Testing the trade-off between family size and human capital of children: Evidence from census data for five Latin American countries *

Lucila Berniell[†] Cristian Bonavida[‡] Dolores de la Mata[§]

August 31, 2023

Abstract

We study the relationship between family size and children's human capital, commonly referred to as the "quantity-quality trade-off" following Becker and Lewis (1973), in Latin American countries. Despite a history of elevated fertility rates in this region and a rapid convergence to the fertility levels in high-income countries, little is known on whether family size negatively affects investments in children's human capital. The evidence for other regions have challenged the existence of such trade-off. We use 10% public-use micro-data samples from the census of five Latin American countries with relatively high fertility rates. We employ a twin instrumental variable approach to estimate the causal effect of family size on educational attainment of children born prior to the twin birth. We shed light on the potential mechanisms affecting the relationship between family size and children's human capital outcomes when: i) financial restrictions are binding, ii) returns to education are low, iii) parental investments decisions are affected by gender norms; iv) formal education supply is scarce. Our results indicate the existence of a quantity-quality trade-off that is related to households' financial restrictions, gender norms and the provision of school services at the district of residence.

Keywords: human capital investments, fertility, Latin America, twins-instrument.
JEL Classification: J13, I21, J24.

*We are grateful to Iván Albina for helpful discussions regarding the analysis of IPUMS-I data. Views expressed here do not necessarily correspond to those of our affiliation.

[†]CAF-Development Bank of Latin America, Research Department.
[‡]CAF-Development Bank of Latin America, Research Department.
[§]CAF-Development Bank of Latin America, Research Department.

1 Introduction

Understanding the interplay between family size and children's human capital accumulation has been an active area of research in social sciences. Correlational evidence shows that children in larger families tend to have on average worse human capital outcomes. Regarding the theoretical analysis, the seminal works of Becker and Lewis (1973) and Becker and Tomes (1976) argued that the number of children and the level of parental investments are closely related, as they both stem from constrained household choices. This framework implies a trade-off between fertility and child human capital, usually referred to as the *quantity-quality trade-off*. However, empirical evidence have challenged the role of family size as a relevant input for human capital formation.

In this paper, we estimate the relationship between fertility and human capital formation of children in Latin America. Despite the region having experienced high fertility rates for many decades, along with a rapid convergence towards the fertility rates of high-income countries, there is very scarce evidence on how family size affects human capital formation. Interestingly for developing regions, where fertility rates were much higher in poorer families but they have been persistently declining decade by decade, if the quantity-quality trade-off were to operate, the smaller family sizes in poorer families would contribute to closing the large socioeconomic gaps in human capital formation that are pervasive in developing countries like those in Latin America (De La Mata et al., 2022).

We use 10% public-use micro-data samples from the census of five Latin American countries that covers (Colombia 2005, Dominican Republic 2002, 2010, Ecuador 2010, El Salvador 1992, 2007, and Nicaragua 1995, 2005). We employ multiple births as an instrument for family size and estimate its causal effects on educational attainment of children born prior to the twin birth. More specifically, in our twins instrumental variable (IV) approach, we look at the outcomes of first-, and first- and second-born children, respectively, in families of two or more, and three or more children, using the birth of twins at the second birth and third birth order as the instrumental variable. We address the possible selection of healthy mothers into twinning controlling for mother's socioeconomic characteristics and indicators of their health, as suggested by Bhalotra and Clarke (2020).

The richness of our data and the large sample sizes enable us to analyze heterogeneous effects of family size on children education outcomes based on parents' socioeconomic characteristics (proxied by education), child characteristics (gender), and place of residence characteristics (population and formal education supply, measured as the ratio teachers to school age children). In particular, with this analysis we intend to shed light on the potential mechanisms affecting the relationship between family size and children's human capital outcomes when: i) financial restrictions are more binding (i.e., for larger and lower socioeconomic status families) ii) returns to education are lower (i.e., in less populated areas), iii) parental investments decisions are affected by gender norms (i.e., different investments on females versus males children); iv) formal education supply is lower (i.e. in districts with lower ratio teachers to school-age children).

The use of the twins methodology is longstanding in the empirical quantity-quality literature (Rosenzweig and Wolpin, 1980; Black et al., 2005; Cáceres-Delpiano, 2006; Angrist et al., 2010). The results of this literature are not conclusive about the existence of a negative trade-off and results vary over time, across regions and sub-populations, across birth order and across the exact outcome of interest studied. In developed countries, most studies tend to find no effects on children education attainment (Black et al., 2005; Cáceres-Delpiano, 2006; Angrist et al., 2010; Åslund and Grönqvist, 2010; Fitzsimons and Malde, 2014), with the exception of Bhalotra and Clarke (2020), who find a trade-off for US data.[1] Studies in developing countries do not abound, but some of them even indicate the existence of a positive relation: Marteleto and de Souza (2012) uses data from Brazil and finds that the causal effect of family size on adolescents' schooling resembles a gradient that ranges from positive to no effect, trending to negative;[2] Alidou and Verpoorten (2019), using data for Sub-Saharan

[1] Black et al. (2010) also find with data from Norway that family size negatively affects IQ.

[2] The effect is positive in periods and regions in the earlier stages of socioeconomic development and with high fertility; but the effect disappears for recent periods when the opportunities for child farm work have declined, education has expanded, and fertility has declined to below-replacement levels.

countries, find no relation between family size on schooling on average and a positive effect in a sample of families with three or more children. On the contrary, Bhalotra and Clarke (2020) using a sample covering 68 developing countries find a null quantity-quality trade-off using the usual control variables in IV twins regressions while a negative quantity-quality relation emerges once selection into twinning—which has been shown to be related to mothers' health by Bhalotra and Clarke (2019)—is accounted for by controlling for variables related to mother's health.

Our IV estimates show that a fertility shock negatively affects first and second born children's years of schooling and their likelihood of completing primary education. The probability of attending formal education at the moment of the census and the probability of completing secondary education are not affected by family size. These results point to family size having primarily an effect in early stages of the education trajectory, affecting the outcome of years of schooling mainly through its effects on primary education completion. Interestingly, the negative effects are only statistically significant for the second-born children in families of 3 or more children. For the second-born, the probability of attending formal education at the time of the census is also negatively affected by the fertility shock.

Parental decisions regarding human capital investment tend to differ depending on the gender of the older children upon the occurrence of the fertility shock: our results show that female children are less affected than male children. Even more, female children in families of 2 or more children tend to be positively affected by the fertility shock. Gender norms may be guiding these results: to overcome financial restrictions male children maybe required to work outside home for a pay job increasing their likelihood of dropping out from school. Indeed, our IV results are in line with this hypothesis. We find that larger families increase male children probabilities to be in the labor force while female children are not affected.

The evidence we provide also indicates that the fertility shock mainly affects families with the tightest financial constraints, as the negative effect is observed when the shock occurs in the third birth and not the 2nd birth (i.e, in already larger families), and in the sample of families of lower socioeconomic status (proxied by parents education). For smaller families and of higher socioeconomic status, the fertility shock has null or even positive effects in some outcomes. If fertility rates of lower educated mothers continue to converge towards that of more educated mothers, our results would imply that the family size channel may contribute to close the high children's human capital socioeconomic gaps that are observed across all Latin American countries.

We also study the heterogeneous effects of family size by district characteristics, since the richness of the census data permits to identify around 750 of these subnational districts in the five countries included in the analysis. This feature of the data also allows to construct representative measures of district characteristics that are of high interest for the study of some of the proposed mechanisms. We find that the negative effects of family size on educational outcomes tend to be concentrated in districts with an intermediate level of formal education supply (measured by the ratio of teachers to school-age children). This result is consistent with an inverted U-shaped effect with respect to the provision of education services, where in districts with high supply restrictions both large and small families face limitations to school access while in districts with many teachers both types of families are not constrained in this dimension. Additionally, we do not find a clear pattern when we estimate heterogeneous effects by population sizes of the districts where children reside.

Finally, we estimate our regressions separately by country (except for Nicaragua and El Salvador, which we need to pool together to maintain reasonable sample sizes). Overall, the sign of the IV estimates for each outcome and country are similar to that of the pooled estimations, although not statistically significant. However, we do find large and statistically significant negative effects of family size on children's education outcomes among low socioeconomic status families in all countries, with the only exception of Dominican Republic.

The remainder of the paper is organized in the following manner. Section 2 provides an overview of the data employed in the analysis, outlines the sample selection process, and engages in a discussion about the observed trends in fertility rates within the countries under examination, thus providing justification for their inclusion in the study. Section 3 presents the identification strategy employed in the analysis, showing how it is an appropriate approach to establish the causal relationship of interest. Section 4 presents the main results, while Section 5 explores the heterogeneity of these results along a discussion of the potential mechanisms explaining the results we find. Section 6 describe the country-specific results, and Section 7 concludes.

2 Data

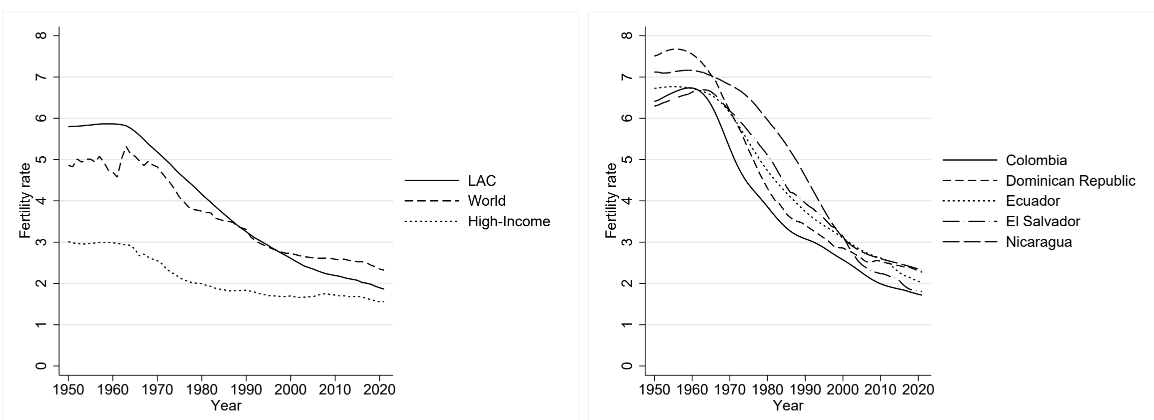
We use data from the IPUMS International (IPUMS-I) datasets, which consist of a collection of 10% public-use microdata samples from censuses conducted in different countries and harmonized across various years. Specifically, our sample comprises the census for Colombia (2005), Ecuador (2010), El Salvador (1992, 2007), Nicaragua (1995, 2005) and Dominican Republic (2002, 2010).[3]

2.1 Fertility rates in the selected LAC countries

Latin American and Caribbean countries are experiencing an important demographic transition, marked by a very rapid decline in fertility (Figure 1a). While the fertility rate per woman almost reached 6 children per woman—2 times larger than in high-income countries— in the 1960s, this rate plummeted to 2 in the decade of 2010 according to the United Nations estimates (UN, 2022), nearly closing the gap with high-income countries. This decline has been even more marked in Colombia, Dominican Republic, Ecuador El Salvador, and Nicaragua (Figure 1b). However, except for Colombia, the countries considered in this study have fertility rates above the regional average.

The decline in fertility rates estimated by the UN are also observed in census data. Graph 2 shows the evolution of the total number of children born to women by their year of birth in each country (Figure A) and by year of birth and education attainment (Figure B), for women who were between 40 and 60 years old in each census.[4] Clearly, family sizes are much larger for the less educated women, although the declining fertility occurs across all education levels.

Figure 1: Fertility rate



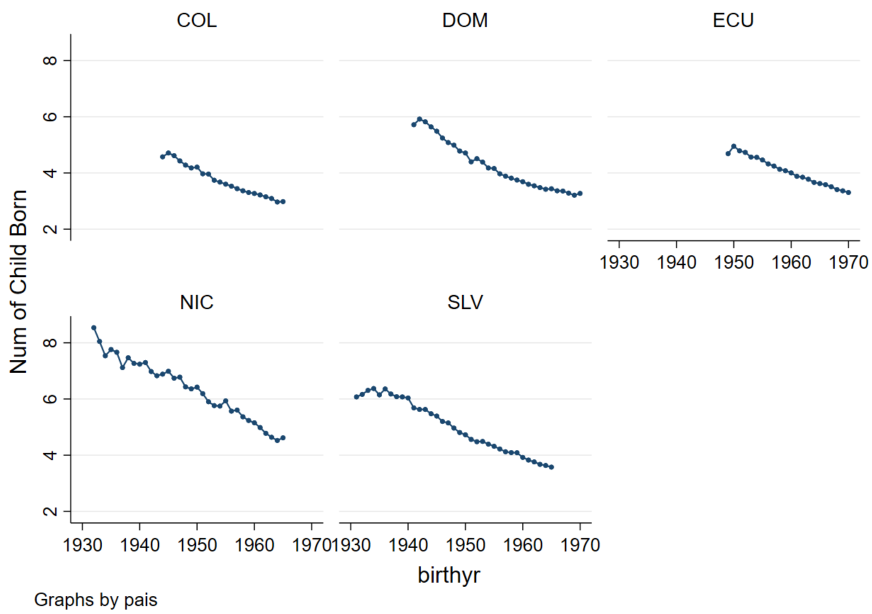
(a) Regions

(b) LAC countries

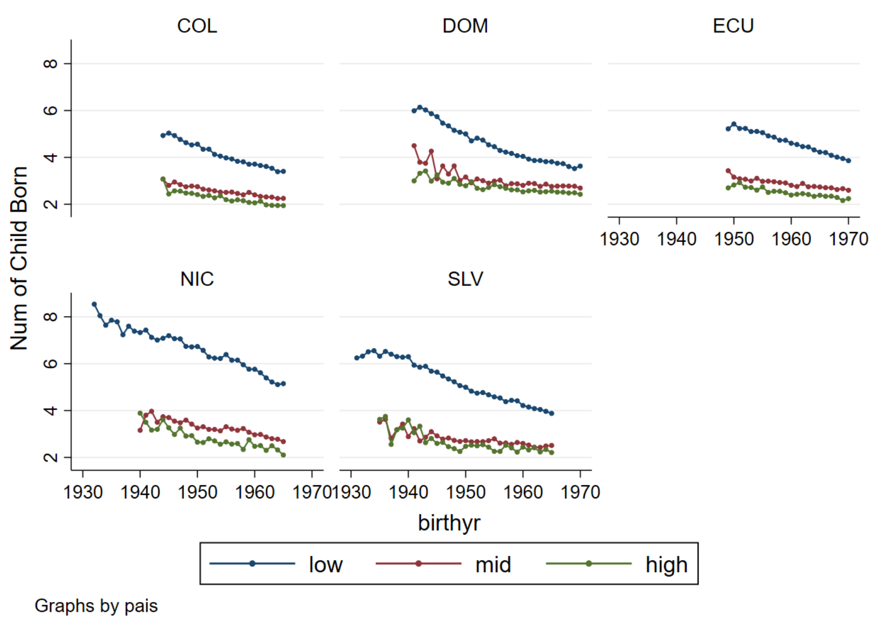
[3]Unfortunately, it is not possible to include all Latin American countries for which IPUMS-I has harmonized censuses available. Our analysis requires some key variables to identify twins (for example, childbirth month to construct a reliable tag for siblings who, having the same age, are in fact a twin pair) or to control in our IV specifications.

[4]At those ages most women reach the end of their reproductive life.

Figure 2: Number of surviving children born to women between 40 and 60 years old at census date



(a) By country



(b) By education level

Note: The graphs show the total number of children born to women between 40 and 60 years old at the time of the census (Colombia 2005, Ecuador 2010, El Salvador 1992 and 2007, Nicaragua 1995 and 2005, and Dominican Republic 2002 and 2010) by country (panel A) and by year of birth and educational attainment (Panel B). “Low” education refers to at most primary education complete or lower; “Medium” education refers to secondary education complete; “High” education refers to university complete.

Table 1: Fertility rate by decades

Country/region	1960	1970	1980	1990	2000	2010	Var % 2010-1960
Colombia	6.29	4.53	**3.43**	**2.87**	2.30	1.87	-70.18
Dominican Republic	7.06	5.34	**3.80**	**3.13**	**2.66**	2.44	-65.44
Ecuador	6.55	5.50	4.26	**3.44**	**2.85**	**2.33**	-64.35
El Salvador	6.57	**5.72**	**4.50**	**3.61**	2.59	2.05	-68.75
Nicaragua	7.04	**6.50**	**5.40**	**3.88**	2.82	2.49	-64.60
Latin America and the Caribbean (UN)	5.66	4.72	3.73	2.95	2.38	2.07	-63.32
High-income countries	2.82	2.22	1.88	1.74	1.70	1.67	-40.89
World	4.95	4.22	3.55	2.94	2.65	2.53	-48.93

Note: The fertility rate refers to the average number of live births a hypothetical cohort of women would have at the end of their reproductive period if they were subject during their whole lives to the fertility rates of a given period and if they were not subject to mortality. It is expressed as live births per woman. Source: United Nations, World Population Prospects (2022)

2.2 Sample selection

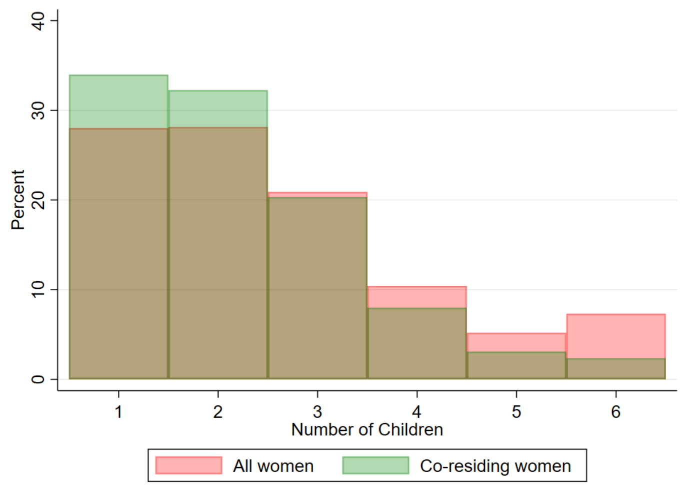
Our sample consists of 6- to 18-year-old first or second-born children living with their biological mother at the time of the census in families with two or more births. We restrict the analysis to mothers of childbearing age (15 to 45 years old). The father may or may not be present, but the presence of the biological mother is required to identify the total number of siblings of the same mother living in the household. Since the links between children and their biological parents within the household are typically unavailable in the censuses, we rely on IPUMS-I "Constructed Family Interrelationship Variables" which identify these links using a well-proven and carefully probabilistic methodology (Ruggles et al., 2015). We restrict the analysis to families where the total number of living children born to the mother coincides with the total number of children co-residing with the mother at the time of the census. This guarantees that we observe the first- and second-born children of the mother.[5]

Fertility (the total number of children) is the sum of children living in the household linked to the same biological mother. Among them, we classify as twins those individuals who have the same age or were born in the same date (month-year), resulting in 6383 twins in 667859 families included in the sample. The order of birth of each child is defined according to children's age.

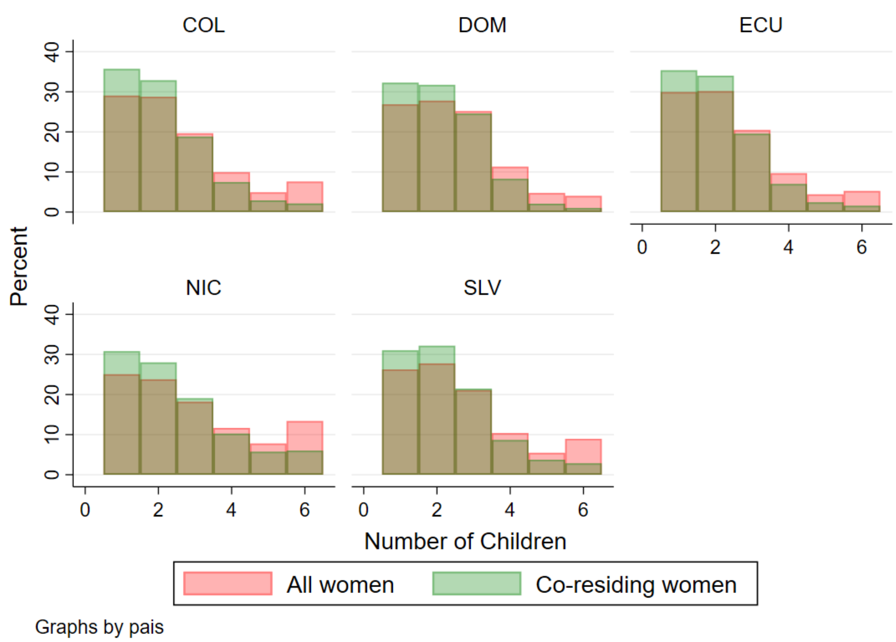
Among families where the mother is between 15 and 45 years old in a given census, those with two or more children represent 72% of all families. As fertility rates were high in the countries and periods we analyze, our sample selection does not imply leaving out a large share of families. Restricting further to the cases in which all children co-reside with the mother, we are left with 66% of all families (Figure 3a). Figure 4a shows the fraction of women that co-resides with all their children by age of the mother. Considering all women, the co-residence rate is high and stable around 75% for women between 20 and 35 years old, and then, as expected, starts to decrease (older mothers also have older children, who are more likely to leave the home). Co-residence rates are relatively lower for women below 20 years old, but specially for those who have three or more children. Figure 4b shows that the average number of children increases with the age of the mother. This pattern partially reflects the declining fertility rate that is observed in the region, but it is also due to the fact that younger mothers are further away of reaching the end of their reproductive life (i.e., number of children is a truncated variable). The figure shows that the average number of children from mothers co-residing with all their children is very similar to the average for all mothers between 15 and 32 years of age. However, the average number of children for those women above 32 years old that co-reside with all their children is lower than the average corresponding to all women.

[5] We cannot observe those children who are not living with their mother. In families where the number of children living with the mother does not coincide, the oldest children not necessarily are the first- and second-born children.

Figure 3: Distribution of the number of children born to women between 15 and 45 years old at census date, by co-residence status



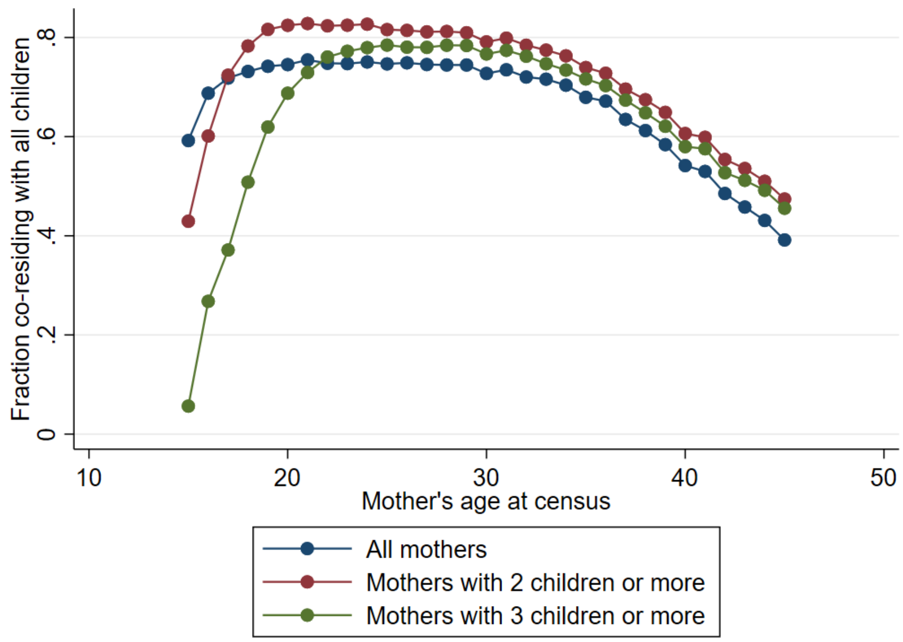
(a) Pool



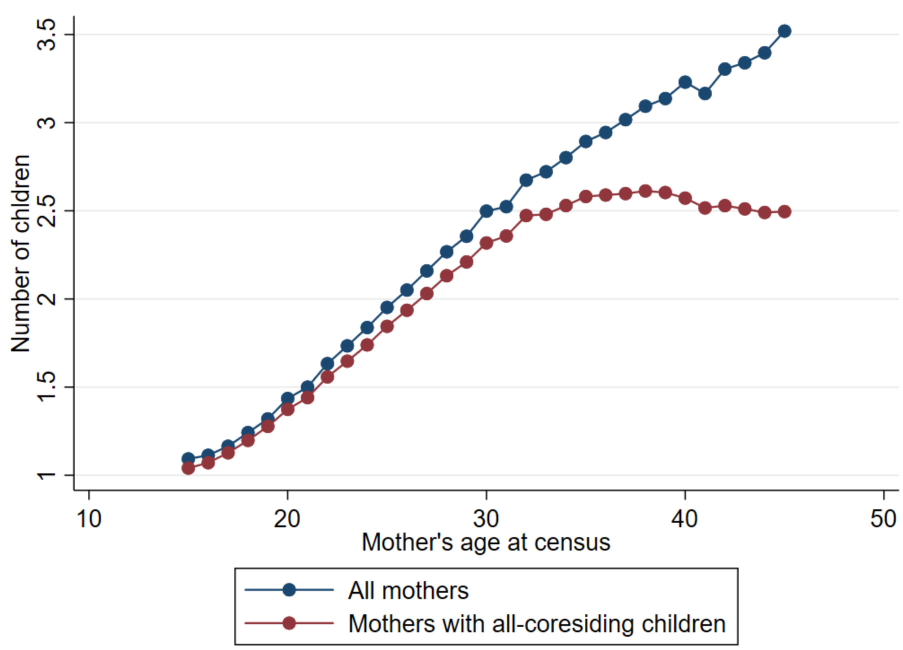
(b) By country

The graph shows the distribution of the number of surviving children born to women between 15 and 45 years old at the census date (Colombia 2005, Ecuador 2010, El Salvador 1992 and 2007, Nicaragua 1995 and 2005, and Dominican Republic 2002 and 2010). Number of children above 6 are collapsed at the value 6. Co-residing families correspond to those where all children live with the mother.

Figure 4: Cor-residence rate and number of children born to women between 15 and 45 years old at census date



(a) Fraction of mother's living with all their children



(b) Number of children by co-residence status

The graphs show the average number of surviving children born to women between 15 and 45 years old at the census date (Colombia 2005, Ecuador 2010, El Salvador 1992 and 2007, Nicaragua 1995 and 2005, and Dominican Republic 2002 and 2010).

2.3 Estimation samples

Following the literature of the twins IV approach (Angrist et al., 2010), we split the full sample into two subsamples. A first subsample, which we call "2+ sample", focuses on the first non-twin born child in families with 2 or more births. A second subsample, which we call "3+ sample", focuses on the first and second non-twin born children in families with 3 or more births. Table 2 describes these samples.[6]

As human capital outcome variables of children we consider age-standardized years of schooling and a dummy variable that indicates whether they are enrolled in formal education at the time of the census. We focus in these two measures since we restrict the analysis to children of school-age, that is, they are still in the process of acquiring education. The years of schooling are standardized (z-scores) within country-year and age of the children. Table 14 in the Appendix shows the mean and standard deviation of years of schooling in the whole population by country, year and age of children. Additionally, we consider whether children have completed primary or secondary education, but only for subsamples of those with ages above the country-specific cutoffs at which children are expected to have finished these levels of education.

Parents' and households' characteristics include the mother's age at the time of the census and at the time of the first birth, the parents' education, population size of the district of residence, and a dummy variable that indicates whether a child ever born to the women has died. This variable results from the difference between the reported number of children ever born to each woman and the number of children born who were alive at the time of the census. Although the age and moment at which born children died are not available in the census, most child deaths typically occur during the first years of life. Since child mortality is negatively associated with maternal health (Bhalotra and Rawlings, 2011), it is expected that healthier mothers in our dataset are less likely to have experienced the loss of a child. Therefore, we use this child mortality variable as an indirect measure of maternal health.

3 Empirical strategy

We first examine the relation between education outcomes of the first-born or first- and second-born children and family size using ordinary least squares (OLS). Then, we use twinning at the second or third births to instrument the number of children in the 2+ sample and the 3+ sample, respectively.

Specifically, our OLS specification in the sample of families with n or more children is the following:

$$y_{ihjd} = \alpha + \beta_1 Nchild_{ih} + \beta_2 X_i + \beta_3 X^M_{ih} + \beta_3 X_d + \gamma_j + u_i, \tag{1}$$

where y_{ihjd} represents the outcome of interest of child i in family h, in district d and country-year j. $Nchild$ is equal to the total number of children in the family (also referred to as fertility). X_i is a set of characteristics for the child i (gender, age, birth order, age difference with the nth sibling), X_d are characteristics of the district of residence (population), X^M_{ih} are mothers' characteristics (age at census, age at first birth, education, and a proxy of her health), and γ_j are country-year of census fixed effects.

The parameter of interest β_1 will be biased in the OLS estimation as potential non observed third factors correlated with fertility decisions may also affect families' investments in child human capital formation. The argument for using twins as an instrumental variable for fertility is that the birth of twins results in an increase in family size that is out of parents' control. However, a recent paper by Bhalotra and Clarke (2019) shows that twin births are not as random as they were presumed to be. The authors find that the distribution of twins in the population is skewed in favour of healthier women with healthier behaviors, both in developed and developing countries. They also found a positive association of twin births with the mother's education, which is coherent with the fact that education facilitates the access and the uptake of health-related information and practices. We follow Bhalotra and Clarke (2020), and include a set of mother's characteristics X^M_h, including her education and the

[6] Table 12 in the Appendix shows descriptive statistics only for children in families that have twins in the nth birth.

Table 2: Sample descriptive statistics

	2+ Sample	3+ Sample	
	first-born	**first-born**	**second-born**
Child characteristics			
Age	12.56	13.29	12.02
Female	0.48	0.48	0.48
Birth month	6.61	6.59	6.58
Age difference (with nth birth)	3.67	6.45	3.81
Family size			
Number of children	2.94	3.66	3.75
More than 3 children	0.23	-	-
More than 4 children	0.09	0.16	0.18
Family Composition			
Father or mother's spouse at home	0.83	0.85	0.83
Father at home	0.81	0.83	0.82
Mother's characteristics			
Mother's age at first child	21.32	22.83	23.41
Mother's age at census	33.88	33.76	35.43
Years of schooling	7.56	6.51	6.54
Primary complete	0.70	0.79	0.78
Secondary complete	0.23	0.17	0.17
University complete	0.07	0.05	0.05
Mother's health (% have a not alive child)	0.07	0.08	0.09
Father's characteristics			
Years of schooling	7.19	6.31	6.34
Primary complete	0.73	0.80	0.79
Secondary complete	0.20	0.15	0.15
University complete	0.08	0.05	0.06
Child Education			
Primary complete (*)	0.83	0.80	0.78
Secondary complete (*)	0.48	0.41	0.41
Attending school	0.84	0.80	0.84
Child in labor force (+15)			
Labor force Participation	0.25	0.28	0.26
Place of residence			
Population of district of residence (in logs)	12.03	11.88	11.92
Census			
Colombia 2005	42.4%	39.6%	39.7%
Dominican Republic 2002	8.8%	9.8%	9.5%
Dominican Republic 2010	9.9%	9.8%	10.1%
Ecuador 2010	17.0%	15.8%	15.9%
Nicaragua 1995	4.4%	6.2%	5.9%
Nicaragua 2005	5.6%	6.2%	6.3%
El Salvador 1992	5.1%	6.0%	5.6%
El Salvador 2007	6.8%	6.6%	6.9%
Observations	414,821	231,402	225,393

Note: * The length of primary and secondary education differ by country. We use these country specific cut-offs to compute primary and secondary education completion on the population who is above this age cutt-offs.

measure of mothers' health described in the previous section. Table 13 in the Appendix shows that the probability of giving birth to twins is positively and significantly associated with mothers' health as well as with mothers' education, conditional on maternal age, age of first birth, place of residence, and year-country fixed effects.[7]

The first stage for our instrumental variable approach in the sample of families with n or more children is as follows:

$$Nchild_{ih} = \pi_0 + \pi_1 \times Twins_{ih} + \pi_2 X_i + \pi_3 X^M_{ih} + \pi_3 X_d + \gamma_j + \epsilon_i, \tag{2}$$

where $Twins_h$ is a binary variable that takes value 1 if the nth birth in family h was a twin birth and 0 if it was a singleton, for n=2 (the "2+ sample") and for n=3 (the "3+ sample").

Bhalotra and Clarke (2020) shows that the omission of mother's characteristics that are correlated with the probability of having twins biases the IV estimate of the β_1 parameter towards zero. We find similar results with our data.[8]

3.1 First stage

Table 3 shows the effect of twin birth in the nth birth and the Kleibergen-Paap rk test of weak instrument. The results demonstrate that the twin instrument is highly predictive of family sizes in both samples. The point estimates indicate that the incidence of twins raises total fertility by about 0.72 to 0.79 births. In line with previous literature, this estimate is less than one.[9] Additionally, these results indicate that the variable chosen consistently passes weak instrument tests.

Table 3: First stage

	Sample 2+	Sample 3+
twin	0.718***	0.790***
	(0.0156)	(0.0178)
Observations	380,191	419,951
Kleibergen-Paap rk statistic	2114.66	1966.33
p-value of rk statistic	0.000	0.000

Note: This table shows first-stage effects of multiple second and third birth, respectively, on number of children. The "2+ sample" comprises first-born non-twins aged 6-18, while the "3+ sample" includes first- and second-born non-twins of the same age range (refer to Section 2.3 for further details). All regressions include indicators for child's age and gender, month of birth, mother's age at the census, mother's age at her first birth, mother's education level, mother's health indicator, birth spacing relative to the nth birth, district of residence's population (log and log square), and country-year fixed effects. In the "3+ sample" regression, an additional control for child birth order is included. Robust standard errors are reported in parentheses. For the "3+ sample," standard errors are clustered by mother's ID. The rk test statistic and corresponding p-value reject that the twin instruments are weak in each case. Statistical significance is indicated as follows: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

[7] We are aware that these variables may only adjust partially for all dimensions characterizing healthier mothers, and, as a result our IV estimates may be upward biased (too close to zero) as described by Bhalotra and Clarke (2020). In this sense, we think of our results as lower bounds.

[8] Results available upon request.

[9] Bhalotra and Clarke (2020) suggest that this is evidence of partial reduction of future fertility following twin births (i.e., a compensating fertility behavior).

4 Results

Table 4 presents the OLS and IV estimates of family size on education. For all outcomes, OLS estimates show a strong negative correlation between family size and children's educational outcomes. After instrumenting for family size with the twin instrument, we find no impacts in the 2+ sample. Some coefficients are even positive, although not statistically significant. This result is in line with much of the previous literature (Black et al. (2005); Cáceres-Delpiano (2006); Angrist et al. (2010); Marteleto and de Souza (2012); Alidou and Verpoorten (2019); Bhalotra and Clarke (2020)), which generally rejects the quantity-quality trade off for first-born children in the 2+ sample.

We find, however, that a negative quantity-quality relation holds true for some outcomes in our 3+ sample, although the coefficients are much lower than in the OLS specification. In particular, according to the IV estimates in the 3+ sample, one additional child in the household causes a reduction of about 0.035 units in the z-score. This magnitude corresponds to the equivalent of between 0.077 and 0.094 years of schooling for a child of 15 years old in the countries considered[10] and a reduction in 2.3 percentage points in the probability of completing primary education, equivalent of a 3% reduction in primary completion, taking as reference the sample mean of primary completion. The probability of attending formal education at the moment of the census and the probability of completing secondary education are not affected by family size. These results point to family size having primarily an effect in early stages of the education trajectory, affecting the years of schooling outcome mainly trough its effects on primary education completion as children in larger families do not show different probabilities of completing secondary education. Notice that this is plausible given that secondary education coverage was relatively low for all the cohorts of children analyzed in this study. Table 5 additionally shows that the effects of larger families appear only for the second-born children in the 3+ sample.

Table 4: OLS and Twins IV estimates

	Yrs. of School (z-score)		Attend		Primary		Secondary	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
A. Sample 2+	-0.129***	0.040*	-0.029***	-0.004	-0.042***	0.015	-0.044***	0.004
	(0.002)	(0.022)	(0.001)	(0.008)	(0.001)	(0.014)	(0.002)	(0.050)
	380,191	380,191	377,513	377,513	206,195	206,195	35,892	35,892
B. Sample 3+	-0.142***	-0.035*	-0.034***	-0.009	-0.049***	-0.023**	-0.036***	0.040
	(0.002)	(0.021)	(0.001)	(0.008)	(0.001)	(0.012)	(0.002)	(0.037)
	419,951	419,951	416,232	416,232	233,388	233,388	39,121	39,121

Note: This table shows the OLS and twins IV estimates of the coefficient β_1 of Equation 1 for the "2+ sample" (panel A) and "3+ sample" (panel B). The "2+ sample" comprises first-born non-twins aged 6-18, while the "3+ sample" includes first and second-born non-twins of the same age range (refer to Section 2.3 for further details). The outcome "Primary" refers to primary education complete and "Secondary" refers to secondary education complete. The regressions for "Primary" and "Secondary" are conducted on a subset of individuals corresponding to country-specific ages at which the respective education levels are expected to be completed. All regressions include as control variables child's age fixed effects, gender, age differences with the nth birth, month of birth fixed effects, mother's age at the census fixed effects, mother's age at her first birth fixed effects, mother's education level, mother's health, interactions between mother's education and mother's health, district of residence's population (log and log square), and country-year fixed effects. In the "3+ sample" regressions, birth order of the child is also included as control variable. Robust standard errors are reported in parentheses. For the "3+ sample," standard errors are clustered by mother's ID. Statistical significance is indicated as follows: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

[10] The coefficient is multiplied by the standard deviation of completed years of education of children of 15 years old (see Table 14 in the Appendix).

Table 5: IV estimates by birth order

	Yrs- of School (z-score)		Attend		Primary		Second	
	First	Second	First	Second	First	Second	First	Second
Sample 3+	-0.021	-0.049*	-0.002	-0.018*	-0.018	-0.030*	0.0211	0.0864
	(0.026)	(0.026)	(0.010)	(0.010)	(0.014)	(0.017)	(0.0419)	(0.0741)
Observations	207,938	212,013	206,103	210,129	128,002	105,386	24,090	15,031

Note: This table shows the twins IV estimates of the coefficient β_1 of Equation 1 birth order of the child (first or second) in the "3+ sample". The "3+ sample" comprises first and second-born non-twins aged 6-18 (refer to Section 2.3 for further details). The outcome "Primary" refers to primary education complete and "Secondary" refers to secondary education complete. The regressions for "Primary" and "Secondary" are conducted on a subset of individuals corresponding to country-specific ages at which the respective education levels are expected to be completed. All regressions include as control variables child's age fixed effects, gender, age differences with the nth birth, month of birth fixed effects, mother's age at the census fixed effects, mother's age at her first birth fixed effects, mother's education level, mother's health, interactions between mother's education and mother's health, district of residence's population (log and log square), and country-year fixed effects. In the "3+ sample" regressions, birth order of the child is also included as control variable. Robust standard errors are reported in parentheses. For the "3+ sample," standard errors are clustered by mother's ID. Statistical significance is indicated as follows: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

5 Heterogeneous effects

We next look at heterogeneous effects of family size according to parents' characteristics, children's characteristics and key characteristics of the district of residence. This analysis allows us to shed light on the potential mechanisms affecting the relationship between family size and children's human capital. First, we estimate the heterogeneous effects by parent's characteristics. A first dimension we look at is parents' educational attainment as a proxy of their socioeconomic status. Financial restrictions may be more binding for households of lower socioeconomic backgrounds; hence, fertility shocks could have a larger effect on children's human capital investments of these families.

A second dimension in the heterogeneity analysis refers to children's characteristics. We look at the heterogeneous effects by the gender of the child. Gender norms may affect parental investment decisions in their children's human capital upon the onset of a fertility shock. Finally, we analyze heterogeneous effects by characteristics of the district of residence.[11] First, we split the sample according to population size of the district of residence. If returns to education in sparsely populated districts are smaller, the tension between fertility and investments to develop human capital in children may loosen. Second, we split the sample according to the provision of education services across districts, which we measure as the ratio of teachers to school-age children.[12] We expect that in districts with high supply restrictions both large and small families face limitations to school access, while in districts with many teachers both types of families are not constrained in this dimension. Hence, in both of these cases the effect of family size on human capital of children is expected to be weaker.[13]

5.1 Parents' socioeconomic characteristics

Our evidence also suggest that the fertility shock mostly affects families with tightest financial restrictions. Results in table 6 show that the children in families of larger size are those with less educated parents. Our results show that in families with low educated parents, one additional child reduces the z-score of schooling years in 0.049 units for those children, as well as it reduces in 3.8 percentage points the probability of completing primary school. The effects are null or even positive for children in families with highly educated parents.

5.2 Children's gender

Parental decisions regarding human capital investments tend to differ depending on the gender of the older children upon the occurrence of the fertility shock. Table 7 shows that female children are less affected than male children. Even more, female children in families of two or more children tend to be positively affected by the fertility shock. Gender norms may guide the underlying decisions: to overcome financial restrictions, male children may be required to work outside the home in a paid job, which, in turn, increases their likelihood of dropping out of school. To explore if this hypothesis is plausible we estimate the same IV regressions but using as an outcome the labor force participation status of children. Table 8 shows that a larger family increases male children probabilities to be in the labor force at the time of census, while female children are not affected.

[11] In general, the district refers to the lowest level of sub-national government (municipalities or departments) in each country. The number of districts in the IPUMS-I data are: 433 in Colombia, 65 in Dominican Republic, 77 in Ecuador, 68 in Nicaragua, and 103 in El Salvador.

[12] We identify teachers through occupation codes available in the census data. We do not distinguish teachers according to the educational level in which they teach.

[13] In both cases, population size and provision of education services, we classify the districts within country and year in three terciles (q1, q2, and q3) according to their location in the distribution of the respective measure.

Table 6: IV estimates by maximum education level of parents

	Yrs. School. (z-score)		Attend		Primary		Secondary	
	Low	High	Low	High	Low	High	Low	High
A. Sample 2+	0.031	0.046*	-0.004	-0.007	0.018	0.017	0.034	-0.0355
	(0.033)	(0.026)	(0.013)	(0.008)	(0.022)	(0.014)	(0.069)	(0.069)
Observations	238,760	141,431	236,322	141,191	130,779	75,416	23,869	12,023
B. Sample 3+	-0.049*	-0.009	-0.013	0.004	-0.038**	-0.001	0.061	-0.002
	(0.027)	(0.028)	(0.011)	(0.008)	(0.016)	(0.013)	(0.046)	(0.058)
Observations	303,565	116,386	300,042	116,190	167,150	66,238	28,504	10,617

Note: This table shows the twins IV estimates of the coefficient β_1 of Equation 1 by the maximum education level of the parents (Low and High). "Low" education refers to at most primary education complete or lower; 'High" education refers to secondary education complete or more. The "2+ sample" comprises first-born non-twins aged 6-18, while the "3+ sample" includes first and second-born non-twins of the same age range (refer to Section 2.3 for further details). The outcome "Primary" refers to primary education complete and "Secondary" refers to secondary education complete. The regressions for "Primary" and "Secondary" are conducted on a subset of individuals corresponding to country-specific ages at which the respective education levels are expected to be completed. All regressions include as control variables child's age fixed effects, gender, age differences with the nth birth, month of birth fixed effects, mother's age at the census fixed effects, mother's age at her first birth fixed effects, mother's education level, mother's health, interactions between mother's education and mother's health, district of residence's population (log and log square), and country-year fixed effects. In the "3+ sample" regressions, birth order of the child is also included as control variable. Robust standard errors are reported in parentheses. For the "3+ sample," standard errors are clustered by mother's ID. Statistical significance is indicated as follows: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 7: IV estimates by gender of the Child

	Yrs. of school (z-score)		Attend		Primary		Secondary	
	Female	Male	Female	Male	Female	Male	Female	Male
A. Sample 2+	0.070**	0.010	-0.005	-0.005	0.037**	-0.007	0.054	-0.027
	(0.031)	(0.031)	(0.011)	(0.011)	(0.018)	(0.021)	(0.085)	(0.060)
Observations	183,058	197,133	181,881	195,632	98,623	107,572	16,388	19,504
B. Sample 3+	-0.030	-0.038	-0.003	-0.014	-0.020	-0.027*	0.104*	-0.010
	(0.029)	(0.027)	(0.011)	(0.011)	(0.017)	(0.015)	(0.061)	(0.047)
Observations	202,129	217,822	200,458	215,774	111,395	121,993	17,699	21,422

Note: This table shows the twins IV estimates of the coefficient β_1 of Equation 1 by gender of the child. The "2+ sample" comprises first-born non-twins aged 6-18, while the "3+ sample" includes first and second-born non-twins of the same age range (refer to Section 2.3 for further details). The outcome "Primary" refers to primary education complete and "Secondary" refers to secondary education complete. The regressions for "Primary" and "Secondary" are conducted on a subset of individuals corresponding to country-specific ages at which the respective education levels are expected to be completed. All regressions include as control variables child's age fixed effects, age differences with the nth birth, month of birth fixed effects, mother's age at the census fixed effects, mother's age at her first birth fixed effects, mother's education level, mother's health, interactions between mother's education and mother's health, district of residence's population (log and log square), and country-year fixed effects. In the "3+ sample" regressions, birth order of the child is also included as control variable. Robust standard errors are reported in parentheses. For the "3+ sample," standard errors are clustered by mother's ID. Statistical significance is indicated as follows: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Table 8: IV estimates by gender of the Child for Labor Force Participation

	Labor Force Participation	
	Female	Male
A. Sample 2+	0.017	-0.049
	(0.028)	(0.033)
Observations	49,741	56,562
B. Sample 3+	0.026	0.061**
	(0.025)	(0.025)
Observations	55,418	63,432

Note: This table shows the twins IV estimates of the coefficient β_1 of Equation 1 by gender of the child. The "2+ sample" comprises first-born nontwins aged 15-18, while the "3+ sample" includes first and second-born nontwins of the same age range (refer to Section 2.3 for further details). The labor force participation outcome is measured only for children 15 to 18 years old. All regressions include as control variables child's age fixed effects, age differences with the nth birth, month of birth fixed effects, mother's age at the census fixed effects, mother's age at her first birth fixed effects, mother's education level, mother's health, interactions between mother's education and mother's health, district of residence's population (log and log square), and country-year fixed effects. In the "3+ sample" regressions, birth order of the child is also included as control variable. Robust standard errors are reported in parentheses. For the "3+ sample," standard errors are clustered by mother's ID. Statistical significance is indicated as follows: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

5.3 Characteristics of the district of residence: population and education supply

Table 9 shows the heterogeneous effects by two key districts's characteristics: population size and the ratio of teachers to school-age children. Districts are divided in terciles according to the country-year distribution of each of these variables. We find that the negative effects of family size on educational outcomes tend to be concentrated in districts with an intermediate supply of formal education, especially when we focus on the outcome of years of schooling. As expected, in undersupplied areas, the trade-off does not seem to operate, as acquiring more education may be extremely expensive for parents independently of their socioeconomic status (e.g., they must seek education in other locations). Also, a non-statistically significant trade-off is observed in districts where the provision of formal education is high, as neither large nor small families are likely to be constrained in this dimension on those locations. On the contrary, we do not find a clear pattern in the analysis by population sizes of the districts where children reside.

Table 9: IV estimates by population size and education supply of district of residence (terciles)

	Yrs. of school (z-score)			Attend			Primary			Secondary		
	q1	q2	q3	q1	q2	q3	q1	q2	q3	q1	q2	q3
1. Population size												
A. Sample 2+	-0.002	0.024	0.026	-0.033	-0.006	-0.001	0.053	0.014	-0.004	-0.099	0.054	-0.016
	(0.064)	(0.045)	(0.027)	(0.028)	(0.017)	(0.010)	(0.037)	(0.026)	(0.018)	(0.126)	(0.072)	(0.073)
Observations	48,426	73,722	224,283	47,971	72,996	222,786	27,245	41,147	121,151	5,045	7,610	21,587
B. Sample 3+	-0.009	-0.045	-0.024	-0.043*	0.032	-0.012	-0.059	-0.025	-0.006	0.099	0.085	0.019
	(0.063)	(0.049)	(0.028)	(0.025)	(0.021)	(0.011)	(0.044)	(0.028)	(0.014)	(0.112)	(0.085)	(0.045)
Observations	56,361	84,681	238,192	55,710	83,644	236,161	32,283	47,719	132,821	5,680	8,435	22,994
2. Formal educ. supply												
A. Sample 2+	0.011	-0.021	0.055*	0.013	-0.031**	0.000	-0.002	0.010	0.006	-0.018	0.104	-0.042
	(0.049)	(0.045)	(0.037)	(0.017)	(0.016)	(0.012)	(0.026)	(0.025)	(0.021)	(0.070)	(0.128)	(0.076)
Observations	87,553	112,666	146,212	86,555	111,850	145,348	49,258	62,788	77,497	9,522	11,131	13,589
B. Sample 3+	-0.003	-0.083**	0.018	-0.010	-0.011	0.000	-0.030	-0.032	0.014	-0.011	0.022	0.090
	(0.044)	(0.042)	(0.032)	(0.019)	(0.016)	(0.013)	(0.024)	(0.024)	(0.018)	(0.070)	(0.068)	(0.058)
Observations	99,968	125,061	154,205	98,533	123,896	153,086	56,775	70,844	85,204	10,467	12,115	14,527

Note: This table shows the twins IV estimates of Equation 1 by population size of the district of residence and the educaiton supply of the district (in terciles). "q1" reffers to the first tercile, "q2" reffers to the second tercile, and "q3" reffers to the third tercile. Education supply is measured as the ratio of teachers to school-age children in the district. Terciles are constructed within country-year. The "2+ sample" comprises first-born non-twins aged 6-18, while the "3+ sample" includes first and second-born non-twins of the same age range (refer to Section 2.3 for further details). The outcome "Primary" refers to primary education complete and "Secondary" refers to secondary education complete. The regressions for "Primary" and "Secondary" are conducted on a subset of individuals corresponding to country-specific ages at which the respective education levels are expected to be completed. All regressions include as control variables child's age fixed effects, gender, age differences with the nth birth, month of birth fixed effects, mother's age at the census fixed effects, mother's age at her first birth fixed effects, mother's education level, mother's health, interactions between mother's education and mother's health, district of residence's population (log and log square), and country-year fixed effects. In the "3+ sample" regressions, birth order of the child is also included as control variable. Robust standard errors are reported in parentheses. For the "3+ sample," standard errors are clustered by mother's ID. Statistical significance is indicated as follows: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

6 Country analysis

Finally, we estimate our regressions separately by country in the 3+ sample, except for Nicaragua and El Salvador, which we pool together in order to maintain reasonable sample sizes. Overall, the signs of the IV estimates for each outcome and country are similar to those of the pooled estimation (Table 4, although they are not statistically significant. However, we do find large and statistically significant negative effects of family size on children's educational outcomes among low socioeconomic status families in all countries, with the only exception of Dominican Republic.

Table 10: OLS and Twins IV estimates by country (3+ sample)

	Yrs. of school (z-score)		Attend		Primary		Secondary	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
A. COL	-0.031***	-0.014	-0.159***	-0.046	-0.051***	-0.024	-0.042***	0.057
	(0.001)	(0.014)	(0.003)	(0.034)	(0.001)	(0.017)	(0.002)	(0.048)
Observations	166,384	166,384	167,278	167,278	109,067	109,067	20,992	20,992
B. ECU	-0.028***	-0.026	-0.104***	-0.074	-0.030***	-0.013	-0.049***	-0.018
	(0.002)	(0.019)	(0.005)	(0.053)	(0.002)	(0.028)	(0.007)	(0.108)
Observations	66,230	66,230	66,230	66,230	37,339	37,339	4,233	4,233
C. NIC + SLV	-0.040***	0.003	-0.136***	-0.001	-0.055***	-0.028	-0.023***	0.106
	(0.002)	(0.021)	(0.003)	(0.043)	(0.002)	(0.036)	(0.002)	(0.105)
Observations	101,870	101,870	104,695	104,695	42,889	42,889	9,027	9,027
D. DOM	-0.031***	-0.009	-0.117***	-0.032	-0.042***	-0.024	-0.040***	0.010
	(0.002)	(0.011)	(0.005)	(0.041)	(0.003)	(0.022)	(0.006)	(0.103)
Observations	81,748	81,748	81,748	81,748	44,093	44,093	4,869	4,869

Note: This table shows the OLS and twins IV estimates of the coefficient β_1 of Equation 1 for "3+ sample" by country. The "3+ sample" comprises first and second-born non-twins aged 6-18 (refer to Section 2.3 for further details). The outcome "Primary" refers to primary education complete and "Secondary" refers to secondary education complete. The regressions for "Primary" and "Secondary" are conducted on a subset of individuals corresponding to country-specific ages at which the respective education levels are expected to be completed. All regressions include as control variables child's age fixed effects, gender, age differences with the nth birth, month of birth fixed effects, mother's age at the census fixed effects, mother's age at her first birth fixed effects, mother's education level, mother's health, interactions between mother's education and mother's health, district of residence's population (log and log square), and country-year fixed effects. In the "3+ sample" regressions, birth order of the child is also included as control variable. Robust standard errors are reported in parentheses. For the "3+ sample," standard errors are clustered by mother's ID. Statistical significance is indicated as follows: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 11: IV estimates by maximum education level of parents by country

	Yrs. School. (z-score)		Attend		Primary		Secondary	
	Low	High	Low	High	Low	High	Low	High
A. COL	-0.087**	0.075	-0.017	-0.001	-0.043**	0.017	0.053	0.070
	(0.042)	(0.046)	(0.017)	(0.017)	(0.022)	(0.015)	(0.057)	(0.081)
Observations	127,104	40,174	126,324	40,060	83,076	25,991	16,023	4,969
B. ECU	-0.148*	0.032	-0.034	-0.008	-0.034	0.009	-0.106	0.093
	(0.083)	(0.054)	(0.030)	(0.017)	(0.044)	(0.030)	(0.141)	(0.125)
Observations	39,899	26,331	39,899	26,331	22,138	15,201	2,368	1,865
C. NIC + SLV	-0.007	0.018	0.003	0.017	-0.037	-0.014	0.229	-0.118
	(0.054)	(0.059)	(0.028)	(0.016)	(0.047)	(0.042)	(0.170)	(0.145)
Observations	85,757	18,938	83,014	18,856	34,710	8,179	7,180	1,847
D. DOM	0.015	-0.085	-0.020	0.003	-0.036	-0.016	0.088	-0.085
	(0.057)	(0.056)	(0.018)	(0.013)	(0.036)	(0.026)	(0.141)	(0.140)
Observations	50,805	30,943	50,805	30,943	27,226	16,867	2,933	1,936

Note: This table shows the twins IV estimates of the coefficient β_1 of Equation 1 in the 3+ sample by country and by the maximum education level of the parents (Low and High). "Low" education refers to at most primary education complete or lower; 'High" education refers to secondary education complete or more. The "3+ sample" comprises first and second-born non-twins aged 6-18 (refer to Section 2.3 for further details). The outcome "Primary" refers to primary education complete and "Secondary" refers to secondary education complete. The regressions for "Primary" and "Secondary" are conducted on a subset of individuals corresponding to country-specific ages at which the respective education levels are expected to be completed. All regressions include as control variables child's age fixed effects, gender, age differences with the nth birth, month of birth fixed effects, mother's age at the census fixed effects, mother's age at her first birth fixed effects, mother's education level, mother's health, interactions between mother's education and mother's health, district of residence's population (log and log square), and country-year fixed effects. In the "3+ sample" regressions, birth order of the child is also included as control variable. Robust standard errors are reported in parentheses. For the "3+ sample," standard errors are clustered by mother's ID. Statistical significance is indicated as follows: $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

7 Conclusions

The causal relationship between family size and children's human capital accumulation has been extensively analyzed worldwide, yielding surprisingly inconclusive results. Some of the limitations in arriving at definitive answers to the significant question of the existence of a quantity-quality trade-off arise from methodological challenges, which we can address by implementing a credible identification strategy.

Another notable limitation in this type of analysis pertains to data constraints, which frequently hinder the thorough exploration of the numerous theoretical mechanisms underlying the complex interaction between decisions concerning fertility and investments in human capital for children. Given the rich data that we use, we are also capable of illuminating the mechanisms at play within five developing countries in Latin America –a region characterized by a dramatic demographic transition and substantial socioeconomic disparities in children's human capital development. In this context, our findings gain significance, particularly in light of the extensive evidence highlighting the pivotal role that circumstances linked to the family of origin have on the magnitude and persistence of human capital disparities.

Our results reveal that a fertility shock negatively affects schooling years and primary education completion for first- and second-born children. These effects are more pronounced for second-born children in families with three or more children. The fertility shock tends to impact male children more, potentially due to gender norms leading them to work rather than pursue education. Our results also suggest that the family size effect is strongest in financially constrained families and districts with intermediate education supply, where public provision of school services can make the difference for children of disadvantaged households. Country-specific analyses mostly show consistent negative family size effects, particularly for low socioeconomic status families.

We can extract a key policy lesson from the analysis of the mechanisms at work in the quantity-quality trade-off that we found: aspects related to norms as well as to the provision of goods and services that can complement parental efforts are crucial in the interplay between fertility decisions and the development of children's human capital. Therefore, the reduction in fertility rates, specially among the most disadvantaged families, will not *per se* ensure a closing socioeconomic gap in educational attainment in regions like Latin America, were these gaps are large and persistent.

References

Alidou, S. and Verpoorten, M. (2019). Family size and schooling in sub-saharan africa: testing the quantity-quality trade-off. *Journal of Population Economics*, 32(4):1353–1399.

Angrist, J., Lavy, V., and Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28(4):773–824.

Åslund, O. and Grönqvist, H. (2010). Family size and child outcomes: Is there really no trade-off? *Labour Economics*, 17(1):130–139.

Becker, G. S. and Lewis, H. G. (1973). On the interaction between the quantity and quality of children. *Journal of political Economy*, 81(2, Part 2):S279–S288.

Becker, G. S. and Tomes, N. (1976). Child endowments and the quantity and quality of children. *Journal of political Economy*, 84(4, Part 2):S143–S162.

Bhalotra, S. and Clarke, D. (2019). Twin birth and maternal condition. *Review of Economics and Statistics*, 101(5):853–864.

Bhalotra, S. and Clarke, D. (2020). The twin instrument: Fertility and human capital investment. *Journal of the European Economic Association*, 18(6):3090–3139.

Bhalotra, S. and Rawlings, S. B. (2011). Intergenerational persistence in health in developing countries: The penalty of gender inequality? *Journal of Public Economics*, 95(3-4):286–299.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (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.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2010). Small family, smart family?: family size and the iq scores of young men. *Journal of Human Resources*, 45(1):33–58.

Cáceres-Delpiano, J. (2006). The impacts of family size on investment in child quality. *Journal of Human Resources*, 41(4):738–754.

De La Mata, D., Berniell, L., Schargrodsky, E., Álvarez, F., and Alves, G. (2022). Inherited inequalities: The role of skills, employment, and wealth in the opportunities of new generations.

Fitzsimons, E. and Malde, B. (2014). Empirically probing the quantity–quality model. *Journal of Population Economics*, 27:33–68.

Marteleto, L. J. and de Souza, L. R. (2012). The changing impact of family size on adolescents' schooling: Assessing the exogenous variation in fertility using twins in brazil. *Demography*, 49(4):1453–1477.

Rosenzweig, M. R. and Wolpin, K. I. (1980). Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica: journal of the Econometric Society*, pages 227–240.

Ruggles, S., McCaa, R., Sobek, M., and Cleveland, L. (2015). The ipums collaboration: integrating and disseminating the world's population microdata. *Journal of Demographic Economics*, 81(2):203–216.

8 Appendix

Table 12: Descriptive statistics of children in families that have twins in the nth birth

	2+ Sample	3+ Sample	
	first-born	**first-born**	**second-born**
Child characteristics			
Age	12.59	13.47	12.09
Female	0.49	0.48	0.48
Birth month	6.72	6.56	6.65
Age difference (with nth birth)	4.05	6.94	4.23
Family size			
Number of children	3.53	4.36	4.42
More than 3 children	0.34	-	-
More than 4 children	0.11	0.23	0.26
Family Composition			
Father or mother's spouse at home	0.82	0.84	0.83
Father at home	0.80	0.81	0.82
Mother characteristics			
Mother's age at first child	22.27	23.28	23.84
Mother's age at census	34.86	34.36	35.92
Years of schooling	8.17	6.67	6.72
Primary complete	0.64	0.77	0.77
Secondary complete	0.26	0.19	0.18
University complete	0.10	0.05	0.05
Mother's health (% have a not alive child)	0.06	0.06	0.07
Father characteristics			
Years of education	7.53	6.59	6.57
Primary complete	0.70	0.78	0.78
Secondary complete	0.21	0.16	0.16
University complete	0.09	0.06	0.06
Child Education			
Primary complete (*)	0.86	0.79	0.77
secondary complete (*)	0.52	0.45	0.43
Attending school	0.22	0.30	0.29
Child in Labor force +15			
Labor force participation	0.22	0.30	0.29
District of residence			
Population of district of residence (in logs)	12.03	11.88	11.92
Census			
Colombia 2005	44.6%	40.2%	40.2%
Dominican Republic 2002	11.0%	13.1%	12.3%
Dominican Republic 2010	11.9%	12.3%	11.9%
Ecuador 2010	14.4%	11.4%	12.5%
Nicaragua 1995	3.4%	5.6%	5.5%
Nicaragua 2005	4.5%	5.4%	5.6%
El Salvador 1992	4.3%	5.1%	5.0%
El Salvador 2007	5.9%	6.9%	7.0%
Observations	2876	1775	1732

Note: * The length of primary and secondary education differ by country. We use these country specific cut-offs to compute primary and secondary education completion on the population who is above this age cutt-offs.

Table 13: Probability of giving birth to twins

	P(Twin = 1)
Mother Health	-0.001**
	(0.000)
Mother Years School==1	0.002**
	(0.001)
Mother Years School==2	0.002***
	(0.001)
Mother Years School==3	0.001
	(0.001)
Mother Years School==4	0.001
	(0.001)
Mother Years School==5	0.001
	(0.000)
Mother Years School==6	0.001*
	(0.000)
Mother Years School==7	-0.000
	(0.001)
Mother Years School==8	0.001*
	(0.001)
Mother Years School==9	0.001
	(0.001)
Mother Years School==10	0.001
	(0.001)
Mother Years School==11	0.001***
	(0.000)
Mother Years School==12	0.001**
	(0.001)
Mother Years School==13	0.002**
	(0.001)
Mother Years School==14	0.001*
	(0.001)
Mother Years School==15	0.001
	(0.001)
Mother Years School==16	0.001
	(0.001)
Mother Years School==17	0.000
	(0.001)
Mother Years School==18	0.003***
	(0.001)
Observations	667,727

*Note:*This table shows marginal effects of a Probit regression where the dependent variable takes value one if the mother had twins. The estimation sample considers all mothers in the "2+ sample" and "3+ sample". To test positive selection into twins, marginal effects for mother´s education and health measures are reported. The regression includes as control variables the mother's age at the census fixed effects, mother's age at her first birth fixed effects, district of residence's population (log and log square) and census fixed effects. Robust standard errors are reported in parentheses. Statistical significance is indicated as follows: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 14: Mean and standard deviation of years of schooling for census-age group

	COL-2005		**DOM-2002**		**DOM-2010**		**ECU-2010**		**NIC-1995**		**NIC-2005**		**SLV-1992**		**SLV-2007**	
age	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
6	0.42	0.60	0.05	0.23	0.43	0.49	4.54	6.96	0.29	0.45	0.34	0.57	0.10	0.30	0.00	0.00
7	1.13	0.84	0.10	0.35	1.02	0.74	2.13	3.16	0.67	0.75	0.86	0.88	0.53	0.50	0.66	0.47
8	1.92	1.03	1.54	1.00	1.73	1.02	2.52	1.63	1.15	1.05	1.41	1.13	1.07	0.85	1.40	0.71
9	2.73	1.21	2.26	1.23	2.49	1.28	3.34	1.23	1.66	1.34	2.07	1.41	1.66	1.18	2.17	0.94
10	3.53	1.41	2.91	1.45	3.22	1.54	4.19	1.21	2.17	1.61	2.81	1.74	2.33	1.47	2.97	1.16
11	4.35	1.60	3.85	1.77	4.06	1.78	4.76	1.12	2.77	1.90	3.45	1.96	2.93	1.76	3.79	1.37
12	5.13	1.81	4.43	2.00	4.73	2.08	5.21	2.62	3.24	2.18	4.08	2.19	3.56	2.09	4.49	1.67
13	5.89	1.99	5.21	2.24	5.84	2.20	6.75	2.36	3.86	2.48	4.73	2.44	4.25	2.36	5.31	1.87
14	6.66	2.24	5.89	2.51	6.78	2.48	7.22	2.29	4.31	2.77	5.35	2.70	4.85	2.71	5.99	2.20
15	7.30	2.56	6.63	2.69	7.49	2.64	7.85	2.20	4.71	3.04	5.89	2.99	5.35	3.16	6.73	2.46
16	7.95	2.84	7.25	2.93	8.26	2.88	8.57	2.37	4.95	3.36	6.27	3.25	5.88	3.45	7.24	2.81
17	8.43	3.16	7.72	3.20	8.78	3.14	9.25	2.69	5.20	3.60	6.50	3.59	6.11	3.78	7.66	3.21
18	8.82	3.47	8.16	3.49	9.20	3.41	9.66	3.03	5.23	3.83	6.76	3.85	6.34	4.12	8.11	3.68