Family size and schooling in Sub-Saharan Africa:

Testing the quantity-quality trade-off^a

Sahawal Alidou¹

Marijke Verpoorten²

This is a pre-print of the article “ Family size and schooling in Sub-Saharan Africa: Testing the quantity-quality trade-off”. Journal: Journal of Population Economics

DOI: 10.1007/s00148-019-00730-z

Abstract

Many family planning programs are based on the idea that small families lead to improved development outcomes, such as more schooling for children. Because of endogeneity issues, this idea is however difficult to verify. A handful of studies have made use of twin birth to deal with the endogeneity of family size. We do so for Sub-Saharan African countries. In a compilation of 86 survey rounds from 34 countries, we exploit the birth of twins to study the effect of a quasi-exogenous increase in family size on the schooling of children at the first, second and third birth order. Our findings do not support the generally assumed negative effect of family size on schooling.

Keywords : Family size, schooling, quantity-quality trade-off, Sub-Saharan Africa

^a We received much appreciated comments from Elena Briones Alonso, Philip H. Ross, Pieter Serneels, Nik Stoop and participants at seminars, conferences and workshops in Leuven (LICOS-KU Leuven seminar), Antwerp (Doctoral Day), and Oxford (2017 CSAE conference), as well as from the Editor and two anonymous referees of this journal.

¹ IOB – Institute of Development Policy and Management (University of Antwerp, Prinsstraat 13, 2000 Antwerp), LICOS – Centre for Institutions and Economic Performance (KU Leuven), sahawal.alidou@uantwerp.be

² Corresponding author. IOB–Institute of Development Policy and Management (University of Antwerp, Prinsstraat 13, 2000 Antwerp), marijke.verpoorten@uantwerp.be, Tel. 003232655297, Fax 003232655741

1. Introduction

Countries worldwide have devoted much effort and resources to family planning programs (Bongaarts, 2009). Most of these programs have been voluntary, but some have left little choice to parents, such as China's one-child program and India's sterilization camps. A major assumption underlying these programs is that "a small family is a happy family"³, or that a reduction in family size enables families to raise investments in human capital per child, leading in its turn to a stronger economy (Bongaarts, 2009). Intuitively the assumed causality between small family sizes and high schooling attainments makes sense: dividing scarce resources among less children, leaves each child with more resources.

This intuition found support in well-known social science research. Judith Blake (1989), studying U.S. families, famously concluded that children from one- and two-child families are better educated and more successful than children in larger families because their parents have more time and money to invest in them. This 'resource dilution model' was backed up by economic theory, in particular in a pioneering paper written by Gary Becker (1960), in which the quantity and quality of children are modelled as substitutes from the parents' point of view.

However, there also exist theories that support a *positive* causal effect of family size on children's schooling. These theories break with Blake's and Becker's assumptions that children only imply a cost to parents, and that more children imply higher costs. As such, the quantity-quality trade-off needs no longer hold when allowing for part-time child work (Mueller, 1984a; Marteleto and de Souza, 2013), or when older children work to provide for the younger ones (the so-called chain arrangement), or when there are economies of scale in raising children, with children sharing clothes, text books, transport to school, or knowledge and skills (Guo and Van Wey, 1999; Rosenzweig and Zhang, 2009; Steelman et al., 2002; Qian, 2009). Economies of scale can also be

³ This text featured on a 2 Rupee Coin issued by India in 1993, to promote family planning.

present in household chores, such that the time each child spends on chores reduces with the number of siblings, thus freeing up time for school.

Despite the diversity of theoretical predictions, it is hard to move away from the idea of a negative causal relation. An important reason for its stickiness lies in the strong negative correlation between family size and children's schooling, and the difficulty to empirically distinguish this correlation from the actual causal effect of family size on schooling. To make this distinction, one needs to purge the correlation of confounding factors. Most importantly, parents' characteristics determine preferences both for the number of children and their years of schooling. For instance, mothers who enjoyed more years of schooling generally prefer smaller families and at the same time are likely to give more importance to their children's schooling. Other confounding factors include wealth, social norms regarding fertility and child labor, labor market opportunities for adults and children, and the availability and quality of old-age security schemes and education policies (Rosenzweig and Wolpin, 1980; Angrist et al., 2010; Black et al., 2010). To the extent that these confounding factors are not perfectly observed and controlled for, the estimated relation between family size and children's schooling is plagued by endogeneity issues.

In this paper, we aim to remedy this endogeneity problem in a sample of children from 208,729 households across 34 Sub-Saharan African countries. In 3,844 of these households twins were born, causing a quasi-exogenous increase in household size. Provided controls for certain mother characteristics (Smits and Monden, 2011; Bhalotra and Clarke, 2016)⁴, and for the health condition of twins (Rosenzweig and Zhang, 2009)⁵, twin birth can be used as a plausibly exogenous instrumental variable to isolate the causal effect of family size on the educational outcomes of

⁴ According to Smits and Monden (2011) twinning is associated with maternal age, maternal height, smoking, oral contraceptive use, race and ethnicity.

⁵ Rosenzweig and Zhang (2009) point out that twins have lower average birth weight than singletons, and perhaps worse health or cognitive achievement later on, and this may threaten the exclusion restriction if parents therefore allocate resources away from twins, towards older singleton-birth children.

children born *prior* to the twin birth (Angrist et al., 2010; Conley et al., 2012). The same does not apply for children born *after* the twins, because their birth can be the result of parental choice. Concretely, in our IV approach, we look at the outcomes of first-, first- and second-, and first-, second- and third- born children respectively in families of two or more, three or more, and four or more children, using the birth of twins at the second, third birth and fourth order as the instrumental variable.

Our empirical investigation adds to the body of literature that has tried to empirically unearth the quantity-quality trade-off by relying on a number of techniques, such as tracing children's intellectual abilities in a longitudinal analysis of families (Guo and Van Wey, 1999), exploiting the gradual roll-out and subsequent relaxation of family planning programs (Liu, 2014; Qian, 2009), a randomized controlled trial in family planning (Sinha, 2005; Joshi and Schultz, 2007), and instrumental variable approaches based on reported miscarriages (Maralani, 2008), siblings' sex-composition (Angrist et al., 2010; Conley and Glauber, 2006; Black et al., 2010; Fitzsimons and Malde, 2014; Lee, 2008), and twin birth (Rosenzweig and Wolpin, 1980; Rosenzweig and Zhang, 2009; Angrist et al., 2010; Black et al., 2005, 2010; Marteleto and de Souza, 2012, 2013; Mogstad and Wiswall, 2016; Bhalotra and Clarke, 2016). These exercises in causal identification have not uniformly yielded negative estimates of the effect of sibship on schooling. Instead, the effect turns out to vary over time, across regions and subpopulations, across birth order, and across the exact outcome of interest studied (e.g. private schooling, school enrolment, educational attainment, or IQ)⁶.

⁶ In the first study using twins as an instrument, Rosenzweig and Wolpin (1980) find that schooling levels of Indian children decrease with exogenous increases in fertility. Consecutive studies using twins find mixed results. Caceres (2006), relying on US Census data, finds a negative effect of family size on the probability of enrollment in private schooling; Black et al. (2010) find that family size negatively affects the IQ of younger cohorts in Norway; and Li et al. (2008), relying on Chinese twins for identification, find a negative effect of sibship on schooling attainments. However, family size is found to have no effect on children's educational attainment in Norway (Black et al., 2005) and in Israel (Angrist et al. 2010). Furthermore, Marteleto and de Souza (2012), studying the effect of family size on adolescents' schooling in Brazil, find a positive effect in periods and regions in the earlier stages of socioeconomic development;

None of these exercises in causal identification has however looked specifically at Sub-Saharan African countries.⁷ Our paper fills this gap. There are several reasons why SSA provides an interesting setting for such analysis. First, in SSA, the majority of households face tight budget constraints, schooling is barely compulsory, and children's participation in the labor market and in time-consuming household chores is socially still largely accepted (Bass, 2004). Combined, these features make it very likely that a household's decision to invest in children's formal education involves important trade-offs. Second, in most African cultures, family members are bound to act for the benefit of the collective, be it the nuclear family or the extended family, the clan or ethnic group (Lloyd and Blanc, 1996). Regarding the decision to invest in schooling, this implies that the benefits of schooling are expected to be shared – giving for instance way to the ‘chain arrangement’ in which earlier-born children are sent to school and use their wage earnings to invest in their younger siblings later on, rendering a quantity-quality trade-off superfluous (Baland et al., 2016; Mueller, 1984b). Third, in many of the least developed regions in SSA, the quality of schooling may be low (e.g. Chaudhury et al., 2006), or labor market opportunities may be lacking (e.g. Garcia and Fares, 2008), both of which depress the returns to education. Hence, additional schooling may not be a way to invest in children's ‘quality’. Finally, SSA still is the region with the highest fertility and lowest educational enrolments and attainments⁸, increasing the relevance of research on these issues.

Our use of DHS data comes with both pros and cons. Among the pros, we count the availability of demographic and health data of mothers, allowing us to explicitly control for factors such as

but this effects disappears for recent periods “when the opportunities for child farm work have declined, education has expanded, and fertility has declined to below-replacement levels” (p. 1473).

⁷ Bhalotra and Clarke (2016) tested the quantity-quality trade-off in a large sample, with data from 72 countries among which many Sub-Saharan African countries, but they did not specifically focus on SSA.

⁸ Data on 46 SSA countries reveal an average Total Fertility Rate (TFR) of 4.97 in 2014 (World Bank, 2016) and a Gross School Enrolment rate (GSE) of 98.96 in 2013 (UNESCO, 2016). The 2014 TFR is markedly lower than in 1985 (6.62) but still much higher than the 2014 TFR in South-Asia (2.56) and in Latin-America (2.11). Similarly, the 2013 GSE has greatly improved from its level in 1985 (79.24), but still compares unfavorably with the 2013 GSE in Southern Asia (109.06) Latin-America and the Caribbean (109.03).

ethnicity, mothers' height and health that are likely to affect twinning. In doing so, we further purge this instrument of potential sources of endogeneity. Second, the detailed information on children's health allows us to verify that parents do not allocate resources away from twins – who may suffer from poorer health at birth – toward older singleton-birth children, thus further providing confidence in our instrument. Third, the DHS data allows us to distinguish between three distinct proximate causes that underlie differences in educational attainment: school enrolment, school starting age, and dropout. Among the cons, we face the constraint that, beyond the fourth birth order, there are insufficient observations on twin births in the DHS to provide enough power in the first stage. Our analysis therefore only focuses on the effect of family size on outcomes of children of the first, second and third birth order, and its findings cannot readily be generalized to siblings at higher birth orders (Booth and Kee, 2009; Qian, 2009; Mogstad and Wiswall, 2016). Another point of attention is that the DHS focuses on mothers of childbearing age, 15-49, such that the observed number of children may be below the eventual number of children and the reported level of schooling may be below the eventual schooling attainment. Consequently, the relation that we observe between sibship and schooling captures a snapshot in time of a process in motion, not its final stage. Furthermore, the DHS provides no systematic information on health for children above five years old, nor information on test scores or other proxies for the quality of schooling. We therefore limit our analysis to the quantity of schooling, measured by the number of years of schooling. Finally, as will be explained further on, we need to address the complexity of family structure in SSA, which includes polygamy and a non-negligible number of children living outside the household, often with extended family.

In the next section we explain the empirical strategy. Then, we describe our data and present the results. In our results section, we find overall no evidence of a quantity-quality trade-off. In particular, we cannot reject the null hypothesis of no relation between family size and schooling in

the subsamples of families with two or more, or four or more children, while we find a significant *positive* effect of family size on schooling in the sample of families with three or more children.

2. Empirical strategy

In our empirical analysis, we use an age-standardized z-score⁹ for the years of schooling as our main outcome variable. We consider 3 sub-samples: firstborn from families with at least 2 children (2+Sample), first- and second- born from families with at least 3 children (3+Sample) and first-, second- and third- born from families with at least 4 children (4+Sample). We first examine the relation between schooling and family size using ordinary least squares (OLS). Then, we respectively use twinning at 2nd, 3rd and 4th birth to instrument the number of children in the ‘2+Sample’, ‘3+Sample’ and ‘4+Sample’. Focusing on the schooling outcomes of $n-1$ children born *prior* to twins at n^{th} birth, avoids selection problems that arise because “families who choose to have another child after a twin birth may differ from families who choose to have another child after a singleton birth” (Black et al. 2005). To ensure the validity of our instrument, we duly control for a battery of mother-level characteristics that may affect twinning and at the same time correlate with children’s schooling.

Concretely, the OLS specification takes the following form:

$$Education_{hmf i} = \beta_0 + \beta_1 \text{Number of children}_h + \beta_2 X_{hmf i} + \beta_3 X_{hm} + \beta_4 X_{hf} + \beta_5 X_h + \varepsilon_{hmf i} \quad (Eq. 1)$$

where h indicates household, m mother, f father and i the individual child. $Education_{hmf i}$ equals the child’s z-score. $Number\ of\ children_h$ is a count variable that equals the total number of sons and daughters of the household head, residing in the household; $X_{hmf i}$ is a vector of child-level

⁹ Since the variation in completed grades across country and by child age is so large, it is unlikely to be completely captured by country and age fixed effects. We therefore use age-standardized z-score for which the reference group comprises of children in the same country and birth cohort. In the robustness section, we show that our results remain stable when using completed grades as outcome variable.

characteristics that includes an indicator variable for sex, indicator variables for each age in the range 6 to 18, the child's birth-year and its month of birth; X_{hm} is the set of mother-level characteristics including her years of schooling, age, age squared, height, religion, ethnicity¹⁰, the total number of her children who have died and whether a child of her has died before its first birthday¹¹; X_{hf} is the set of father-level characteristics comprising his age and years of schooling; X_h includes household's residence area (urban/rural) and wealth quintile¹². To account for within-household correlation of the residuals, we cluster all error terms ε_{hmfi} at the DHS cluster level, which equate villages in rural areas and city blocks in urban areas.¹³

The second stage of the IV specification is captured by the following equation:

$$Education_{hmfi} = \delta_0 + \delta_1 \widehat{Number\ of\ children}_h + \delta_2 X_{hmfi} + \delta_3 X_{hm} + \delta_4 X_{hf} + \delta_5 X_h + \omega_{hmfi} \text{ (Eq. 2)}$$

in which family size is instrumented in the first stage equation:

$$Number\ of\ children_h = \alpha_0 + \alpha_1 Twin_h + \alpha_2 X_{hmfi} + \alpha_3 X_{hm} + \alpha_4 X_{hf} + \alpha_5 X_h + \vartheta_h \text{ (Eq. 3)}$$

In the n+Sample, the indicator variable $Twin_h$ equals 1 if the nth birth is a multiple birth and 0 otherwise¹⁴. The X-vectors include control variables as previously defined. ε_{hmfi} , ϑ_h and ω_{hmfi} are error terms.

In our results section, we scrutinize the exclusion restriction of our instrument, relying on insights of Altonji et al. (2005), Conley et al. (2012) and Bhalotra and Clarke (2016). In a series of nine

¹⁰ Mother's ethnicity is country specific and generated as follows: (country code x 1000) + ethnic group code. The insertion of the complete set of ethnicity fixed effects (as dummy variables) makes country fixed effects superfluous. In some DHS rounds, ethnicity is omitted (e.g. Rwanda, Burundi). These rounds are omitted from our baseline results, but included in a robustness check that uses region of residence of the household as a proxy for ethnicity.

¹¹ Infant deaths proxy among others the reproductive health of the mother.

¹² For each round, DHS separates all interviewed households into five wealth quintiles based on their wealth index. The wealth index is a composite measure of a household's cumulative living standard. It is calculated using principal components analysis on easy-to-collect data on a household's ownership of selected assets, such as televisions and bicycles; materials used for housing construction; and types of water access and sanitation facilities (Standard Recode Manual for DHS 6, 2013).

¹³ The results are robust to clustering at the ethnicity or region level. The number of countries is insufficient to get a robust covariance matrix when clustering at the country level.

¹⁴ For instance, in the 3+Sample, the indicator variable $Twin_h$ equals 1 if the third birth is a multiple birth and 0 otherwise.

robustness checks (cf. *infra*), we modify this empirical framework in several ways, using alternative samples, changing the decision unit (from household head to wife/wives), and modifying the definition of our key variables.

In all cases, we report heteroscedasticity-robust statistics and the usual post-estimation tests (underidentification test, weak identification test and overidentification test).

3. Data

In our empirical analysis, we rely on all DHS surveys implemented in SSA countries in the period 1990-2014, for which we could construct the main variables.¹⁵ In our baseline approach, we consider 59 survey rounds for which information on the ethnicity of the mother is available (Appendix 1 gives an overview of these survey rounds by country and year) and restrict the sample to children whose siblings of schooling age (6-18) all reside within the household.¹⁶ This gives a dataset of 456,068 siblings of schooling age (6-18), to which we will refer as ‘Sample I’.

We focus on the educational attainments of 64,339 firstborn (2+Sample); 99,875 first- and second-born (3+Sample) and 101,848 first-, second- and third- born children (4+Sample) in the age group 6 to 18, who live with their parents. The lower-limit of 6 is the age at which many children in SSA start primary schooling. The upper-limit of 18 is the age of secondary schooling completion, provided a swift grade progression. We do not extend the upper-limit beyond 18 because post-secondary education in SSA still faces important supply side constraints, and because a considerable proportion of children above 18 live outside the household such that their schooling attainment goes unrecorded in the DHS surveys.

¹⁵ In the period 1990-2014, 86 DHS survey rounds were administered in 34 SSA countries. For the following surveys, we could not construct and merge the key variables because of issues with the unique identifiers of surveyed individuals: 1991 and 1998 in Cameroon, 1996-1997 in Chad, 1998-1999 in Cote d’Ivoire, 2001 in Mali and 1992 in Niger.

¹⁶ A household is defined as one person or a group of persons who usually live and eat together. This is not the same as a family. A family includes people who are related, but a household includes any people who live together, whether or not they are related (DHS Interviewer’s manual, October 2012).

To capture the educational attainment of these children, we look at their completed years of schooling at the time of the survey and then construct age-standardized education z-score with children of the same country and birth cohort as reference group. Our explanatory variable of interest is the number of children in a household. In our baseline specification, we define this variable as the total number of the household head's sons and daughters residing in the household.

Our instrumental variables are the birth of twins at second, third and fourth birth order. In the DHS birth records, there is a specific variable indicating whether a child is part of a twinship or not. To determine birth order, we consider all children of the household head including those who do not have their mother in the household.

The summary statistics of the principal variables in our analysis can be consulted in Table 1. The summary statistics of the other variables can be consulted in Table A1 of the online Supplementary Appendix.

4. Baseline Results

4.1 Baseline OLS and IV estimation

The estimation of the OLS model (Eq.1) yields a negative and significant relation between family size and children's schooling. As shown in Table 2, the estimated coefficient is rather small in magnitude: one additional child is associated with a reduction of about 0.023 units of the z-score which corresponds to 0.057 years of schooling for a child of 10 years of age.¹⁷ As this result does not isolate a causal link between our variables of interest, we turn to our IV estimations.

The estimates of the first stage (Eq.3) are shown in the second panel of Table 3. Unsurprisingly they indicate that twinning increases average family size. The effect of twinning ranges from an

¹⁷ The coefficients are multiplied by the standard deviation of completed years of education in Sample I (see Table 1) and interpreted with reference to the average age of children in Sample I.

additional family size of 0.356 (in 4+Sample) to 0.509 (in 3+Sample). The coefficients are precisely estimated, significant at 1%, and similar to the range of coefficients found in previous research.¹⁸ In all cases, the twin instrument has reassuring first stage post-estimation statistics, with the Cragg-Donald Wald F-statistic well above 100.

In contrast to the OLS estimates that are uniformly negative in all 3 sub-samples, the second-stage IV estimates (first panel of Table 3) indicate either no impact (in the 2+Sample and 4+Sample) or a positive and significant effect (at 5%) of family size on education z-scores (in the 3+Sample). In the 3+Sample, a one-unit increase in predicted family size increases the z-score by 0.097 units on average which is the equivalent of 0.240 years of schooling for a child of 10 years of age. We demonstrate the robustness of these IV results in Section 5, and tentatively explore plausible explanations in Section 6.

4.2. On instrument validity

A concern when implementing the IV estimation is the violation of the exclusion restriction. The exclusion restriction may be threatened because of the presence of confounding factors, that affect both twinning and children's schooling. In particular, besides a mother's age, ethnicity and height, also less easily measurable characteristics such as her general health condition may affect the probability of twinning (Smits and Monden, 2011). In theory, a mother's health behaviour, in particular smoking and multiple-birth-enhancing fertility treatments could pose another threat to our instrument's validity, but in practice this threat is neutralized in SSA due to its prohibitively high (social and monetary) cost (Inhorn, 2003).

Table 4 explores the determinants of twinning in our sample. The first column shows the results of a regression of twinning on the mother-level characteristics included in our baseline model: her

¹⁸ For instance Bhalotra and Clarke (2016), Marteleto and de Souza (2012), Angrist et al. (2010) and Black et al. (2005).

years of schooling, age, age squared, height, ethnicity, religion, the total number of her children who have died and a dummy variable capturing whether these children died before their first birthday. Out of these 8 characteristics, only mother's religion does not significantly affect the probability of twinning at third birth. However, the overall explanatory power of the regression is very low (R^2 of 0.009). The second column adds four additional mother-level regressors - mothers' BMI, access to prenatal health care, access to a doctor and access to a nurse.¹⁹ All four additional regressors turn up significant while mother's education becomes non-significant.

Based on subsamples of individuals for which all additional four control variables are available, Table 5 shows how the estimated coefficient of interest changes when adding the additional four control variables to our baseline controls. Using the baseline specification, the estimates are 0.003 in the 2+Sample (Panel A, column I), 0.137** in the 3+Sample (Panel B, column I) and -0.043 in the 4+Sample (Panel C, column I). When adding the four additional mother-level controls, our coefficient of interest only slightly decreases in all three panels, to -0.005, 0.126** and -0.061, respectively. Thus, without the additional controls, our estimates are (slightly) biased upward, which is expected if the included mother characteristics positively correlate with both twinning and children's schooling. In the spirit of Altonji et al. (2005), we however argue that, if observed additional controls change the value of our estimated coefficients only to such a small extent, then it is unlikely that there exist unobserved controls that would turn our results upside down.²⁰

To further safeguard our results, we follow Conley et al. (2012) and Bhalotra and Clarke (2016) in deriving bounds for our coefficient of interest using the 'Plausexog' command in Stata. To do so, we first acquire insight into the direct effect of twinning on children's schooling (γ) by simply

¹⁹ To measure access to prenatal care, we use the percentage of births (occurring within 5 years prior to the survey) with prenatal care in the DHS cluster. Access to a doctor or nurse are proxied by the DHS cluster-level percentage of births with prenatal care given by a doctor or nurse.

²⁰ We do not formally verify this intuition, since the formalized extension by Oster (2017) only applies to OLS estimates.

comparing education z-scores of children from twin-mothers to those of children from non-twin mothers, controlling for the set variables in our baseline specification (see Table A2 in Appendix). We then take the upper value of the estimated 95% confidence interval which is 0.005.²¹ The standard deviation of γ (sd=0.007) is obtained from a 100 replications bootstrap estimation of γ . Table 6 shows our bounds' estimates of the family size effect on children's schooling using Conley's Union of Confidence Interval (UCI) and Local To Zero (LTZ) approaches. The results point to a rejection of a quantity-quality trade-off in all three samples, and a confirmation of the positive effect of family size on children schooling in the 3+Sample.

Another potential violation of the exclusion restriction stems from parental behavior towards twins due to twins' lower average birth weight and the potential consequences for their health or cognitive achievements later on (Rosenzweig and Zhang, 2009). If the future earning potential of twins is thought to be lower, parents may divert resources from twins to singletons. Rosenzweig and Zhang (2009) suggest to include twins in the analytical sample and to include birth weight in the regression in order to control for this potential bias. This could indeed be a straightforward solution, were it not that the DHS only includes birth weight for the under-five year olds. As a second best, we verify whether this 'diversion of resources' hypothesis is supported in the sample of under-five year olds.

Using information on 70,902 under-five years olds in our 59 DHS waves, we regress *birth weight* (in log terms) on a twin indicator and find that twins indeed have a significantly lower average birth weight than singletons (see column I of Table A3 in the Supplementary Appendix).²² When looking at the weight (column II) and body mass index percentile (column III) of under-five year olds *at*

²¹ Note that our prior value on γ is more conservative than the 0.004 used in Bhalotra and Clarke (2016). Our results hold even with γ set at 0.008.

²² We control for several characteristics likely to affect birth weight such as a mother's age, education, body mass, age at first birth and ethnicity; and the child's sex, preceding birth interval and year of birth; and the household's residence and wealth quintile. All error terms are clustered at the DHS cluster level.

the time of the survey, we still find that twins have significantly lower weight and body mass index (BMI) than singletons.²³ However, the estimated coefficient is smaller, indicating that the gap has become smaller over time. In fact, when looking at the BMI, twins and singletons belong to the same decile on average.²⁴ The closing of the gap suggests that parents do not divert resources away from twins. Furthermore, controlling for the entire set of age dummies, wealth quintiles, region of residence, child sex, parental education and ethnicity fixed-effect, we find that twins enjoy as much education as singletons (Column IV of Table A3 in the Supplementary Appendix).

4.3. Heterogeneity

As mentioned in the introduction, the studies that have set out to identify the causal relationship between family size and schooling have produced mixed results, suggesting that the relation is context dependent. Hence, we explore heterogeneity of our result, by running separate regressions across subsamples with respect to gender, poverty status and region.

To explore regional variation in the estimated relation, we contrast West and Central Africa with East and Southern Africa. This division is inspired by the regional clustering of TFR, which is on average relatively high in West and Central Africa (5.09), and lower in East and Southern Africa (4.45). To compare across poor and non-poor, we define households in the 1st and the 2nd asset ownership quintiles as poor and those in the 4th and the 5th quintiles as non-poor. We discard the 3rd quintile to achieve a sharper contrast between poor and non-poor. The results are lined up in Table 7. In our discussion below, we highlight the various significant coefficient estimates.

²³ We regress BMI and BMI percentile on the twin indicator, controlling for the factors mentioned in the previous footnote, as well as for birth weight, child age, number of months of breastfeeding, number of under-5 year olds in the households and whether or not the mother lives with her husband.

²⁴ The difference in BMI is measured in terms of percentiles. A child is considered underweighted if its BMI is below the 10th percentile of the World Health Organization's reference BMI distribution (see Cole et al, 2007). According to this definition, 13.75% of twins are underweighted compared to only 10% of singletons. But, as shown in Table 3, on average twins and singletons belong to the same decile.

When only looking at the schooling of boys, we find a sizeable and significantly positive coefficient estimate in the 3+Sample (0.122**). For girls, the estimated coefficient is found to be negative and (slightly) significant (-0.183*) in the 4+Sample. In the subsample of poor households, the estimated coefficient is insignificant across the board. For the subsample of non-poor, we observe a significantly positive estimate in the 3+Sample (0.139***). In the regional subsamples, the estimated family size coefficient appears to be only slightly significant (and positive) in the 3+Sample in West and Central Africa (0.108*). It remains non-significant elsewhere.

When comparing the effect of family size on schooling across countries with persistently high fertility²⁵ and countries with either a low fertility or a downward fertility trend²⁶, we find a non-significant effect in low-fertility countries in all 2+, 3+ and 4+Samples, but a positive and slightly significant (of 0.128*) effect in the 3+Sample of high-fertility countries.

Finally, we explore whether the relation between family size and schooling has changed over time. To do so, we focus on the year 2000, which marked the adoption of the Education For All initiative by 189 countries.²⁷ We compare the effect of family size on schooling across children born prior to 2000 and children born from 2000 onwards. In none of the samples, we find a differential effect of family size over time, but our results indicate that the association of parents' education, gender and residence area with children's educational outcomes has weakened considerably over time, suggesting a democratization of schooling (see Table 8).

²⁵ Burkina Faso, Burundi, Congo DR, Guinea, Malawi, Mali, Niger, Nigeria, Chad (see Lesthaeghe, 2014, p. 2)

²⁶ Ghana, Lesotho, Liberia, Madagascar, Namibia, Rwanda, Senegal, South Africa, Swaziland, Uganda (see Lesthaeghe, 2014, p. 2)

²⁷ The initiative aimed at a global commitment to provide quality basic education for all children, youth and adults, and was first launched in 1990 (<http://www.worldbank.org/en/topic/education/brief/education-for-all>). Admittedly, children born in 2000 reached schooling age only in 2006. On the other hand, the initiative's adoption itself is unlikely to have had an immediate impact on the ground. In any case, any cut-off year would be somewhat arbitrary. Choosing slightly different cut-off years (1998 and 1999 which give almost balanced samples across cohorts) yields similar results.

In sum, in our various subsample analyses, the zero-result is confirmed for the 2+ and the 4+Sample (with the notable exception of the subsample of girls in the 4+Sample²⁸), while the positive result for the 3+Sample is shown to be mainly driven by children from non-poor families, living in high-fertility countries. We will come back to this latter result in Section 6. Now we first discuss a series of robustness checks.

5. Robustness checks

We check the robustness of our results in ten different ways. Table 9 gives a line-up of the estimated coefficients on family size for each check in the 2+, 3+ and 4+Samples. The full results are reported in Tables A4-A13 of the Supplementary Appendix, for the total sample, as well as for the poor and non-poor subsamples.

In the first robustness check, we follow Angrist et al. (2010) in allowing for heterogeneity across subsamples in the predictive power of twin birth in the first stage. We do so by including interaction terms in our first stage regressions between twin birth and a set of indicator variables (rural, mother is Muslim, rural West & Central Africa), thus sequentially adding the following regressors: $Twin_h * Rur_h$; $Twin_h * Mother_Islam_m$ and $Twin_h * Rural_West\&Central_Africa_h$. The rationale for including these regressors is that fertility is higher in rural areas (compared to urban areas), in Muslim families, and in West and Central Africa (compared to East and Southern Africa), and the twin instrument tends to perform less well in larger families (Angrist et al., 2010).²⁹ When including

²⁸ Family size might have a different impact on girls' schooling (no impact or negative impact) because of time spent in households chores by older daughters (Mueller, 1984; Tiefenthaler, 1997) which might negatively affect their educational achievements.

²⁹ Also following Angrist et al. (2010), we try out a second instrument, i.e. sibship sex-composition. This alternative instrument relies on the idea that parents prefer to have both boys and girls rather than only children of the same sex. Hence, in the latter case, they may be more likely to have an additional child. However, in our case, only twinning performs well in the first stage. For instance, where the F-statistic surpasses 100 for twinning, it barely reaches 20 for the sex composition instrument (results not reported, but available on request). We therefore focus on twinning in our analysis.

all three interaction terms, the zero-result remains in the 2+ and 4+Samples, while we still find a positive and significant in the 3+Sample.

In the second robustness check, we use an unrestricted specification with the ‘partially missing instruments’ method as described in Mogstad and Wiswall (2016, p.174:175). The partially missing instruments are constructed based on a polynomial function of mother age, mother education, father age and father education, controlling for child characteristics, households characteristics, mother ethnicity and her health conditions. This change in specification does not alter our results: family size remains insignificant in the 2+ and 4+Samples and significantly positive in the 3+Sample, in particular when captured by the indicator variable ‘more than 3 children’ that apprehends the marginal effect of moving from 3 to 4 children.

Third, instead of restricting the sample to those children that are part of households where all children reside within the household, we expand the sample to include also children that live in households where one or more school-aged siblings reside outside the household. This approach yields a 2+Sample of 68,259 children ; a 3+Sample of 112,285 children and a 4+Sample of 119,544 children. We find that the coefficient on family size is still positive and slightly significant in the 3+Sample, positive and non-significant in the 2+Sample, but turns negative and slightly significant in the 4+Sample.

Fourth, recognizing the complexity of SSA households, we change the decision unit. Among the 525,646 children in our Sample I, we count 165,418 living in polygamous households. While our baseline approach considers the household head as the unit of decision making, in this robustness check we assume decisions to be taken at the level of each mother. In this decentralized approach, the number of children is defined for each of the household head’s wives as her total number of sons and daughters living in the household. Birth order is also defined at the level of the mothers. Doing so, we find that family size loses its significance in the 3+Sample, and remains so in the 2+ and 4+Samples.

In checks five, six and seven, we use alternative definitions of our key variables. Instead of defining the number of children as the sons and daughters of the household head, we define them as the total number of births given by the household head's wives (unless the household head is female). To reduce measurement error in our schooling variable (for instance parents reporting years in kindergarten as schooling), education z-score are based on censored years of education.³⁰ And, estimates using completed grades (years of schooling) rather than educational z-scores are provided in robustness check seven. Our findings remain similar in all three cases: no significant effect in the 2+ and 4+Samples and a positive and significant effect in the 3+Sample.

In robustness checks eight and nine, we use region of residence of the household and country-by-urban/rural fixed effects instead of mother's ethnicity fixed-effects to take into account DHS rounds in which ethnicity is not included (e.g. Rwanda, Burundi).³¹ In the former case, the estimated coefficient loses significance in the 3+Sample. Apart from that, the results remain qualitatively similar: a positive and significant coefficient in the 3+Sample and non-significant ones in the 2+ and the 4+Samples.

Finally, we use an alternative definition to discriminate between poor and non-poor families, defining the poverty line as the average value of the wealth index in each DHS round. This last approach confirms the positive effect observed only in non-poor families of the 3+Sample and the absence of effect both in poor and non-poor families across 2+ and 4+Samples.

Overall, our results remain fairly robust in all three subsamples. The zero-result in the 2+Sample remains so across all ten robustness checks while it becomes significantly negative in the 4+Sample in only one case (inclusion of households in which some school-aged children reside outside the

³⁰ We censor the completed years of education to the child's age minus 6, assuming that schooling can start only from 6 years onwards since children are admitted to primary school at 6 in most of the countries (from <http://data.worldbank.org/indicator>). For instance, if a 12-year old in our sample is reported to have completed 9 years of schooling, we set years of schooling to 6.

³¹ This yields a more balanced sample between West & Central Africa (57.80%) and East & Southern Africa (42.20%).

household). In the 3+Sample, the positive coefficient is no longer significant in only two cases (decentralized approach and region of residence of the household instead of mother's ethnicity fixed effect). When running the robustness checks on the non-poor sample, the estimated coefficient on family size in the 3+Sample remains positive and significant across the board.³²

Taken together, these results bolster the case against a quantity-quality trade off in SSA, when quality is measured as educational attainment. At the same time, however, we note the heterogeneity of coefficient estimates, not only across gender, asset wealth and region, but also across the 2+, 3+ and 4+Samples. Whether or not this heterogeneity is a statistical artefact needs to be determined by future work.

6. Mechanisms

To further guide future work we fully exploit the DHS data, in order to provide some cues for the possible mechanisms underlying our results and their heterogeneity

First, we use the DHS data to distinguish between three proximate causes that underlie differences in educational attainment: school enrolment, school starting age and dropout. We explore these proximate causes in all three analytical samples, by estimating (Eq. 2) and (Eq.3) with school enrolment, school starting age and dropout as dependent variables. Table 10 lines up the coefficients of interest. The full results are given in Tables A14-A20 of the Supplementary Appendix.

When using the total sample and the subsample of poor households, we find a zero-effect of family size on enrolment, dropout and school starting age in the 2+, 3+ and 4+ Sample, with one exception (in the total 3+Sample, the dropout of second born is slightly reduced). For the

³² In the case of the 4+Sample, we obtain a negative and slightly significant (at 10%) coefficient in the subsample of non-poor households when using censored education z-scores (Panel C of Table A9) and when using the household's region of residence instead of mother's ethnicity (Panel C of Table A11).

subsample of non-poor households, we find various significant coefficients. In the 2+Sample, firstborn's school starting age is reduced (-0.499*) with an exogenous increase in family size. In the 3+Sample, it are second born that seem to start school earlier on. A closer examination of this effect reveals that it is largest and significant when the second born is relatively close in age with the first born (3 years apart or less³³) (see column III Table 10). In the 4+Sample, we find that family size reduces the probability of enrolment of the second born (-0.072*). For the third born, results show a reduction in the probability of dropout (-0.076*)³⁴.

Overall, this tentative exploration of the proximate causes suggests that, in response to an exogenous increase in family size at birth order n , relatively small and wealthy households tend to send the $n-1^{\text{th}}$ child earlier to school, a finding that is not replicated in poor households. This could indicate that, when financially possible, some households opt to speed the schooling of earlier born children upon twin birth. Whether this finding can be broadly replicated, and whether it is explained by an attempt on the part of parents to maximize economies of scale³⁵, or simply to relieve the caregiver so s/he can focus on younger siblings, is a question left for future research. The non-linearity of economies of scale (see e.g. Holmes and Tiefenthaler, 1997, Tiefenthaler, 1997) together with the negative effect of reduced care time on children cognitive abilities (Lehmann et al, 2018) may account for the heterogeneity across the 2+, 3+ and 4+ samples.

Next to economies of scale, the introduction of the article mentioned three other mechanisms that could explain the absence of a quantity-quality trade-off, or even a positive effect of family size on schooling: child labor, the chain arrangement, and support from the extended family. While we do

³³ This result remains significant but is lower when considering an age difference of ' ≤ 4 years', but it loses significance when the age difference is 2 years or less, both in the 3+ and 2+Samples.

³⁴ Enrolment, school starting age and dropout of first born in the 4+Sample are not affected by family size (results not reported).

³⁵ When sharing books, tutoring or transport to school, the per child cost of schooling declines when two or more children can be sent to school simultaneously (Qian, 2009).

not have the data to explore the likelihood of the latter two mechanism, we tentatively discuss (and dismiss) the role of child labor.

Child labor, both at home and in the labor market,³⁶ still is common in many SSA countries, but the group of children that are working is increasingly made up of children who combine part-time employment and schooling (Guarcello et al., 2015). The combination of work and schooling may allow for child labor to contribute to schooling rather than crowd it out, by providing resources for schooling fees, their own as well as those of their siblings. Should child labor explain the positive effect of family size in the 3+Sample, we would however expect the effect to be larger in poor households and lower in non-poor ones, given that the latter rely less on resources provided by children. Instead, we find the reverse. To test more formally for the child labor mechanism, we exploit the available child labor information in a subset of the DHS surveys.³⁷ If child labor contributes to schooling and explains the positive effect (or zero-effect) of family size, the estimated coefficient on family size would be reduced after controlling for child labor in our model. Results in Table A21 of the Supplementary Appendix show that, rather than attenuating the positive effect of family size, the inclusion of child labor (both own labor of the labor of its siblings) slightly reinforces the positive effect in 3+Sample while it leaves the estimated coefficients in the 2+ and the 4+Samples almost unaltered.

7. Conclusion

The aim of this study was to test the quantity-quality trade-off in SSA, focusing on children's schooling. To do so, we investigated how an increase in family size affects schooling using twin

³⁶ In our analysis child labor is broadly defined as the sum of time spent in household chores, number of working hours for a family member (including him/her self) or for someone outside the household. In an alternative definition, we exclude household chores; our findings remain robust to this change.

³⁷ In 21 surveys of 15 countries, information on children's time allocation is provided (the rounds and countries are listed in Appendix 1). The average total hours worked by a child, including time spent on household chores, is 9.78 per week and is significantly larger in poor families in a sample mean t-test: 11.73 hours compared to 7.59 in richer ones.

birth as an instrumental variable to deal with endogeneity issues. Overall, we find no significant effect of family size on children's schooling, thus casting doubt on the generally assumed negative relation between family size and schooling.

In the subsample of first and second born from relatively rich households with three or more children, we find a positive effect of family size on schooling and this effect survives various robustness checks. To exclude that this result is a statistical artifact, its replication in other samples is required.

In a tentative exploration of the underlying mechanisms, it emerges that upon a fertility shock at the n^{th} birth, relatively small and rich families tend to send their $n-1^{\text{th}}$ born to school relatively early on. Doing so may optimize economies of scale in schooling and/or relieve the caregiver, although both the replicability of the finding and its explanation need further study.

Exploring heterogeneity of our results, we find that the effect of family size does not vary substantially across time. This is in sharp contrast to the role of other factors - parental education, gender, residence area (urban/rural) – that significantly decreased over time, in line with the 'democratization' of education. Only in the 3+Sample we find a regional difference: family size positively impacts children's schooling in countries with persistently high fertility while we find no effect in countries with low or declining fertility.

Our research suffers from a number of limitations. First, the positive impact of family size on schooling of first and second born children in the 3+Sample cannot readily be generalized to higher parity children, which is clear from our results in the 4+Sample. Second, the DHS provides only a snapshot in time of children whose mothers are of childbearing age (15-49). The number of children observed as well as their schooling attainment reflect therefore only an intermediate situation, not the final one. This leaves open the possibility that, in the longer run, the positive effect of family size on schooling in non-poor households with 3 or more children (driven by early

enrolment of the second born) fades away. Third, the available data are not well suited to distinguish between competing underlying mechanisms to the differential effect of family size across samples. The short-term horizon does not allow for explicitly testing of the chain arrangement. And, lacking more detailed data on household consumption and transfers, we cannot thoroughly test for the economies of scale and extended family mechanisms. Finally, the number of years of schooling is only one way in which parents can invest in their children. Important omitted dimensions include the quality of schooling, and health care.

These gaps should be filled by future research, relying on other types of data such as pooled census data in which families and their split-offs are traced over time, and micro-economic surveys that provide detailed information on household members' consumption of schooling inputs and their time allocation, as well as surveys that include more information on health and the quality of schooling of a children in various age cohorts. A more open line of questioning, in qualitative research, could also reveal the reasoning underlying parent's decision making.

Compliance with Ethical Standards:

No conflict exists. The authors declare that they have no conflict of interest.

References

- Altonji JG, Elder TE, Taber CR (2005) Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. *Journal of Political Economy*, 113: 151–184.
- Angrist JD, J-S Pischke (2009) *Mostly Harmless Econometrics: An Empiricist's Companion*, 1 edition. ed. Princeton, NJ: Princeton University Press.
- Angrist J, Lavy V, Schlosser A (2010) Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28(4), pp.773-824.
- Baland JM, Bonjean I, Guirkinge C, Ziparo R (2016) The economic consequences of mutual help in extended families. *Journal of Development Economics*, 123, pp. 38-56
- Bass LE (2004) *Child labor in sub-Saharan Africa*. Lynne Rienner Publishers.
- Becker GS (1960) An Economic Analysis of Fertility. In: National Bureau of Economic Research (ed.). *Demographic and Economic Change in Developed Countries*. Princeton: Princeton University Press: 209-231.
- Bhalotra S R, Clarke D (2016) The twin instrument. IZA Discussion Paper No. 10405.
- Black SE, Devereux PJ, Salvanes KG (2005) The more the merrier? The effect of family composition on children's education. *Quarterly Journal of Economics*, 120, pp. 669–700.
- Black SE, Devereux PJ, Salvanes KG (2010) Small family, smart family? Family size and the IQ scores of young men. *Journal of Human Resources*, 45(1), pp.33-58.
- Blake J (1989) *Family Size and Achievement*. Berkeley: Univ. Calif. Press
- Bongaarts J (2009) Human population growth and the demographic transition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1532), pp.2985-2990.

Booth AL, Kee HJ (2009) Birth order matters: the effect of family size and birth order on educational attainment. *J Popul Econ* 22:367–397

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

Chaudhury N, Hammer J, Kremer M, Muralidharan K, Rogers FH (2006) "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *Journal of Economic Perspectives*, 20 (1): 91-116.

Cole TJ, Flegal KM, Nicholls D, Jackson AA (2007) Body mass index cut offs to define thinness in children and adolescents: international survey. *BMJ : British Medical Journal*, 335(7612), 194. <http://doi.org/10.1136/bmj.39238.399444.55>

Conley D, Glauber R (2006) Parental educational investment and children's academic risk estimates of the impact of sibship size and birth order from exogenous variation in fertility. *Journal of Human Resources*, 41(4), pp.722-737.

Conley, T. G., Hansen, C. B., Rossi, P. E. (2012) Plausibly exogenous. *Review of Economics and Statistics*, 94(1), 260-272.

Deaton A, Paxson C (1998) Economies of scale, household size, and the demand for food. *Journal of Political Economy*, 106(5), pp.897-930.

Fitzsimons E, Malde B (2014) Empirically probing the quantity–quality model. *J Popul Econ* 27: 33-68.

Foster A, Roy N (1997) The dynamics of education and fertility: Evidence from a family planning experiment. Unpublished manuscript.

- Garcia M, Fares J (2008) Youth in Africa's labor market (English). Directions in development ; human development. Washington, DC: World Bank. Gibbs, B., Workman, J., and Downey D (2016) The (Conditional) Resource Dilution Model:State- and Community-Level Modifications. *Demography* (2016) 53:723–748
- Guarcello L, Lyon S, Valdivia C (2015) Evolution of the relationship between child labour and education since 2000 Evidence from 19 developing countries. Working Paper March 2015. Understanding Children's Work (UCW) Programme.
- Guo G, Van Wey L (1999) Sibship size and intellectual development: is the relationship causal? *American Sociological Review* 64:169–87.
- Holmes, J., & Tiefenthaler, J. (1997). Cheaper by the dozen? The marginal time costs of children in the Philippines. *Population Research and Policy Review*, 16(6), 561-578.
- Inhorn MC (2003) Global infertility and the globalization of new reproductive technologies: illustrations from Egypt. *Social science & medicine*, 56(9), pp.1837-1851.
- Joshi S, Schultz TP (2007) Family planning as an investment in development: evaluation of a program's consequences in Matlab, Bangladesh. Yale University Economic Growth Center Discussion Paper, (951).
- Lee J (2008) Sibling size and investment in children's education: an Asian instrument. *J Popul Econ* 21: 855–875
- Lehmann, J. Y. K., Nuevo-Chiquero, A., & Vidal-Fernandez, M. (2018). The early origins of birth order differences in children's outcomes and parental behavior. *Journal of Human Resources*, 53(1), 123-156.

Lesthaeghe R (2014) The fertility transition in sub-Saharan Africa into the 21st century. Ann Arbor, MI: University of Michigan.

Li H, Zhang J, Zhu Y (2008) The quantity-quality trade-off of children in a developing country: Identification using Chinese twins. *Demography*, 45(1), pp. 223-243.

Liu H (2014) The quality–quantity trade-off: evidence from the relaxation of China’s one-child policy. *J Popul Econ* 27: 565-602.

Lloyd CB, Blanc AK (1996) Children's schooling in sub-Saharan Africa: The role of fathers, mothers, and others. *Population and development review*, 22(2), pp.265-298.

Maralani V (2008) The changing relationship between family size and educational attainment over the course of socioeconomic development: Evidence from Indonesia. *Demography*, 45(3), pp.693-717.

Marteletto LJ, de Souza LR (2012) The changing impact of family size on adolescents’ schooling: Assessing the exogenous variation in fertility using twins in Brazil. *Demography*, 49(4), pp.1453-1477.

Marteletto LJ, de Souza LR (2013) The implications of family size for adolescents' education and work in Brazil: Gender and birth order differences. *Social forces*, 92(1), pp.275-302.

Mogstad M, Wiswall M (2016) Testing the quantity–quality model of fertility: Estimation using unrestricted family size models. *Quantitative Economics*, 7, pp. 157–192.

Mueller E (1984a) The value and allocation of time in rural Botswana. *Journal of Development Economics*, 15(1), pp.329-360.

Mueller E (1984b) Income, aspirations, and fertility in rural areas of less developed countries. In W. A. Schutjer & C. Shannon Stokes (Eds.), *Rural development and human fertility* (pp. 121–150). New York: Macmillan.

Narayan D, Pritchett L. (1999) Cents and Sociability: Household Income and Social Capital in Rural Tanzania, *Economic Development and Cultural Change*, 47(4): 871-897.

Oster E (2017) Unobservable selection and coefficient stability: Theory and validation. *Journal of Business Economics and Statistics*. Forthcoming.

Qian N(2009) Quantity-quality and the one child policy: The only-child disadvantage in school enrollment in rural China (No. w14973). National Bureau of Economic Research.

Rosenzweig MR, Wolpin KI (1980) Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica: journal of the Econometric Society*, pp.227-240.

Rosenzweig MR, Zhang J (2009) Do population control policies induce more human capital investment? Twins, birth weight and China's “one-child” policy. *The Review of Economic Studies*, 76(3), pp.1149-1174.

Sinha N (2005) Fertility, child work, and schooling consequences of family planning programs: Evidence from an experiment in rural Bangladesh. *Economic Development and Cultural Change*, 54(1), pp.97-128.

Smits J, Monden C (2011) Twinning across the developing world. *PLoS One*, 6(9), p.e25239.

Steelman LC, Powell B, Werum R, Carter S (2002) Reconsidering the effects of sibling configuration: Recent Advances and Challenges. *Annual Review of Sociology* 28:243–69.

Tiefenthaler, J. (1997). Fertility and family time allocation in the Philippines. *Population and Development Review*, 377-397.

UIS Stat. UNESCO. 2016. <http://data.uis.unesco.org>

World Bank (2016) World Development Indicators.

<http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators>

Accessed 6 February 2016.

Tables

Table 1: Sample means and proportions of main variables

Variables	SAMPLE I	2+SAMPLE	3+SAMPLE	4+SAMPLE
Number of children	5.30 (2.54)	3.86 (1.53)	4.39 (1.46)	5.14 (1.43)
Child completed years of schooling	2.05 (2.47)	2.47 (2.67)	2.29 (2.52)	2.31 (2.55)
Child Education Z-score	0.00 (1.00)	0.156 (1.04)	0.104 (1.02)	0.086 (1.02)
Child age	10.09 (3.04)	10.38 (3.11)	10.19 (2.94)	10.44 (2.96)
Child sex				
<i>Male</i>	<i>51.76</i>	<i>51.03</i>	<i>51.36</i>	<i>51.68</i>
<i>Female</i>	<i>48.24</i>	<i>48.97</i>	<i>48.64</i>	<i>48.32</i>
Month of birth	6.05 (3.33)	6.05 (3.36)	6.05 (3.34)	6.06 (3.33)
Mother age	35.39 (6.31)	30.17 (4.91)	30.91 (4.74)	32.05 (4.66)
Mother education (single years)	2.65 (3.84)	3.61 (4.42)	3.20 (4.18)	2.80 (3.90)
Mother religion				
<i>Mother is muslim</i>	<i>43.49</i>	<i>38.53</i>	<i>40.33</i>	<i>42.22</i>
<i>Mother is not muslim</i>	<i>56.51</i>	<i>61.47</i>	<i>59.67</i>	<i>(57.78)</i>
Father age	44.73 (9.43)	38.18 (7.76)	39.23 (7.79)	40.79 (7.88)
Father education (single years)	4.16 (4.85)	5.22 (5.20)	4.83 (5.08)	4.44 (4.93)
Residence				
<i>Urban</i>	<i>26.39</i>	<i>31.98</i>	<i>29.32</i>	<i>26.93</i>
<i>Rural</i>	<i>73.61</i>	<i>68.02</i>	<i>70.68</i>	<i>73.07</i>
Wealth quintile				
<i>Poorest</i>	<i>25.00</i>	<i>22.10</i>	<i>23.62</i>	<i>25.36</i>
<i>Poor</i>	<i>21.21</i>	<i>19.58</i>	<i>20.24</i>	<i>21.05</i>
<i>Middle</i>	<i>19.77</i>	<i>18.76</i>	<i>19.38</i>	<i>19.56</i>
<i>Rich</i>	<i>18.03</i>	<i>18.60</i>	<i>18.39</i>	<i>18.09</i>
<i>Richest</i>	<i>16.00</i>	<i>20.97</i>	<i>18.37</i>	<i>15.94</i>
Twin birth				
<i>Twin at 2nd birth</i>	<i>1.10</i>	<i>1.37</i>	<i>1.24</i>	<i>1.13</i>
<i>Twin at 3rd birth</i>	<i>1.53</i>	<i>1.60</i>	<i>1.61</i>	<i>1.71</i>
<i>Twin at 4th birth</i>	<i>1.64</i>	<i>1.88</i>	<i>1.82</i>	<i>1.93</i>
<i>Twin at 5th birth</i>	<i>1.89</i>	<i>2.28</i>	<i>2.16</i>	<i>2.12</i>
N (number of observations)	456,068	64,339	99,875	101,848

Source : Authors, based on data from 59 DHS rounds. The twin indicators are expressed in percentages of the total number of households, while the other percentages relate to the total number of children. Sample I includes all children of schooling age (6-18). 'N+Sample' is composed of lower birth order children from families with N or more children, whose siblings of schooling age (6-18) all reside within the household.

Table 2: OLS estimates

	DEPENDENT VARIABLE : EDUCATION Z-SCORE		
	2+SAMPLE	3+SAMPLE	4+SAMPLE
Number of children	-0.023*** (0.003)	-0.024*** (0.003)	-0.023*** (0.003)
Clusters	20,339	20,117	18,307
Observations	64,339	99,875	101,848
R-squared	0.398	0.375	0.363

*** p<0.01, ** p<0.05, * p<0.1

Note.— 2+Sample includes first born from families with two or more children; 3+Sample includes first and second-born from families with three or more children and 4+Samples comprises first, second and third-born from families with four or more children. Control variables include child-level characteristics (sex, age, birth-year and month of birth), mother-level characteristics (mother's education, mother's age, mother-age squared, mother height, mother ethnicity, her religion, the number of her children who have died and an indicator variables taking one if any of the mother's children died before their first birthday), father-level characteristics (his age and education) and household-level characteristics (residence area and wealth quintile). Standard errors are clustered at the DHS-Cluster level and reported in parentheses.

Table 3: IV estimates of the effect of family size on children schooling

	2+SAMPLE	3+SAMPLE	4+SAMPLE
2ND STAGE			
DEPENDENT VARIABLE : EDUCATION Z-SCORE			
Number of children	0.019 (0.055)	0.097** (0.047)	-0.060 (0.064)
Clusters	20,339	20,117	18,307
Observations	64,339	99,875	101,848
R-squared	0.339	0.297	0.304
1ST STAGE			
DEPENDENT VARIABLE : NUMBER OF CHILDREN			
Twin at 2 nd birth	0.503*** (0.042)		
Twin at 3 rd birth		0.509*** (0.040)	
Twin at 4 th birth			0.356*** (0.038)
F-statistic (excluded instrument)	142.21	160.71	85.41
Under identification test p-value	0.000	0.000	0.000
Weak identification test:			
Cragg-Donald Wald F	143.41	259.91	136.89
(Stock-Yogo critical values)	(16.38)	(16.38)	(16.38)
Ho: equation is weakly identified			

*** p<0.01, ** p<0.05, * p<0.1

Note.— 2+Sample includes first born from families with two or more children; 3+Sample includes first and second-born from families with three or more children and 4+Samples comprises first, second and third-born from families with four or more children. Control variables include child-level characteristics (sex, age, birth-year and month of birth), mother-level characteristics (mother's education, mother's age, mother-age squared, mother height, mother ethnicity, her religion, the number of her children who have died and an indicator variables taking one if any of the mother's children died before their first birthday), father-level characteristics (his age and education) and household-level characteristics (residence area and wealth quintile). Standard errors are clustered at the DHS-cluster level and reported in parentheses.

Table 4: Determinants of twinning

DEPENDENT VARIABLE: PROBABILITY OF TWINNING		
	(I) With H+E	(II) With H+E+AC
Mother age	0.223*** (0.062)	0.208*** (0.062)
Mother age squared	-0.002** (0.001)	-0.002* (0.001)
Mother completed years of education	0.037** (0.015)	0.012 (0.015)
Mother height	0.005*** (0.001)	0.005*** (0.001)
Number of dead children	0.478*** (0.056)	0.497*** (0.056)
Infant death	0.453*** (0.127)	0.452*** (0.127)
Mother is Muslim	-0.136 (0.145)	-0.153 (0.145)
Mother BMI	-	0.000* (0.000)
Access to prenatal care	-	0.450* (0.234)
Access to a doctor	-	0.613* (0.326)
Access to a nurse	-	0.454** (0.199)
Clusters	121,597	121,597
Observations	250,837	250,837
R-squared	0.009	0.009

*** p<0.01, ** p<0.05, * p<0.1

Note.—This table shows linear probability estimates of twinning. The sample includes all children in Sample I. H+E stands for Mother's health characteristics and her ethnicity for which we control for in our baseline specification: her years of education, age, age squared, height, ethnicity, religion, number of her children who have died, whether children died before their first birthday. AC stands for four additional controls: mother's BMI, access to prenatal health care, access to a doctor and access to a nurse. We additionally control for mother ethnicity and the year in which the birth occurred. Robust standard errors are clustered at the mother level and reported in parentheses.

Table 5: Stability of the estimated coefficient of interest and post-estimation statistics when expanding the set of mother-level control variables

	DEPENDENT VARIABLE : EDUCATION Z-SCORE	
	(I) <i>With H+E</i>	(II) <i>With H+E +AC</i>
PANEL A: 2+SAMPLE		
Number of children	0.003 (0.063)	-0.005 (0.063)
<i>Clusters</i>	<i>16,935</i>	<i>16,935</i>
<i>Observations</i>	<i>42,789</i>	<i>42,789</i>
<i>R squared</i>	<i>0.345</i>	<i>0.337</i>
<i>Cragg-Donald Wald F</i>	<i>102.09</i>	<i>103.92</i>
<i>(Stock-Yogo critical values)</i>	<i>(16.38)</i>	<i>(16.38)</i>
PANEL B: 3+SAMPLE		
Number of children	0.137** (0.056)	0.126** (0.054)
<i>Clusters</i>	<i>16,940</i>	<i>16,940</i>
<i>Observations</i>	<i>66,654</i>	<i>66,654</i>
<i>R squared</i>	<i>0.273</i>	<i>0.287</i>
<i>Cragg-Donald Wald F</i>	<i>178.99</i>	<i>182.76</i>
<i>(Stock-Yogo critical values)</i>	<i>(16.38)</i>	<i>(16.38)</i>
PANEL C: 4+SAMPLE		
Number of children	-0.043 (0.073)	-0.061 (0.071)
<i>Clusters</i>	<i>14,996</i>	<i>14,996</i>
<i>Observations</i>	<i>68,144</i>	<i>68,144</i>
<i>R squared</i>	<i>0.304</i>	<i>0.309</i>
<i>Cragg-Donald Wald F</i>	<i>93.97</i>	<i>97.56</i>
<i>(Stock-Yogo critical values)</i>	<i>(16.38)</i>	<i>(16.38)</i>

*** p<0.01, ** p<0.05, * p<0.1

Note : H+E stands for Mother's health characteristics and her ethnicity for which we control in our baseline specification: her years of education, age, age squared, height, ethnicity, religion, number of her children who have died and whether children died before their first birthday. AC stands for five additional controls: mothers' BMI, access to prenatal health care, access to a doctor and access to a nurse. In Col. I we control H+E in Col. (II) and H+E and AC in Col. (III). Robust standard errors are clustered at the DHS-cluster level and reported in parentheses. 2+Sample, 3+Sample and 4+Sample are here restricted to subsamples for which the four additional control variables (AC) are available.

Table 6: Bounds' estimates of family size effect on children schooling using Conley's UCI and LTZ approaches

	Union of confidence interval (UCI)		Local to zero (LTZ)		
	Lower bound	Upper bound	Lower bound	Coefficient	Upper bound
2+SAMPLE	-0.108	0.127	-0.109	0.020	0.149
3+SAMPLE	-0.002	0.176	0.017	0.119**	0.221
4+SAMPLE	-0.196	0.048	-0.162	-0.033	0.095

Note : The bound estimates are derived using the 'Plausxog' command in Stata and are based on the prior that being from a twin family has a direct effect $\gamma = 0.005$ on educational outcomes (which for UCI bounds, is more conservative compared to the 0.004 used in Bhalotra and Clarke, 2016 for developing countries). The UCI bounds are derived based on $\gamma\text{-min} = 0.000$ and $\gamma\text{-max} = 0.010$ while the LTZ bounds are derived based on $\gamma = 0.005$ with a sd of 0.007 (the sd results from 100 replications bootstrap estimations and we perform test of normal distribution of γ ; see details in Table A3 in Appendix). Since the LTZ does not allow for factor variables, we exclude mother's ethnicity from the equation and all other variables enter as continuous variables. When applied to the baseline specification, this does not qualitatively change our results.

Table 7: Subsample analyses with respect to gender, asset wealth and region

	DEPENDENT VARIABLE: EDUCATION Z-SCORE							
	Only Boys	Only Girls	Only children from poor families	Only children from non-poor families	Only children living in West & Central Africa	Only children living in East & Southern Africa	High fertility	Low Fertility
PANEL A: 2+SAMPLE								
Number of children	0.042 (0.077)	-0.023 (0.079)	-0.094 (0.072)	0.020 (0.095)	0.050 (0.076)	-0.059 (0.066)	0.026 (0.103)	0.244 (0.225)
Clusters	15,963	15,825	11,080	11,287	13,549	6,790	7,901	2,924
Observations	32,831	31,508	26,814	25,458	44,276	20,063	30,767	6,960
R-squared	0.325	0.354	0.150	0.385	0.326	0.387	0.360	0.255
Cragg-Donald Wald F	72.25	69.98	57.60	60.97	71.19	109.35	39.38	15.07
PANEL B: 3+SAMPLE								
Number of children	0.122** (0.060)	0.064 (0.065)	0.060 (0.085)	0.139** (0.061)	0.108* (0.060)	0.075 (0.072)	0.128* (0.067)	0.058 (0.140)
Clusters	17,404	17,240	11,286	11,020	13,554	6,563	7,934	10,597
Observations	51,297	48,578	43,802	36,717	69,185	30,690	48,672	2,847
R-squared	0.282	0.315	0.150	0.333	0.286	0.347	0.306	0.326
Cragg-Donald Wald F	135.96	120.22	59.51	175.07	136.81	175.96	110.79	28.75
PANEL C: 4+SAMPLE								
Number of children	0.031 (0.072)	-0.183* (0.102)	-0.076 (0.108)	-0.132 (0.087)	-0.066 (0.084)	-0.046 (0.088)	-0.041 (0.101)	0.107 (0.283)
Clusters	15,994	15,900	10,314	9,372	12,518	5,789	7,546	2,039
Observations	52,637	49,211	47,269	34,662	71,406	30,442	50,298	3,886
R-squared	0.298	0.273	0.146	0.337	0.306	0.331	0.338	0.290
Cragg-Donald Wald F	95.68	120.22	33.06	83.93	67.14	124.04	45.44	28.75

*** p<0.01, ** p<0.05, * p<0.1

Note.— The control variables are as specified in the note of Table 3. Robust standard errors are clustered at the DHS-cluster level and reported in parentheses. Low and High fertility countries are as defined in Lesthaeghe (2014). Stock-Yogo critical value is 16.38 in all columns.

Table 8: Effect of family size across older and younger cohorts

	DEPENDENT VARIABLE: EDUCATION Z-SCORE		
	2+SAMPLE	3+SAMPLE	4+SAMPLE
Number of children	-0.067 (0.074)	0.079 (0.070)	-0.138 (0.091)
Number of children * young cohort	0.222 (0.170)	0.030 (0.120)	0.152 (0.149)
Female	-0.018*** (0.008)	-0.029*** (0.007)	-0.040*** (0.007)
Female * young cohort	0.042*** (0.014)	0.045*** (0.011)	0.028*** (0.011)
Mother's completed years of education	0.036*** (0.004)	0.040*** (0.003)	0.032*** (0.004)
Mother's completed years of education * young cohort	-0.003 (0.006)	-0.007* (0.004)	-0.003 (0.006)
Father's completed years of education	0.032*** (0.001)	0.034*** (0.001)	0.033*** (0.002)
Father's completed years of education* young cohort	-0.004* (0.002)	-0.005* (0.002)	-0.002 (0.003)
Rural	-0.164*** (0.021)	-0.206*** (0.017)	-0.175*** (0.021)
Rural * young cohort	0.081*** (0.025)	0.091*** (0.019)	0.076*** (0.020)
Wealth quintile	0.116*** (0.007)	0.122*** (0.005)	0.124*** (0.005)
Wealth quintile * young cohort	0.048** (0.020)	0.023** (0.012)	0.026*** (0.010)
Clusters	20,339	20,117	18,307
Observations	64,339	99,875	101,848
R-squared	0.328	0.301	0.292
Cragg-Donald Wald F (Stock-Yogo critical values)	31.40 (7.03)	58.73 (7.03)	33.02 (7.03)

*** p<0.01, ** p<0.05, * p<0.1

Note.— This table shows the effect of the number of children and other relevant variables on children's schooling for children belonging to a young (born after 2000) and an older cohort (born before 2000). Control variables include child-level characteristics (age, birth-year and month of birth), mother-level characteristics (mother's age, mother-age squared, mother height, mother ethnicity, her religion, the number of her children who have died and an indicator variables taking one if any of the mother's children died before their first birthday) and father's age. Robust standard errors are clustered at the DHS-cluster level and reported in parentheses.

Table 9: Summary of robustness checks

Robustness check	Table	Description	Coefficient on number of children		
			2+SAMPLE	3+SAMPLE	4+SAMPLE
1	Table A4	Heterogeneity of twin at 3 rd birth			
		<i>Interaction with Rural</i>	0.019 (0.056)	0.098** (0.047)	-0.061 (0.064)
		<i>Interaction with Mother is Muslim</i>	0.008 (0.054)	0.092** (0.046)	-0.047 (0.063)
		<i>Interaction with Rural West & Central Africa</i>	0.018 (0.054)	0.094** (0.047)	-0.064 (0.062)
2	Table A5	Partially missing instruments with non-linear specification			
		<i>Number of children > 2</i>	-0.050 (0.264)		
		<i>Number of children > 3</i>	0.156 (0.122)	0.275** (0.127)	
		<i>Number of children > 4</i>	-0.083 (0.140)	-0.104 (0.115)	-0.110 (0.120)
		<i>Number of children > 5</i>	0.021 (0.166)	0.023 (0.129)	0.043 (0.116)
3	Table A6	Expand sample incl. households in which some school-aged children reside outside the household	0.029 (0.056)	0.091* (0.048)	-0.114* (0.070)
4	Table A7	Decentralized approach (mother as decision unit)	0.001 (0.066)	0.059 (0.042)	0.003 (0.044)
5	Table A8	Alternative definition of the number of children (as the total number of births given by the household head's wives)	0.018 (0.051)	0.092* (0.045)	-0.055 (0.059)
6	Table A9	Education Z-score calculated with censored completed years of education (to the child's age minus 6 at most)	0.050 (0.045)	0.113** (0.041)	-0.054 (0.057)
7	Table A10	Completed grade (years of schooling) as outcome variable	0.075 (0.095)	0.144* (0.079)	-0.031 (0.111)
8	Table A11	Region of residence of the household instead of mother's ethnicity FE	0.030 (0.048)	0.063 (0.041)	-0.080 (0.054)
9	Table A12	Country-by-urban/rural instead of mother's ethnicity FE	0.040 (0.048)	0.097** (0.044)	-0.054 (0.061)
10	Table A13	Alternative definition to discriminate between poor and non-poor families			
		<i>Poor</i>	-0.030 (0.071)	0.034 (0.073)	-0.072 (0.088)
		<i>Non-Poor</i>	0.042 (0.087)	0.138** (0.059)	-0.099 (0.093)

Note.— Detailed tables of robustness checks are provided in the Supplementary Appendix A4-A13.

Table 10: Exploring the proximate cause underlying the positive effect of family size on schooling

	(I) Enrolment	(II) Starting age	(III) Starting age when difference in age between 1 st and 2 nd born ≤3years	(IV) Drop out
PANEL A: 2+SAMPLE				
Effect on 1 st born (All)	-0.009 (0.023)	-0.172 (0.149)	-0.214 (0.245)	-0.024 (0.026)
Effect on 1 st born (Poor)	0.000 (0.039)	0.134 (0.169)	0.192 (0.182)	-0.058 (0.057)
Effect on 1 st born (Non-poor)	-0.028 (0.023)	-0.499* (0.260)	-0.965* (0.528)	0.008 (0.039)
PANEL B : 3+SAMPLE				
Effect on 1 st born (All)	0.001 (0.022)	-0.196 (0.139)	-0.048 (0.181)	-0.014 (0.030)
Effect on 1 st born (Poor)	-0.019 (0.047)	-0.039 (0.404)	0.550 (0.534)	0.065 (0.092)
Effect on 1 st born (Non-poor)	0.030 (0.023)	-0.128 (0.154)	-0.179 (0.210)	-0.017 (0.030)
Effect on 2 nd born (All)	0.037 (0.024)	-0.018 (0.180)	-0.222 (0.208)	0.063* (0.035)
Effect on 2 nd born (Poor)	-0.010 (0.050)	0.524 (0.405)	0.691 (0.490)	0.062 (0.081)
Effect on 2 nd born (Non-poor)	0.023 (0.027)	-0.285 (0.180)	-0.623** (0.275)	0.060 (0.040)
PANEL C : 4+SAMPLE				
Effect on 2 nd born (All)	-0.010 (0.050)	0.524 (0.405)	0.691 (0.490)	0.062 (0.081)
Effect on 2 nd born (Poor)	0.034 (0.063)	-0.462 (0.291)	-0.248 (0.326)	-0.008 (0.085)
Effect on 2 nd born (Non-poor)	-0.072* (0.042)	0.342 (0.441)	0.469 (0.435)	-0.048 (0.047)
Effect on 3 rd born (All)	-0.004 (0.035)	-0.204 (0.234)	-0.109 (0.177)	-0.047 (0.040)
Effect on 3 rd born (Poor)	-0.027 (0.068)	0.240 (0.276)	-0.147 (0.213)	0.016 (0.087)
Effect on 3 rd born (Non-poor)	0.020 (0.046)	-0.557 (0.429)	-0.159 (0.380)	-0.076* (0.040)

*** p<0.01, ** p<0.05, * p<0.1

Note.— The control variables are as specified in the note of Table 3. Robust standard errors are clustered at the DHS-cluster level and reported in parentheses. In Cols (II) to (V) we consider only children enrolled in school at the time of the survey. In Cols (IV) and (V), the school starting age is generated based on censored years of education (see robustness check 5). Following Angrist & Pischke (2009: 102-107), we use manual 2SLS as a linear probability model in Cols (I) and (VI), although 'enrolment' and 'drop out' are binary variables. In Panel C, the age difference in Cols (III) and (V) is between the 2nd and the 3rd born. Full results are reported in the Supplementary Appendix Tables A14-A20.

Country	DHS round	Exploitable data on child labor
Benin	1996	No
	2001	No
	2006	No
	2012	No
Burkina	1993	No
	2003	No
	2010	Yes
Cameroon	2004	No
	2011	Yes
Central African Rep.	1995	No
Chad	2004	Yes
Congo	2012	Yes
Congo Dem.	2007	Yes
	2014	Yes
Cote d'Ivoire	2012	Yes
Ethiopia	2000	No
	2005	No
	2011	Yes
Gabon	2000	No
	2012	Yes
Gambia	2013	No
Ghana	1993	No
	1998	No
	2003	No
	2008	No
	2014	No
Guinea	1999	No
	2005	No
	2012	Yes
Kenya	1993	No
	1998	No
	2003	No
	2014	No
Liberia	2013	No
Malawi	2000	Yes
	2004	Yes
	2010	No
Mali	1996	No
	2006	Yes
	2013	Yes
Mozambique	1997	No

Namibia		2011	No
		1992	No
		2000	No
Niger		1998	No
		2006	Yes
Nigeria		2008	No
		2013	No
		1993	No
Senegal		2005	Yes
		2011	Yes
		2013	Yes
		2014	No
		2008	Yes
Sierra Leone		2013	Yes
		1998	No
Togo		2014	Yes
		2011	No
Uganda		2011	No
Zimbabwe		1994	No
All Together		59	

Source.— Authors, based on DHS data