.142
	(0.061)	(0.060)	(0.067)	(0.067)	(0.151)	(0.148)
Excellent Health	0.139	0.131	-0.046	-0.026	0.326	0.353
	(0.181)	(0.178)	(0.227)	(0.229)	(0.603)	(0.607)
Very good Health	0.141	0.134	-0.048	-0.027	0.294	0.323
	(0.181)	(0.178)	(0.227)	(0.229)	(0.603)	(0.607)
Good Health	0.080	0.086	-0.100	-0.065	0.248	0.289
	(0.181)	(0.178)	(0.227)	(0.229)	(0.603)	(0.607)
Fair Health	0.006	0.024	-0.185	-0.139	0.200	0.249
	(0.181)	(0.179)	(0.228)	(0.230)	(0.603)	(0.607)
Poor Health	-0.098	-0.070	-0.291	-0.231	-0.020	0.047
	(0.186)	(0.183)	(0.234)	(0.235)	(0.610)	(0.613)
Mother's Height	0.079	0.061	0.187	0.168	0.123	0.127
	(0.102)	(0.102)	(0.139)	(0.138)	(0.240)	(0.240)
Mother's Height Squared	-0.001	-0.000	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Smoked Prior to Birth	-0.047***	-0.041***	-0.051**	-0.046**	-0.055	-0.051
	(0.015)	(0.015)	(0.021)	(0.021)	(0.040)	(0.040)
No Response to Smoking	0.046*	0.041	0.062*	0.052	0.094*	0.079
	(0.026)	(0.025)	(0.034)	(0.033)	(0.055)	(0.055)
Birth Order 2			-0.039*	-0.033	0.022	0.027
			(0.023)	(0.023)	(0.043)	(0.042)
Birth Order 3					0.007	0.018
					(0.083)	(0.080)
Observations	61,267	61,267	47,308	47,308	21,352	21,352
R-Squared	0.000	0.003	0.004	0.008	-0.005	-0.004
Joint Test Maternal Education (χ^2)		80.468		68.916		17.029

Notes: Full output is presented from IV regressions displayed in Table 7 on health and socioeconomic controls from models denoted "+H" (adding health controls) and "+S&H" (adding health and socioeconomic controls). Additionally, fixed effects for years of education of the mother are included in regressions though are not displayed in the interests of space. These fixed effects show a positive gradient with higher education associated with additional child education. Full notes are available in Table 7.

TABLE A8. Full Output on Health and Socioeconomic Controls from IV Estimates (USA Health)

Dependent Variable	2+		3+		4+	
Excellent Health	+H	+S&H	+H	+S&H	+H	+S&H
Fertility	0.027	0.026	-0.058*	-0.057*	-0.025	-0.031
	(0.021)	(0.021)	(0.032)	(0.032)	(0.053)	(0.051)
Excellent Health	0.501***	0.499***	0.450***	0.454***	0.089	0.090
	(0.090)	(0.090)	(0.136)	(0.136)	(0.133)	(0.127)
Very good Health	-0.022	-0.023	-0.076	-0.072	-0.435***	-0.433***
	(0.090)	(0.090)	(0.136)	(0.136)	(0.134)	(0.128)
Good Health	-0.112	-0.107	-0.172	-0.164	-0.547***	-0.541***
	(0.090)	(0.090)	(0.136)	(0.136)	(0.133)	(0.127)
Fair Health	-0.096	-0.087	-0.146	-0.137	-0.492***	-0.485***
	(0.090)	(0.090)	(0.136)	(0.136)	(0.134)	(0.128)
Poor Health	-0.097	-0.085	-0.132	-0.120	-0.598***	-0.588***
	(0.092)	(0.091)	(0.139)	(0.138)	(0.137)	(0.132)
Mother’s Height	-0.018	-0.024	-0.001	-0.003	0.013	0.022
	(0.046)	(0.046)	(0.068)	(0.068)	(0.120)	(0.120)
Mother’s Height Squared	0.000	0.000	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Smoked Prior to Birth	0.016**	0.019***	0.008	0.011	0.008	0.011
	(0.007)	(0.007)	(0.010)	(0.010)	(0.019)	(0.019)
No Response to Smoking	0.001	-0.001	-0.004	-0.005	-0.025	-0.027
	(0.011)	(0.011)	(0.016)	(0.016)	(0.027)	(0.027)
Birth Order 2			0.009	0.009	0.005	0.006
			(0.008)	(0.008)	(0.012)	(0.011)
Birth Order 3					0.007	0.010
					(0.022)	(0.022)
Observations	70,277	70,277	53,393	53,393	24,358	24,358
R-Squared	0.295	0.298	0.295	0.296	0.304	0.306
Joint Test Maternal Education (χ^2)		113.866		31.169		18.390

Notes: Full output is presented from IV regressions displayed in Table 7 on health and socioeconomic controls from models denoted “+H” (adding health controls) and “+S&H” (adding health and socioeconomic controls). Additionally, fixed effects for years of education of the mother are included in regressions though are not displayed in the interests of space. These fixed effects show a positive gradient with higher education associated with additional child education. Full notes are available in Table 7.

TABLE A9. First Stages for Non-Linear IV Estimates

Instrument	Siblings ≥ 2	Siblings ≥ 3	Siblings ≥ 4	Siblings ≥ 5
Panel A: Two Plus Sample				
$Twin_2$	0.291***	0.229***	0.135***	0.039***
	(0.005)	(0.012)	(0.011)	(0.007)
$Twin_3^*$	0.000	0.439***	0.185***	0.082***
	(0.008)	(0.007)	(0.013)	(0.010)
$Twin_4^*$	-0.006	0.004	0.532***	0.184***
	(0.008)	(0.017)	(0.008)	(0.017)
$Twin_5^*$	-0.001	-0.018	-0.003	0.660***
	(0.009)	(0.021)	(0.033)	(0.012)
Panel B: Three Plus Sample				
$Twin_3$		0.393***	0.196***	0.084***
		(0.004)	(0.009)	(0.007)
$Twin_4^*$		0.011	0.520***	0.184***
		(0.009)	(0.005)	(0.011)
$Twin_5^*$		-0.012	0.003	0.651***
		(0.010)	(0.020)	(0.007)
Panel C: Four Plus Sample				
$Twin_4$			0.479***	0.183***
			(0.004)	(0.009)
$Twin_5^*$			0.009	0.638***
			(0.012)	(0.005)

Notes: Each column reports the first stage estimate of fertility at each parity on twin births from the IV regressions displayed in Table 8. In each case we report the first stages for the baseline specification of the Non-Linear IV, although results are quantitatively similar in the case of the +S&H specification. Standard errors are clustered by family (three plus and four plus samples), or robust to heteroscedasticity when only one child from each family is included in the regressions (two plus sample).

TABLE A10. Assessing Bias with Covariate Adjustment – Splitting Sample by Maternal Health

	OLS			IV		
	Base	+H	+S&H	Base	+H	+S&H
Panel A: Developing Country Results						
Fertility (Less Healthy)	-0.134***	-0.112***	-0.073***	-0.030	-0.036*	-0.038**
	(0.002)	(0.002)	(0.002)	(0.021)	(0.019)	(0.018)
Observations	311,395	311,395	311,395	311,395	311,395	311,395
Fertility (More Healthy)	-0.142***	-0.121***	-0.076***	-0.026	-0.037**	-0.035**
	(0.002)	(0.002)	(0.002)	(0.018)	(0.017)	(0.016)
Observations	302,662	302,662	302,662	302,662	302,662	302,662
Panel B: US Results						
Dependent Variable = School Z-Score						
Fertility (Less Healthy)	-0.034***	-0.033***	-0.026***	-0.093*	-0.096*	-0.100*
	(0.008)	(0.008)	(0.008)	(0.056)	(0.055)	(0.055)
Observations	60,594	60,594	60,594	60,594	60,594	60,594
Fertility (More Healthy)	-0.021**	-0.022**	-0.018*	-0.035	-0.027	-0.033
	(0.010)	(0.010)	(0.010)	(0.100)	(0.100)	(0.100)
Observations	27,692	27,692	27,692	27,692	27,692	27,692
Dependent Variable = Excellent Health						
Fertility (Less Healthy)	-0.008***	-0.006**	-0.004	-0.000	-0.004	-0.005
	(0.003)	(0.003)	(0.003)	(0.025)	(0.025)	(0.024)
Observations	70,011	70,011	70,011	70,011	70,011	70,011
Fertility (More Healthy)	-0.003	-0.003	-0.002	-0.020	-0.019	-0.020
	(0.002)	(0.002)	(0.002)	(0.018)	(0.018)	(0.018)
Observations	33,109	33,109	33,109	33,109	33,109	33,109

Notes: OLS and IV results are shown for the pooled 2+, 3+ and 4+ samples, splitting samples by the health status of each mother. In the case of IV estimates, fertility is instrumented using the twin instruments with pooling procedure described in Angrist et al. (2010) and refinement discussed in Mogstad and Wiswall (2012). In the developing country sample, less and more healthy refers to mothers whose height is respectively below and above the country-level mean (calculated in each survey) given heterogeneity in educational attainment by countries. In the case of the US, more health refers to mothers who report being in excellent health, while less healthy refers to mothers who report any other health status. All other details follow OLS and IV estimates in Tables 5–7. *p<0.1; **p<0.05; ***p<0.01.

TABLE A11. Bounds Estimates of the Quantity–Quality Trade-off

	IV	Nevo and Rosen (2012) Imperfect IV Bounds		Conley et al. (2012) UCI: $\gamma \in [0, 2\hat{\gamma}]$		Conley et al. (2012) LTZ: $\mathcal{N}(\mu_\gamma, \sigma_\gamma^2)$	
	with Controls	Lower	Upper	Lower	Upper	Lower	Upper
Panel A: DHS (Education Z-Score)							
Two Plus	-0.0115	-0.0780	-0.0115	-0.0210	-0.0115	-0.0164	
	[-0.062;0.039]	[-0.0836	0.0394]	[-0.0719	0.0394]	[-0.0673	0.0346]
Three Plus	-0.0476	-0.0713	-0.0476	-0.0572	-0.0476	-0.0526	
	[-0.086;-0.009]	[-0.0761	-0.0091]	[-0.0957	-0.0091]	[-0.0867	-0.0185]
Four Plus	-0.0365	-0.0587	-0.0365	-0.0457	-0.0365	-0.0412	
	[-0.073;0.000]	[-0.0637	0.0004]	[-0.0826	0.0004]	[-0.0716	-0.0108]
Panel B: USA (Education Z-Score)							
Two Plus	-0.1012	-0.1012	-0.0311	-0.1188	-0.1012	-0.1095	
	[-0.219;0.017]	[-0.2192	-0.0153]	[-0.2368	0.0169]	[-0.2276	0.0086]
Three Plus	-0.0171	-0.0171	-0.0217	-0.0338	-0.0171	-0.0256	
	[-0.149;0.115]	[-0.1491	-0.0001]	[-0.1486	0.0977]	[-0.1405	0.0893]
Four Plus	-0.1415	-0.1415	-0.0230	-0.1567	-0.1415	-0.1493	
	[-0.432;0.149]	[-0.4324	0.0231]	[-0.4654	0.1670]	[-0.4582	0.1595]
USA (Excellent Health)							
Two Plus	0.0256	0.0256	0.0002	0.0089	0.0256	0.0172	
	[-0.016;0.067]	[-0.0158	0.0061]	[-0.0324	0.0669]	[-0.0244	0.0588]
Three Plus	-0.0568	-0.0568	-0.0118	-0.0732	-0.0568	-0.0653	
	[-0.120;0.006]	[-0.1199	-0.0026]	[-0.1214	-0.0088]	[-0.1136	-0.0170]
Four Plus	-0.0308	-0.0308	-0.0180	-0.0459	-0.0308	-0.0387	
	[-0.132;0.070]	[-0.1315	-0.0037]	[-0.1168	0.0400]	[-0.1097	0.0323]

Notes: This table presents upper and lower bounds of a 95% confidence interval for the effects of family size on (standardised) children's educational attainment and health (health in USA only). Nevo and Rosen (2012) bounds are presented in columns 2 and 3, and variants of Conley et al. (2012) bounds are presented in columns 4-7. the IV point estimate with full controls is displayed for comparison in column 1. Nevo and Rosen (2012) bounds are based on the assumption that selection into twinning and into fertility are oppositely signed (eg positive and negative), and twins are "less endogenous" than fertility. Conley et al. (2012) bounds are estimated as described in section 2.2 under various priors about the direct effect that being from a twin family has on educational outcomes (γ). In the UCI (union of confidence interval) approach, it is assumed the true $\gamma \in [0, 2\hat{\gamma}]$, while in the LTZ (local to zero) approach it is assumed that γ follows the empirical distribution estimated in each case. The preferred prior for γ ($\hat{\gamma}$) and its distribution is discussed in Bhalotra and Clarke (2016). Comparisons under a range of priors are presented in Figures 6 and A10-A12. Each estimate is based on the specifications with full controls from Tables 6 and 7.

Table A12. Results for Kitagawa (2015) Tests with Controls (DHS)

	Baseline	Socioeconomic	Socioeconomic plus Health
Kitagawa Test Statistic	14.559	15.963	16.558
Instrumental Validity (p-value)	0.028	0.224	0.462
Coefficient (IV model)	-0.013	-0.032	-0.042
	(0.073)	(0.068)	(0.068)
Observations	251,831	251,831	251,831

Notes: Results are presented for the Kitagawa (2015) test of instrumental validity. This test exists for a binary endogenous variable, and as such rather than estimate a model with fertility as the endogenous variable, we estimate a model with the binary variable “greater than 2 births” as the endogenous variable. The instrument considered is twinning at birth order 2. The estimation results of a typical IV model using this specification are presented and indicated as “IV model”. Instrumental validity can not be proven, but can be disproven, with low p-values being evidence against instrumental validity. The first row shows the value for the variance weighted test statistic proposed by Kitagawa (2015), and the second row displays the p-values associated with the Kitagawa test. Baseline controls consist of mother year of birth fixed effects, continent fixed effects, child sex, and decade of birth fixed effects. Socioeconomic controls add indicators for mother’s education (0 years, 1-6 years, 7-11 years, or 12+ years), and Health controls add indicators for overweight or underweight mothers, and whether the majority of births in the mother’s region were attended by doctors, nurses or unattended. A trimming constant of 0.07 is used for the instrumental validity test, (as laid out in Kitagawa (2015)), and 500 bootstrap replications are run to determine the p-value.

TABLE A13. Developing Country IV Estimates Using Same Sex Twins Only

	2+			3+			4+		
	Base	+H	+S&H	Base	+H	+S&H	Base	+H	+S&H
Panel A: First Stage									
Dependent Variable = Fertility									
Same Sex Twins	0.796***	0.800***	0.802***	0.804***	0.815***	0.810***	1.008***	0.999***	0.995***
	(0.048)	(0.047)	(0.045)	(0.045)	(0.045)	(0.044)	(0.053)	(0.051)	(0.050)
Kleibergen-Paap rk statistic	278.55	285.98	312.31	317.05	330.96	338.52	368.15	377.63	391.22
p-value of rk statistic	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Panel B: IV Results									
Dependent Variable = School Z-Score									
Total number of children in the family	-0.042	-0.048	-0.045	0.021	0.004	0.011	-0.018	-0.022	-0.012
	(0.052)	(0.051)	(0.049)	(0.041)	(0.039)	(0.038)	(0.034)	(0.032)	(0.029)
Observations	259,958	259,958	259,958	395,687	395,687	395,687	409,576	409,576	409,576
Coefficient Difference		0.499	0.858		0.071	0.512		0.623	0.642

Notes: Refer to notes to Table 6. This table follows identical specifications, however now only same sex twins are used as an instrument instead of all twins. In the DHS, 64.1% of twin pairs are of the same gender. Standard errors are clustered by mother. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

TABLE A14. US IV Estimates Using Same Sex Twins Only

	2+			3+			4+		
	Base	+H	+S&H	Base	+H	+S&H	Base	+H	+S&H
Panel A: First Stage									
Dependent Variable = Fertility (School Z-Score Second Stage)									
[1em] Same Sex Twins	0.719***	0.721***	0.722***	0.770***	0.770***	0.773***	0.838***	0.842***	0.850***
	(0.031)	(0.031)	(0.031)	(0.054)	(0.054)	(0.054)	(0.111)	(0.111)	(0.109)
Kleibergen-Paap rk statistic	531.08	534.95	550.13	201.03	203.41	206.54	57.08	57.73	60.53
p-value of rk statistic	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Dependent Variable = Fertility (Excellent Health Second Stage)									
[1em] Same Sex Twins	0.756***	0.758***	0.760***	0.782***	0.783***	0.786***	0.820***	0.828***	0.837***
	(0.030)	(0.030)	(0.029)	(0.050)	(0.050)	(0.050)	(0.105)	(0.104)	(0.103)
Kleibergen-Paap rk statistic	643.69	650.43	667.08	239.82	243.36	248.46	61.42	62.96	66.17
p-value of rk statistic	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Panel B: IV Results									
Dependent Variable = School Z-Score									
[1em] Fertility	-0.065	-0.064	-0.066	-0.014	-0.019	-0.023	0.094	0.088	0.084
	(0.061)	(0.060)	(0.060)	(0.082)	(0.082)	(0.083)	(0.119)	(0.116)	(0.116)
Observations	61,267	61,267	61,267	47,308	47,308	47,308	21,352	21,352	21,352
Dependent Variable = Excellent Health									
[1em] Fertility	-0.002	0.029	0.028	-0.024	-0.066*	-0.065*	0.075	0.001	-0.002
	(0.029)	(0.025)	(0.025)	(0.046)	(0.037)	(0.037)	(0.067)	(0.055)	(0.054)
Observations	70,277	70,277	70,277	53,393	53,393	53,393	24,358	24,358	24,358

Notes: Refer to notes in Table 7. This table follows identical specifications, however now only same sex twins are used as an instrument instead of all twins. In the NHIS, 66.0% of twin pairs are of the same gender. Standard errors are clustered by mother. *p<0.1; **p<0.05; ***p<0.01.

B. Data Definitions

All outcome and control variables used in principal IV and OLS analyses are described in the following table. As well as variable definitions, units and any functional forms are indicated, which refer to the way variables enter IV or OLS models.

TABLE B1. Variable Definitions

Variable	Definition
Panel A: DHS Data	
School Z-score	Z-score of years of schooling, standardised relative to country and year of birth cohort.
Male Child	Binary measure, one for boy, zero for girls
Country	Fixed effect for country of survey
Year of Birth	Fixed effect for year of birth
Child's Age	Fixed effect for child's age
Contraceptive Intent	Fixed effect for mother's use of contraceptive methods
Mother's Age	Fixed effect for mother's age at child birth
Mother's Age at First Birth	Inferred from age at survey time and age of child
Mother's Education	Fixed effect for total years of education achieved
Family Wealth	Fixed effect for DHS-assigned wealth quintile. Where not recorded a separate fixed effect for "no wealth quintile" is included
Mother's Height	Measured in centimetres
Mother's BMI	Measured in units (weight in kilograms divided by height in metres squared)
Prenatal Doctor Availability	Proportion of births in the same DHS cluster which received a prenatal check-up from a doctor
Prenatal Nurse Availability	Proportion of births in the same DHS cluster which received a prenatal check-up from a nurse
No Prenatal Care	Proportion of births in the same DHS cluster which received no prenatal check-ups from health professionals
Panel B: NHIS Data	
Education Z-Score	Z-score of grade progression, standardised relative to month and year of birth cohort
Excellent Health	Indicator of whether a child is classified by the family as being in "excellent health" (chosen from a categorical list)
Male Child	Binary measure, one for boy, zero for girls
Survey Year	Fixed effect for year NHIS wave was run
Child Age	Fixed effect for age at interview in months and years
Region	Fixed effect for census bureau region of residence
Mother's Race	Fixed effect for mother's race
Mother's Age	Fixed effect for mother's age in years
Mother's Age at First Birth	Inferred from age at survey time and age of child
Mother's Education	Fixed effects for mother's highest completed year of education
Mother's Health Status	Self-reported based on categorical list
Mother's Height	Mother's Height in Inches
Smoking Status	Binary variable indicating whether the mother smoked prior to pregnancy
Smoking Status Missing	Binary variable indicating no response to the mother's smoking status

C. Testing for Equality of Coefficients Between IV Models

When estimating subsequent IV models with the progressive inclusion of controls to capture maternal selection, our point is really that column 1 ("Base") is not distinguishable from 0, while column 3 ("+S&H") often is, as this is the important thing in considering the literature and in showing that partial bias adjustment recovers the trade-off. We have nevertheless added a formal test of coefficients between IV models in all IV tables. This is added as a row called "Coefficient Difference" at the bottom of Tables 6 and 7. This computation is not entirely trivial, as these tests must take account of correlations between variance-covariance matrices of each IV regression in the style of seemingly unrelated regression. Thus, we calculate these test statistics by jointly estimating the models with GMM (seemingly unrelated regression is a Feasible Generalised Least Squares technique, and hence not suitable for IV models). To do this we form two equations which are the two models we wish to compare in the following format:

$$y_{ij} = b_0 + b_1 \times fertility_j + \mathbf{baseline}'_{ij} \times \mathbf{b}_b \qquad \text{(C1)}$$

$$y_{ij} = c_0 + c_1 \times fertility_j + \mathbf{baseline}'_{ij} \times \mathbf{c}_b + \mathbf{health}'_{ij} \times \mathbf{c}_h. \qquad \text{(C2)}$$

Our goal is to test the equality of coefficients $b_1 = c_1$. Given that we are using instruments for endogenous $fertility$ in each case, we can thus form the following population moment conditions which hold under the null of instrumental validity in each case (ie, replicate the specifications we are estimating in the paper):

$$twin'_i(y_{ij} - b_0 - b_1 \times fertility_j - \mathbf{baseline}'_{ij} \times \mathbf{b}_b) = 0 \qquad \text{(C3)}$$

$$twin'_i(y_{ij} - c_0 - c_1 \times fertility_j - \mathbf{baseline}'_{ij} \times \mathbf{c}_b - \mathbf{health}'_{ij} \times \mathbf{c}_h) = 0. \qquad \text{(C4)}$$

Using the sample analogues of these moments, we can then estimate the parameters **b** and **c** via GMM. Denoting the two moments as the 2 element vector $g(\widehat{bc})$, we then estimate the parameters $\widehat{b}$ and $\widehat{c}$ using the GMM objective function $J(\widehat{bc}) = ng(\widehat{bc})'Wg(\widehat{bc})$. An unadjusted weight matrix is used which assumes that the moment conditions are independent, which replicates all parameters and standard errors from the original IV model, but now the estimates can be formally tested for equality against one-another using a χ^2 test which also considers the correlation between the observations in the two models when estimating the eventual variance-covariance matrix.

D. Bounds for the ATE using Monotone IV Assumptions

Manski and Pepper (2000) define monotone IV (MIV) bounds, where a monotone IV is a variable which fulfills a "mean montonocity" condition:

$$E[y(t)|v = u_2] \geq E[y(t)|v = u_1]$$

In this notation, v refers to a monotone IV, u_2 and u_1 to levels of the monotone IV, and $y(t)$ the outcome level given treatment t. To derive bounds with a MIV, Manski and Pepper (2000) start with "no assumptions bounds" of Manski (1989). Point identification is not possible due to the lack of observed counterfactual outcome $y(t)$ for those whose treatment is not equal to t. The bounds identification behind "no assumptions bounds" simply comes from assuming that this counterfactual outcome is bounded between some values $K_0 \leq y \leq K_1$. For example, in the case of a binary outcome variable, $K_0 = 0$ and $K_1 = 1$. Using a similar argument in the MIV setting, Manski and Pepper (2000, p. 1000) note that bounds on $y(t)$ with a monotone IV are given by:

$$\sum_{u \in V} P(v=u) \left\{ \sup_{u_1 \leq u} [E(y|v=u_1, z=t) \cdot P(z=t|v=u_1) + K_0 \cdot P(z \neq t|v=u_1)] \right\} \\ \leq E[y(t)] \leq \\ \sum_{u \in V} P(v=u) \left\{ \inf_{u_2 \geq u} [E(y|v=u_2, z=t) \cdot P(z=t|v=u_2) + K_1 \cdot P(z \neq t|v=u_2)] \right\}, \tag{D1}$$

where z refers to a treatment of interest (of level t), and u_1 and u_2 refer to the levels of the monotone IV. For example, following our notation, bounding an outcome of interest y for treatment level $fert = 3$ gives the following bounds on $y(fert = 3)$:

$$\sum_{u \in twin = \{0,1\}} P(twin = u) \times \\ \left\{ \sup_{u_1 \leq u} [E(y|twin = u_1, fert = 3) \cdot P(fert = 3|twin = u_1) + K_0 \cdot P(fert \neq 3|twin = u_1)] \right\} \\ \leq E[y(t)] \leq \\ \sum_{u \in twin = \{0,1\}} P(twin = u) \times \\ \left\{ \inf_{u_2 \geq u} [E(y|twin = u_2, fert = 3) \cdot P(fert = 3|twin = u_2) + K_1 \cdot P(fert \neq 3|twin = u_2)] \right\}.$$

Typically, an MIV assumption alone is non-informative (for example see Brinch et al. (2017)), and as such Manski and Pepper (2000) suggest MIV bounds should be estimated with the imposition of additional assumptions, namely Monotone Treatment Selection (MTS) or Monotone Treatment Response (MTR), where appropriate. However, in our case, both MTS and MTR are, strictly, inappropriate assumptions. MTR is equivalent to assuming that each child's level of human capital is weakly decreasing in conjectured fertility, while the MTS assumption states that parents who choose higher fertility have weakly lower child educational outcomes than those who choose lower fertility. As our application is interested in *estimating* the fertility–human capital trade-off, it is not appropriate to *assume* a sign in this relationship. Thus, in what follows we focus largely on bounds estimated only using a MIV assumption. Nevertheless, we also briefly discuss and present the MTS–MTR bounds, which are generated under assumptions implying that a fertility–human capital trade-off is present, and if this is the case, they provide an estimate of the lower bound on the trade-off. Indeed, we do note that while we likely prefer not to impose these assumptions given that our question of interest is in *determining* whether a human capital–fertility trade-off exists, we also note that our data are not inconsistent with MTR–MTS assumptions, as in all cases examined, child human capital is decreasing in family size which is consistent what one would observe if MTR–MTS assumptions were true.[1]

Below we present estimated bounds on average treatment effects of the impact of fertility on child outcomes. Note that Manski and Pepper (2000) bounds are non-parametric bounds, and so an ATE is calculated as:

$$\Delta(s,t) \equiv E[y(t)] - E[y(s)], \qquad \text{where } s < t.$$

As stated in Manski (2003, p. 148), the lower bound on $\Delta(s,t)$ can be calculated by subtracting the lower bound on $E[y(t)]$ from the upper bound on $E[y(s)]$, and vice versa for the upper bound. These bounds are sharp. Inference on lower and upper bounds is undertaken as described in Manski and Pepper (2000), using a standard percentile bootstrap. When calculating ATE bounds, we consider a series of changes in fertility, in each case examining movements of 1 unit, eg from 2 to 3 births, from 3 to 4 births, and so on, for each of the 2+, 3+ and 4+ samples of interest.

As is often the case where additional MTS or MTR assumptions can be invoked, these MIV bounds are extremely wide, and similar to “no assumption bounds”. For example, In the 2+ sample, the change in school Z-score based on a movement from 2 to 3 siblings is bounded between -3.38 and 1.96 standard deviations. Given that the twin instrument causes the largest change in the distribution at around the parity of birth, bounds for fertility movements at higher birth parities are even wider. In no case are we able to estimate informative bounds in this setting. This is observed when using the DHS data (Table D1) and when using NHIS data from the US (Table D2).

For comparison, we also present the MTS–MTR bounds on the ATE. As documented in Manski and Pepper (2000, Proposition 2, Corollary 2), if we are additionally willing to make these assumptions, the non-parametric bounds on $E[y(t)]$ become:

$$\begin{aligned} &\sum_{u<t} E(y|z=u)\cdot P(z=u) + E(y|z=t)\cdot P(z\geq t) \\ &\qquad\qquad \leq E[y(t)] \leq \\ &\sum_{u>t} E(y|z=u)\cdot P(z=u) + E(y|z=t)\cdot P(z\leq t), \end{aligned} \tag{D2}$$

1. In Manski and Pepper’s notation, this is that $E(y|z=u)$ is weakly decresing in u, or applied to our case, that $E(\text{child human capital}|fertility=u)$ is weakly decreasing in u where u refers to family size.

TABLE D1. Manski and Pepper Monotone IV Bounds on Average Treatment Effect, DHS Data

		Lower Bound on $\Delta(s,t)$			Upper Bound on $\Delta(s,t)$		
s	t	0.025 Bootstrap Quantile	Bound Estimate	0.975 Bootstrap Quantile	0.025 Bootstrap Quantile	Bound Estimate	0.975 Bootstrap Quantile
		Two-Plus Sample, Montone IV: Twin at Birth 2					
2	3	-4.372	-4.366	-4.359	4.170	4.230	4.277
3	4	-4.557	-4.502	-4.446	4.281	4.336	4.385
4	5	-4.965	-4.913	-4.858	4.780	4.830	4.879
5	6	-5.298	-5.251	-5.203	5.300	5.337	5.377
		Three-Plus, Montone IV: Twin at Birth 3					
3	4	-4.206	-4.201	-4.196	4.083	4.088	4.093
4	5	-4.676	-4.672	-4.666	4.551	4.601	4.620
5	6	-5.160	-5.143	-5.098	5.126	5.142	5.146
		Four-Plus, Montone IV: Twin at Birth 4					
4	5	-4.120	-4.108	-4.036	3.960	4.029	4.033
5	6	-4.678	-4.675	-4.606	4.641	4.646	4.651

Notes: ATE bounds are presented following Manski and Pepper (2000), under *only* a monotone IV assumption. In each sample, the twin birth instrument is indicated, and our Monotone IV assumption is that investments in children are weakly higher among women with twin births than those with singleton births, given positive selection of women into twin births. Here the outcome is a school z-score, and we assume $K_0 = -3$ and $K_1 = 3$, each of which are extreme values in the outcome distribution. Upper and lower bounds estimates are provided, along with 95% confidence intervals on these bounds. These are calculated with percentile bootstrap, with 500 bootstrap replications.

where the notation is as in (D1). For example, consider once again the case where we wish to estimate the average outcome for children with a family size of 3:

$$\sum_{u<3} E(y|fert = u) \cdot P(fert = u) + E(y|fert = 3) \cdot P(fert \geq 3) \\ \leq E[y(t)] \leq \\ \sum_{u>3} E(y|fert = u) \cdot P(fert = u) + E(y|fert = 3) \cdot P(fert \leq 3). \tag{D3}$$

Note that in the above, it is clear that we are no longer using twins to derive partial identification. Partial identification is driven by selection assumptions, namely that human capital is weakly decreasing in conjectured fertility (MTR), and that parents who choose higher fertility have weakly lower child educational outcomes (MTS). In this case the upper bound on the ATE $\Delta(s,t)$ is now zero, given that a trade-off is conjectured to exist. We provide estimates of both the lower bound and it's confidence interval in Tables D3 (for DHS) and D4 (for USA). In this case lower bound estimates are much tighter, at around -0.05 to -0.2 standard deviations for child educational attainment, and typically close to lower bounds estimated on LATEs from the twin instrument documented in the body of the paper.

TABLE D2. Manski and Pepper Monotone IV Bounds on Average Treatment Effect, USA Data

		Lower Bound on $\Delta(s,t)$			Upper Bound on $\Delta(s,t)$		
s	t	0.025 Bootstrap Quantile	Bound Estimate	0.975 Bootstrap Quantile	0.025 Bootstrap Quantile	Bound Estimate	0.975 Bootstrap Quantile
PANEL A: DEPENDENT VARIABLE = SCHOOL Z-SCORE							
		Two-Plus Sample, Montone IV: Twin at Birth 2					
2	3	-3.397	-3.387	-3.377	1.849	1.966	2.080
3	4	-3.643	-3.541	-3.436	4.526	4.621	4.703
4	5	-5.436	-5.352	-5.261	5.588	5.631	5.663
5	6	-5.898	-5.865	-5.824	5.880	5.901	5.905
		Three-Plus, Monotone IV: Twin at Birth 3					
3	4	-3.141	-3.024	-2.919	3.270	3.286	3.297
4	5	-5.147	-5.135	-5.117	5.128	5.143	5.154
5	6	-5.749	-5.740	-5.733	5.743	5.750	5.758
		Four-Plus, Monotone IV: Twin at Birth 4					
4	5	-3.081	-2.858	-2.632	3.273	3.313	3.332
5	6	-5.213	-5.194	-5.132	5.196	5.223	5.240
PANEL B: DEPENDENT VARIABLE = EXCELLENT HEALTH							
		Two-Plus Sample, Montone IV: Twin at Birth 2					
2	3	-0.594	-0.590	-0.586	0.322	0.354	0.384
3	4	-0.681	-0.649	-0.620	0.748	0.767	0.786
4	5	-0.917	-0.897	-0.877	0.929	0.938	0.943
5	6	-0.983	-0.978	-0.969	0.980	0.983	0.984
		Three-Plus, Monotone IV: Twin at Birth 3					
3	4	-0.567	-0.563	-0.534	0.527	0.533	0.537
4	5	-0.863	-0.860	-0.852	0.850	0.853	0.856
5	6	-0.959	-0.957	-0.956	0.956	0.958	0.960
		Four-Plus, Monotone IV: Twin at Birth 4					
4	5	-0.561	-0.497	-0.429	0.522	0.539	0.546
5	6	-0.871	-0.867	-0.851	0.865	0.869	0.873

Notes: ATE bounds are presented following Manski and Pepper (2000), under *only* a monotone IV assumption. In each sample, the twin birth instrument is indicated, and our Monotone IV assumption is that investments in children are weakly higher among women with twin births than those with singleton births, given positive selection of women into twin births. Here the outcome in panel A is a school z-score, and we assume $K_0 = -3$ and $K_1 = 3$, each of which are extreme values in the outcome distribution. In panel B, the outcome is a binary variable for “excellent health”, and so we assume $K_0 = 0$ and $K_1 = 1$. Upper and lower bounds estimates are provided, along with 95% confidence intervals on these bounds. These are calculated with percentile bootstrap, with 500 bootstrap replications.

TABLE D3. Manski and Pepper MTS-MTR Lower Bounds on Average Treatment Effect, DHS Data

		Lower Bound on $\Delta(s,t)$	
s	t	0.025 Bootstrap Quantile	Bound Estimate
Two-Plus Sample, MTS–MTR Bounds			
2	3	-0.121	-0.118
3	4	-0.090	-0.086
4	5	-0.145	-0.139
5	6	-0.226	-0.218
Three-Plus, MTS–MTR Bounds			
3	4	-0.089	-0.086
4	5	-0.093	-0.090
5	6	-0.149	-0.144
Four-Plus, MTS–MTR Bounds			
4	5	-0.069	-0.067
5	6	-0.079	-0.076

Notes: ATE bounds are presented following Manski and Pepper (2000), under the MTS–MTR assumption. These bounds conjecture that a fertility–human capital trade-off does exist, and as such, the upper bound is no larger than 0. We present bounds for each of the parity-specific samples, providing the MTS–MTR lower bound, along with the confidence interval on the lower bound. These are calculated with a percentile bootstrap, with 500 bootstrap replications.

TABLE D4. Manski and Pepper MTS-MTR Lower Bounds on Average Treatment Effect, USA Data

		Lower Bound on $\Delta(s,t)$	
s	t	0.025 Bootstrap Quantile	Bound Estimate
PANEL A: DEPENDENT VARIABLE = SCHOOL Z-SCORE			
Two-Plus Sample, MTS–MTR Bounds			
2	3	-0.013	-0.010
3	4	-0.032	-0.020
4	5	-0.087	-0.063
5	6	-0.187	-0.140
Three-Plus, MTS–MTR Bounds			
3	4	-0.011	-0.007
4	5	-0.042	-0.028
5	6	-0.126	-0.091
Four-Plus, MTS–MTR Bounds			
4	5	-0.008	-0.002
5	6	-0.070	-0.045
PANEL B: DEPENDENT VARIABLE = EXCELLENT HEALTH			
Two-Plus Sample, MTS–MTR Bounds			
2	3	-0.006	-0.004
3	4	-0.026	-0.020
4	5	-0.049	-0.038
5	6	-0.083	-0.062
		Three-Plus, MTS–MTR Bounds	
3	4	-0.006	-0.004
4	5	-0.020	-0.013
5	6	-0.053	-0.038
		Four-Plus, MTS–MTR Bounds	
4	5	-0.007	-0.004
5	6	-0.033	-0.022

Notes: Notes: ATE bounds are presented following Manski and Pepper (2000), under the MTS–MTR assumption. These bounds conjecture that a fertility–human capital trade-off does exist, and as such, the upper bound is no larger than 0. We present bounds for each of the parity-specific samples and for each outcome variable of interest in US data (a school z-score and an excellent health indicator), providing the MTS–MTR lower bound, along with the confidence interval on the lower bound. These are calculated with a percentile bootstrap, with 500 bootstrap replications.

References

Angrist, Joshua, Victor Lavy, and Analia Schlosser (2010). "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." *Journal of Labor Economics*, 28(4), 773–824.

Bhalotra, Sonia and Damian Clarke (2018). "Twin Birth and Maternal Condition." *Review of Economics and Statistics*, 0(ja), 1–45.

Bhalotra, Sonia R. and Damian Clarke (2016). "The Twin Instrument." IZA Discussion Papers 10405, Institute for the Study of Labor (IZA).

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2005). "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *The Quarterly Journal of Economics*, 120(2), 669–700.

Brinch, Christian N., Magne Mogstad, and Matthew Wiswall (2017). "Beyond LATE with a Discrete Instrument." *Journal of Political Economy*, 125(4), 985–1039.

Cáceres-Delpiano, Julio (2006). "The Impacts of Family Size on Investment in Child Quality." *Journal of Human Resources*, 41(4), 738–754.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi (2012). "Plausibly Exogenous." *The Review of Economics and Statistics*, 94(1), 260–272.

Fitzsimons, Emla and Bansi Malde (2014). "Empirically probing the quantity-quality model." *Journal of Population Economics*, 27(1), 33–68.

Kitagawa, Toru (2015). "A Test for Instrument Validity." *Econometrica*, 83(5), 2043–2063.

Li, Hongbin, Junsen Zhang, and Yi Zhu (2008). "The quantity-Quality trade-Off of children In a developing country: Identification using Chinese twins." *Demography*, 45, 223–243.

Manski, Charles F. (1989). "Anatomy of the Selection Problem." *Journal of Human Resources*, 24(3), 343–360.

Manski, Charles F. (2003). *Partial Identification of Probability Distributions*. Springer, New York, USA.

Manski, Charles F. and John V. Pepper (2000). "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica*, 68(4), 997–1010.

Mogstad, Magne and Matthew Wiswall (2012). "Instrumental variables estimation with partially missing instruments." *Economics Letters*, 114, 186–189.

Nevo, Aviv and Adam M. Rosen (2012). "Identification with Imperfect Instruments." *The Review of Economics and Statistics*, 94(3), 659–671.

Ponczek, Vladimir and André Portela Souza (2012). "New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country." *Journal of Human Resources*, 47(1), 64–106.

Rosenzweig, Mark R. and Junsen Zhang (2009). "Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's One-Child Policy." *Review of Economic Studies*, 76(3), 1149–1174.