The Twin Instrument*

Sonia Bhalotra† Damian Clarke‡

June 6, 2014

Abstract

The incidence of twins has been used to identify the impact of changes in fertility on measures of investment in children born prior to the twins, and the emerging consensus in this literature is that there is no evidence of a quantity-quality trade-off. We argue that the standard approach is flawed for two reasons. First, even if twin conception is random, bringing twins to term is a function of maternal health which is difficult to fully observe and which tends to be correlated with child quality, rendering the instrument invalid. Second, twins will only constitute a shock to family size if their occurrence takes family size across the desired level. The neglect of both of these considerations in the literature will tend to lead to under-estimation of the quality-quantity (Q-Q) trade-off and so could contribute to explaining the negative results in the literature. Using a large sample of microdata from developing countries, we show that a significant trade-off emerges upon correcting for these biases. We show that this result is likely to be only a *lower* bound of the true Q-Q trade-off and discuss how to estimate the size of these bounds.

JEL codes: J13,J16,J18,I15,O15.

*We are grateful to Paul Devereux, James Fenske, Atheen Venkataramani and Cheti Nicoletti, [WHO ELSE TO ADD?] along with seminar audiences and discussants at CMPO Bristol, ESPE, NEUDC, CSAE, and The University of Essex for helpful comments. We make full data and source code available for examination and replication. This is available at `https://bitbucket.org/damianclarke/twins`.

†The University of Essex. Contact: srbhal@essex.ac.uk

‡The University of Oxford. Contact: damian.clarke@economics.ox.ac.uk

1 Introduction

Include introduction here.

2 The Twin Literature

The occurrence of twin births is a commonly used for identification in economic studies. Specifically, twin births as a shock to total fertility have been used to motivate the estimation of the quantity-quality trade-off (Rosenzweig and Wolpin, 1980a; Black et al., 2005; Angrist et al., 2010; Cáceres-Delpiano, 2006), the effects of childbearing on female labour force participation (Rosenzweig and Wolpin, 1980b; Jacobsen et al., 1999; Angrist and Evans, 1998), and the effects of unwed childbearing on marriage market outcomes, poverty and welfare receipt (Bronars and Grogger, 1994).[1]

The methodology employed in all of these studies requires that twinning be (at least conditionally) exogenous. Typically then, the previous literature has made one of a series of assumptions. The earliest twin studies of the Q-Q trade-off (Rosenzweig and Wolpin, 1980a,b) pointed out that twinning is both a positive function of total parity, and of maternal age. By focusing on twins at first birth and conditioning on maternal age, they thus produced consistent estimates of the tradeoff under the assumption that beyond maternal age and parity, twin births were entirely random. A more recent wave of studies including Black et al. (2005); Cáceres-Delpiano (2006); Li et al. (2008); Angrist et al. (2010) take this one step further, assuming that twinning is conditionally exogenous after controlling for additional family characteristics such as a mother's race and her educational attainment. They thus condition on a mother's age, race, and education, as well as on the birth order of the child in first and second stage regressions.

In some cases the validity of such assumptions is probed by regressing twinning on observable family outcomes, or testing for the equality of means of certain characteristics between twin and non-twin families. Black et al. (2005), Li et al. (2008) and Sanhueza (2009) report joint F-tests suggesting that twinning is not related to parental education in their data samples, while Rosenzweig and Zhang (2009) report t-tests showing equality of means across twin and non-twin groups. However, as is well known and acknowledged in each case, any such tests are at best partial evidence in support of instrumental validity. While twins can be shown to be unrelated to

[1] Twins have been widely used in the economic, medical, biology and psychology literature in a number of ways. In this paper we focus only on the use of twin births as an instrument for total fertility, and not on the so called 'twin studies', which base inference on between-twin comparisons using maternal fixed effects.

observable or measured characteristics, similar tests cannot be run for variables which are either unobservable, or not recorded in survey data. We return to this point in the following sections.

Finally, recent studies aim to control for the fact that multiple births are correlated with fertility treatments. Typically such a treatment requires either focusing on offspring born before the introduction of fertility treatments (Cáceres-Delpiano, 2006; Angrist et al., 2010), or, in the case of sufficiently rich data, removing families undergoing fertility treatment from estimation samples (Braakmann and Wildman, 2014). Once again, consistent estimation in this case is based on the assumption that beyond fertility treatment and family controls listed above, twin births are as good as random.

Critiques of the twin instrument have focused on parental behaviours in response to twins, rather than on the likelihood that parental behaviours (or endowments) may affect the likelihood of twinning. Rosenzweig and Zhang (2009) question the effect that close (or indeed no) birth spacing and an endowment effect—where parental behaviours respond to the lower health at birth of twins compared to single births[2]—has on investments in pre-twin siblings. They demonstrate that if parents behave in such a manner, bounds for the Q-Q trade-off can be calculated. This hypothesis is tested in Angrist et al. (2010), and applied in Fitzsimons and Malde (2010).

3 Data and Estimation Samples

3.1 Data

The data used here come from the Demographic and Health Surveys (DHS). The DHS are a set of nationally representative surveys which have been administered in low- and middle-income countries between 1985 and the present. Women aged between 15–49 in surveyed households respond to an in-depth series of questions reporting their full fertility history (listing all surviving and non-surviving children), their actual and desired contraceptive use and number of births, education level, marital status, plus the measurement of a number of health endowments such as height and body mass index. For all other members living in the household a shorter series of responses are recorded, including the individual's educational attainment.

This results in two distinct sets of data to be merged. One database contains one line for

[2]Using data from the United States, Almond et al. (2005) document that twins have substantially lower birth weight, lower APGAR scores, higher use of assisted ventilation at birth and lower gestion period than singletons.

each birth reported by every 15–49 year-old woman surveyed with a limited number of child-level covariates such as the child's date of birth, type of birth (single or multiple), and the child's survival status. The other database contains one line for each member currently living in the survey household. This database includes each member's educational status. By crossing these two databases we are thus able to generate data for the educational attainment of each of a woman's children currently residing in the household as well as their mother's health and educational status. This database is thus selected in two ways: firstly it only contains children who have survived up until the survey date, and secondly it only contains children who have remained living in the same household as their mother. In order to ensure that this sample is as representative as possible, we restrict our analysis to those children aged 18 and under.

We pool all publicly available DHS data resulting in microdata on 6,697,397 ever-born children. A full list of the DHS countries and years of surveys which make up this sample is provided as appendix table 10. Of the 6,697,397 offspring reported in survey data, 3,803,796 remain living in the same household as their mother. The majority of these 3,803,796 children are aged 18 and under (91.45%) and hence make up our principal estimation sample (in future we will refer to this as the 'household sample'). The remaining 2,893,601 offspring were not recorded as living in the same household as their mother. Of these children *not* in the household, and hence for whom education is not recorded, the majority (51.8%) were aged over 18 or had died prior to the date of survey.[3]

3.2 Estimation Samples

We work with a number of estimation samples. For tests in which twinning is the outcome variable of interest, our sample consists of all children who remain living in the household with their mother, and hence for whom we have full covariates. For tests in which child death is the outcome, we use the full sample of under 18 year-olds, whether or not they remain in the household of their mother. Such a sample is necessary given that children who died prior to the date of the survey no longer appear in our household sample.

In all tests of the Q-Q trade-off using twins, we follow existing literature in defining birth-order-specific estimation samples conditional on the total number of a mother's births. These samples are referred to as the 2+, 3+, 4+ and 5+ samples. These samples are defined $\forall\ n \in \{2,3,4,5\}$ such that they include first-born to $n-1$ born children in families with at least n

[3]Children aged under 18 who are alive but not living in the same household as their mother are statistically quite different to those children who do remain in the household. In our data sample, they are on average 2 years older, born to less educated and older mothers, and are slightly less likely to be males or twins.

births.[4] As an example, the 2+ sample consists of first-borns in families with at least two births, and the 3+ sample consists of first- and second-borns in families with at least 3 births. Such a sample decision is important when estimating the Q-Q trade-off using twinning as an instrument. Given that family size is endogenously chosen by parents and rates of twin birth are not constant by birth-order, twin-births will occur more frequently in families that have a higher fertility preference. This point is addressed by (among others) Rosenzweig and Wolpin (1980a) and Black et al. (2005) who first suggested combining $n+$ groups with twinning at birth order n as a way to ensure that twin and non-twin families in the sample would have similar fertility preferences.

Table 1 presents the sample size along with basic descriptives for each of the samples. Full descriptives are provided in the sub-section which follows. For each of the 2+ to 5+ samples we only focus on children aged over six given that our outcome of interest requires the child to have (at least) reached primary-school age.

3.3 Descriptive Statistics

Table 2 provides summary statistics of this data. Fertility and maternal characteristics are described at the level of the mother, while child education and survival are described at the level of the child (the education and full samples respectively). The number of observations at each level is provided at the bottom of the table.

Survey countries are classified according to country income level in order to allow for a disaggregation of Q-Q results by income group.[5] We present summary statistics by birth type (singleton or twin), and by country income status. Twin births make up 1.85% of all births. A simple comparison of means suggests that healthy mothers (as proxied by height, BMI and probability of being underweight) are more likely to give birth to twins, and that twin births are more likely to occur in low-income countries. This apparent contradiction can be explained given that twins are (both mechanically and biologically) a positive function of fertility, and fertility is higher in the low-income sample. Figure 2 describes this positive relationship: while twins account for less than 1% of all first-borns, they account for greater than 4% of all tenth-born children. As expected, twin families are larger than non-twin families. Figure 3 describes total

[4] Existing studies such as Angrist et al. (2010) focus mainly on the 2+ and 3+ samples. Given the higher fertility in the DHS data, we also include higher birth-order groups.

[5] This classification is obtained from the World Bank, with DHS surveyed countries falling into two broad groups based on their GNI per capita at the moment of the DHS survey. These groups consist of countries classed as low-income economies, and countries classed as middle-income economies (either lower-middle or upper-middle). Details regarding this classification can be found in Appendix Table 10.

fertility in twin and non-twin families. The distribution of family size in families where at least one twin birth has occurred dominates the corresponding distribution for all-singleton families. This is expected given imperfect fertility control and—even were fertility perfectly controlled by families—given that some twins will occur on a family's final desired birth. Such a result is required for instrumental relevance when using twining to estimate a Q-Q trade-off.

Child 'quality' is measured using each child's educational attainment. Our principal outcome variable is a standarised score for schooling (Z-Score). This Z-Score is calculated by comparing each child's total years of completed education to his or her birth cohort in their country of residence. This allows us to express all effect-sizes in terms of a one standard deviation increase in total educational attainment.

4 Methodology

Typically, empirical analyses of the quality-quantity trade-off focus on producing consistent estimates of β_1 in the following specification:

$$educ_{ij} = \beta_0 + \beta_1 fert_j + \boldsymbol{X}\boldsymbol{\beta}_2 + u_{ij}. \tag{1}$$

Here, quality is proxied by the educational attainment of child i in family j, ($educ$) and fertility ($fert$) is measured as the total births in a child's family. A vector of family and child controls is included, denoted $\boldsymbol{X}$. As has been extensively discussed in prior literature, estimation of β_1 using OLS and cross-sectional data will result in biased coefficients given that child quality and quantity are jointly determined (Becker and Lewis, 1973; Becker and Tomes, 1976), and given that unobservable parental behaviours and attributes influence both fertility decisions, and investments in children's education (Qian, 2009).

4.1 Quantity-Quality with Twins

As a result, 2SLS estimation is employed, where fertility is instrumented using twin births.[6] The corresponding first stage is:

$$fert_j = \alpha_0 + \alpha_1 twins_j + \boldsymbol{X}\boldsymbol{\alpha}_2 + \nu_j, \tag{2}$$

[6] Other instruments and methodologies are also used including gender mix of children (Angrist et al., 2010), policy experiments (Qian, 2009), and historical time series variation in schooling (Bleakley and Lange, 2009)

where $twins_j$ is an indicator for whether the n^{th} birth in a family is a twin birth. As described in section 3.2, the sample in each case is the so-called $n+$ group, consisting of children born before birth n in families with at least n births. As such the twins themselves are excluded from the estimation sample.[7] As an alternative specification we do include children from the n^{th} birth (twins in twin families), however generally focus on the typical $n+$ sample as our principal sample.

Consistent estimation of $\hat{\beta}_1$ can thus proceed provided (among other things) that instrumental validity holds:

$$P \lim N^{-1} twin_j u_{ij} = 0 \tag{3}$$

Given the unobservable nature of the error term u, we are unable to test (3). There is however nothing which stops us from partially testing (3) by removing a subset of observable components from u. The error term u in (1) is a function of a large number of elements:

$$\begin{aligned} u = f(&\text{maternal health, fertility behaviour, positive pregnancy investments,} \\ &\text{parental education, negative pregnancy investments,} \ldots) \end{aligned} \tag{4}$$

for now we do not specify the signs on any of these variables. While many of the relevant elements are either completely or partially unobservable to the econometrician, some of these variables, such as maternal education and health can be measured. Thus a partial test of the twin methodology consists of estimating the following regression.

$$twin_j = \gamma_0 + \boldsymbol{X}\boldsymbol{\gamma}_1 + \boldsymbol{S}\boldsymbol{\gamma}_2 + \boldsymbol{H}\boldsymbol{\gamma}_3 + \varepsilon_j. \tag{5}$$

Here $\boldsymbol{X}$ refers to the initial vector of family and child controls, $\boldsymbol{S}$ to additional family socioeconomic variables such as income and parental education, and $\boldsymbol{H}$ to maternal health variables.

If twin birth is indeed an event which is as good as random, the coefficients on maternal health and family socioeconomic variables in the above regression should not be significantly different to zero. We thus test the following hypothesis in our DHS data:

$$H_0 : \boldsymbol{\gamma}_2 = \boldsymbol{\gamma}_3 = 0. \tag{6}$$

Rejection of the null would raise difficulties in proceeding with IV estimation using the twin instrument. Of course, if the rejection of the null were only due to one or a number of *observable* element(s) which predicted twinning, these variables could simply be included in the first and second stages above, much like occurs with maternal age and race in the existing literature.

[7] Typically, the argument is made that twins are different to single births, and hence should not be compared in analysis

However, more generally it would be difficult to be argue for instrumental validity if twinning is shown to depend upon (a limited set of) measurable family characteristic or choice variables, while many similar variables are not observed.

Given the biological demands placed on a mother pregnant with twins, we may expect that healthier mothers, or mothers with more resources to invest in their pregnancy are more likely to take twin conceptions to term. Similarly, we may suspect that mothers more able to invest their pregnancy will also be more able to invest in their child's human capital after birth. If this is the case, we would see that (at the very least) $\boldsymbol{\gamma_2} > 0$.

If we assume additive separability of the elements in the omitted error term, we can re-write u_{ij} from (2) and (4) as:

$$u_{ij} = u_{ij}^S + u_{ij}^H + u_{ij}^*.$$

Here u_{ij}^S and u_{ij}^H correspond to the identical (observable) elements included as $\boldsymbol{S}$ and $\boldsymbol{H}$ in (5), while u_{ij}^* represent the remaining (unobserved) components. We can thus re-write our IV estimate for β_1 as:

$$\hat{\beta}_1^{IV} = \beta_1 + P\lim N^{-1} twin_j u_{ij}^S + P\lim N^{-1} twin_j u_{ij}^H + P\lim N^{-1} twin_j u_{ij}^* \tag{7}$$

Typically, this is the coefficient estimated in the existing twin literature which assumes that twinning is a conditionally exogenous event. If, however, the likelihood of taking twin conceptions to term increases for healthier and/or wealthier mothers, we should include $\boldsymbol{S}$ and $\boldsymbol{H}$ in the first and second stages, giving

$$\hat{\beta}_1^{IV,S+H} = \beta_1 + P\lim N^{-1} twin_j u_{ij}^*, \tag{8}$$

where the superscript $S+H$ signifies that socioeconomic and health variables have been included as additional controls. What's more, if both the likelihood that a family takes twins to term and a family's investment in child human capital are positively correlated with positive health behaviours and other positive socioeconomic variables such as parental education, we would expect that:

$$\hat{\beta}_1^{IV} > \hat{\beta}_1^{IV,S} > \hat{\beta}_1^{IV,S+H} > \beta_1.$$

It should be noted in the above series of inequalities that even conditional upon socioeconomic and health variables, IV estimation will *not* result in a consistent estimate of β_1 if twinning is correlated with unobservable elements in u_{ij}^*. We return to this point, and how to bound β_1 in the following sub-section.

4.2 Bounding the Q-Q Trade-off

In the previous subsection, Q-Q estimation using twins is motivated in equations (1) and (2). Consistent IV estimation imposes the (strong) prior belief that twin births can be excluded from the second stage equation, or that the sign of γ in the following is equal to zero:

$$Quality_{ij} = \beta fert_j + \gamma twin_j + u_{ij}. \tag{9}$$

As we discuss above, this will not be the case if maternal health controls omitted from (1) are correlated with both the likelihood of taking twin conceptions to term, and with eventual measures of child quality.

However, even in cases such as this where we are not confident that $\gamma = 0$, we can still estimate bounds on the Q-Q tradeoff if we are confident in making some statement of prior belief about the distribution from which γ is drawn. Conley et al. (2012) describe such a process, which they refer to as *plausible exogeneity*. We invoke this terminology here, and refer to twins as a plausibly exogenous event, implying that we have reason to believe that γ may be close to, but not necessarily precisely equal to, zero.

In this paper we estimate β under a range of assumptions regarding the true nature of γ. Firstly we estimate β by simply assuming a support assumption for γ: specifically that γ falls between zero (implying instrumental validuty) and some (positive) number δ:

$$\gamma \in [0, \delta]. \tag{10}$$

This is a relatively weak assumption, however, as Conley et al. (2012) show, it allows for us to recover a 'union of confidence intervals' (hereafter UCI) for estimates of β over the entire support of γ. This UCI, then, provides bounds for β even in the case that twin exogeneity does not strictly hold. We also estimate by imposing a stronger prior: specifically we fully specify the distribution of γ as:

$$\gamma \sim U(0, \delta). \tag{11}$$

This stronger assumption allows for a tighter estimate of the bounds on β. Conley et al. (2012) provide a full derivation of this result, and we follow them in referring to this as a local-to-zero (LTZ) approximation.

Assumptions (10) and (11) depend upon the values of δ which we believe hold in the case of twinning and the Q-Q equation. As such, we estimate both specifications over a wide range of values for δ, however our preferred values (those which we present in tables in this paper) are that $\gamma \in [0, 2\hat{\gamma}]$ and $\gamma \sim U(0, 2\hat{\gamma})$. We thus first estimate $\hat{\gamma}$ by running the structural equation (9), and then plug these estimates back into the Conley et al. plausibly exogenous approach to estimate bounds for β.

5 Results

5.1 Twin Exogeneity

Table 3 presents a test of balance of maternal characteristics between mothers who have and who have not given birth to twins. This table presents a comparison of means, and a two-tailed t-test of the null that mothers of twins are not statistically different from mothers of non-twins. These comparisons suggest that twin mothers are older at the moment of birth, begin child-bearing earlier, have more children, and are more (less) likely to come from top (bottom) wealth quintiles. These comparisons of means, however, are not surprising given the well established fact that twins occur with higher frequency as mothers age, and at higher parities (Hall, 2003). Given that mother's who have more births (and hence who are more likely to have twins) are different in other dimensions, these tests may capture the effect of selection into higher fertility, rather than twinning itself. It is, however, worth pointing out that table 3 suggests that healthier mothers as proxied by height and body mass index (BMI) are unconditionally more likely to have twins. Similar comparisons of all births suggest that healthier mothers on average have less births.

In order to examine partial correlations, we regress a child's twin status (1 if a twin, 0 otherwise) on their mother's health, education, and a range of other demographic and family characteristics.[8] In table 4 we report the results from specification (5). These results suggest that twin births are not random, even after conditioning on maternal age and child birth order as is typical in recent twin literature. The inclusion of a full set of country and year-of-birth dummies (not displayed in table 4) will capture any systematic trend in the frequency of twin births across time or regions, and country dummies will absorb all time invariant differences in the probability of a twin birth across countries. The estimated coefficients and signs support the idea discussed in section 4 that higher 'investments' (for example in maternal health) required to maintain multiple healthy fetuses *in utero* may result in non-random twin births. Initially results from the pooled DHS data are presented as this provides a particularly large sample with

[8] In our principal specification, the full set of controls are country, child year of birth, and age dummies; a cubic function of mother's age at time of birth; mother's age at time of first birth; mother's education and education squared; and mother's height and BMI. We cluster standard errors at the level of the mother.

which to test the hypothesis of twin exogeneity. This is represented in table 4 column (1) and provides considerable evidence that live multiple births respond to family 'choice' variables such as education (tests for the joint significance of both socioeconomic variables and health variables are rejected with p-values of 0.0000).

These results are robust to all alternative specifications examined. Significant and quantitatively similar results are found if a logit model is estimated rather than a linear probability model, and when running separate models for twinning at each birth order. Similarly, if we run the regression at the level of the mother or include any combination of fertility measures, similar patterns are observed.

The fact that maternal health is correlated with twinning is supported by medical literature, although is not a point which has been incorporated into prior economic studies of twinning. Hall (2003) for example suggests that follicle-stimulating hormone (FSH) is associated with an increased likelihood of twinning, and is found in higher concentrations in older, heavier and taller mothers. Further, she suggests "that adequate maternal folic acid consumption could affect the number of twins coming to term" (see p. 741, and further discussion in Li et al. (2003)). Given that twinning also increases in cases where the mother undergoes fertility treatment, we run a similar regression for children born in a period not potentially affected by IVF.[9] These results are included in columns (4) and (5). Pre- and post-1990 results are qualitatively similar although education is no longer significant prior to 1990 (in the smaller sample). Mother's height and weight, and family socioeconomic variables remain economically and statistically significant.

If the reason non-random twin births are observed is due to insufficient investment in the developing fetus, it seems likely that twin 'selection' will be more pronounced in lower income settings, and settings where the mother is less well resourced during gestation. This is tested in columns (2) and (3), where it is shown that the violation of the twin exclusion restriction is particularly strong in low income countries. Here maternal health is a more important predictor, and the explained portion of this set of variables is larger that in middle-income countries.[10] Finally, we add measures for the availability of prenatal care (for the subset of children for whom this variable is recorded). The omitted case are those who report prenatal care from relatives or traditional health care providers. We find evidence to suggest that those receiving prenatal care from a doctor are more likely to give live birth to twins, and those who receive no prenatal care are less likely to twin. We treat these results with precaution however, given that accessing

[9]In order to be conservative, we estimate for the period preceeding 1990, the date which coincides with the first reported successful use of IVF in South Africa, an early-adopter among DHS countries.

[10]The low R^2 in these regressions is not at all surprising given that twin conception can be thought of as an approximately random process. The fact that socioeconomic and health variables have *any* power in explaining twin birth however is sufficient to invalidate IV estimations if these or other relevant predictors are not controlled for.

prenatal care is a choice variable, and likely endogenous.

These results call into question the veracity of the conditional exogeneity (or 'as good as random') assumption regarding twin births required to estimate β_1 consistently in (1). This implies that omitting factors such as family socioeconomic and maternal health variables is likely to result in inconsistent IV estimates if these variables are also correlated with child 'quality' outcomes. It seems unlikely, furthermore, that controlling for the factors that we have shown to predict twinning will allow us to recover consistent IV estimates. Maternal health stocks are both difficult to measure, and multi-dimensional in nature.[11] In the following subsections we examine the effect that omitting these factors is likely to have on IV estimates of the Q-Q trade-off.

5.2 Twinning and Quantity-Quality

As is typically found in empirical studies of the Q-Q tradeoff, correlations between family size and child outcome variables are negative, and strongly significant. Table 6 shows OLS estimates of child 'quality' on total fertility. These results suggest that an additional sibling is associated with an approximatelty 0.1 sd decrease in standardised schooling outcomes.

Of course, this empirically observed relationship between the quantity of an individual's siblings and their measured 'quality' does not necessarily imply that such a trade-off exists if parental decisions regarding the production of child quality and quantity are jointly made and possibly influenced by unobserved factors (Becker and Tomes, 1976). Principally here we are concerned with unobserved parental behaviours which may favour both lower family size and higher child quality. The OLS results are consistent with such a result, as the inclusion of maternal education and maternal health controls—likely correlated with desires for smaller family size and higher investments per child—reduce the magnitude of this observed trade-off.

5.2.1 Q-Q Tradeoff: Estimates Using Twin Births

Rather than focusing on OLS estimates which are likely to be biased, we turn to estimates which rely on twin births to identify the Q-Q tradeoff. As we outline in section 5.1, the assumption of 'as good as random' twin births is unlikely to hold, even when conditioning on the augmented set of controls proposed in (5). If this is the case, we will also be unable to consistently estimate

[11] In ongoing work, we find that the probability of twinning falls with maternal alcohol consumption, smoking, drug taking and a range of other behaviours in data from the USA, Chile, UK and Scotland.

β_1 using twin births.

However, it is likely that the β_1 that we estimate using twin births will provide us with a strict lower bound of the magnitude of the Q-Q trade-off as outlined in (4.1). We expect that the bias in this estimate is due to those mothers who invest more in their children in utero, or who have greater initial health endowments, being more likely to give birth to twins resulting in larger family sizes. At the same time, we expect healthier mothers to invest more in their children after birth resulting in higher quality children. By relegating health variables to the error term, these two positive correlations will result in a positive bias on the fertility coefficient estimated via IV. In order to determine the effect that these omitted variables have on estimates of the Q-Q trade-off, we turn to results for equation (1), both first omitting, and the including, maternal health and socioeconomic variables.

We present IV estimates of the Q-Q trade-off in table 7. The main specification is displayed in the top row with separate columns for the 2+, 3+ and 4+ sample groups. In each case the base case (controlling for maternal and child age, country, and year of birth) results in insignificant, and at times weakly positive, estimates of the effect of an additional birth on a child's educational attainment. These results suggest that the inclusion of maternal health and socioeconomic controls may be of considerable importance. Despite the lack of results when using the 'typical' set of twin controls, including health (columns 2, 4 and 6) reduces point estimates on fertility from an effect of approximately 0% of a standard deviation, to -3 or -4% of a standard deviation in standardised educational attainment. Further, conditioning on maternal education results in slightly more precise estimates, suggesting a statistically significant[12] Q-Q trade-off of at least 3 or 4%.

5.2.2 Heterogeneity of the Q-Q Trade-off

Estimates of the Q-Q trade-off are heterogeneous across birth orders, country income level, and gender of the child affected by the additional birth. Such a result is not surprising if parents perceive that returns to education vary for different children or in different circumstances, and invest in human capital in line with this.

The magnitude and significance of the results is lowest when considering the effect on the first-born child of moving from two to three births (the 2+ group), and higher when considering moving from three to four births or four to five births. However, in lower fertility environments

[12] Or close to statistically significant in the case of the 2+ sample.

the effect is, as expected, concentrated on lower birth orders. The third row of table 7 suggests that in middle-income countries that the effect is largest on first borns, and progressively smaller, but still considerable, at higher birth orders.

Estimates of the magnitude of the Q-Q trade-off by country income level suggest that the trade-off is considerably larger in *middle* rather than low-income countries. In low-income countries point estimates on fertility suggest (insignificant) trade-offs centred around 2-3% of a standard deviation, while in middle-income countries results are significant, and considerably larger, reaching as much as 9% of a standard deviation: only slightly lower than OLS estimates for this group. [ADD TEST HERE LOOKING AT COSTS OF EDUCATION? HIGHER IN COUNTRIES WHERE EDUCATION COSTS MORE? SONIA, DO YOU HAVE THE EDUCATION COST DATABASE THAT I THINK YOU MENTIONED TO ME ONCE?]

Similarly, effects of the Q-Q trade-off are considerably different depending upon a child's gender. In table 11 we present specifications estimated separately by the gender of the index child. These results suggest that females may bear the brunt of additional births, with estimates being negative and significant for girls, while insignificant for boy children. Such a result suggests that parents may engage in redistributory behaviour. This idea has been extensively studied by Rosenzweig and Zhang (2009), however their analysis focuses on the possibility that parents reinforce positive birth endowments based on child health, not based on child gender as we find here.

In order for the results we produce to be comparable to recent literature we thus far have only focused on children *preceding* twins. As an alternative specification we also include the n^{th} birth in the analysis sample.[13] Given the lower birth endowments of twins (and potential that reinforcing behaviour moves resources away from twins as per Rosenzweig and Zhang (2009)), we expect to see that the trade-off is significantly stronger when including twins in the sample. These results are displayed in the final row table 7, and suggest that, if we focus not just on pre-twins, that the trade-off of an additional birth is large: on the order of 5-8% of a standard deviation. Finally, in row 4 of the table we perform an alternative consistency check. Rather than using the full fertility reported by a family, we adjust birth order and fertility to only account for those children who survived beyond 1 year of age. Once again, we find that the estimated Q-Q trade-off is larger, and of similar magnitude to the results when including twin births. However, given that child survival could itself be thought of as a 'quality' variable, it is not clear that this result is representative of the entire population of children. Like the results based on twins and

[13]In the case of of the 2+ group we thus compare first and second born children from families which did not have a twin at the second birth to first and second born children from families which did have a twin at the second birth.

pre-twins, this result, while interesting, is less conservative than our main specification.

5.3 Estimating Q-Q Bounds

The results from the previous subsection provide consistent estimates of β_1 via 2SLS *if* the full set of controls completely account for those characteristics and behaviours which predict giving live birth to twins. However, given that we have shown that twinning is predicted by a wide range of health behaviours, and given that maternal health variables in this dataset are limited, it seems unlikely that all relevant variables are included in these specifications. As such, we turn to Conley et al.'s (2012) methodology to estimate bounds for the Q-Q trade-off.

As described in section 4.2, this involves the definition of some prior over the sign that the twin instrument takes in the structural equation 9. Results are displayed in figures 6 and 7, and in table 8. At each point on the horizontal axis of figures 6 and 7, the bounds for β_1 are displayed, along with the corresponding point estimate under the assumption that γ is distributed $U(0, \delta)$. Dashed lines present the 95% confidence interval, while the solid line represents the point estimate.

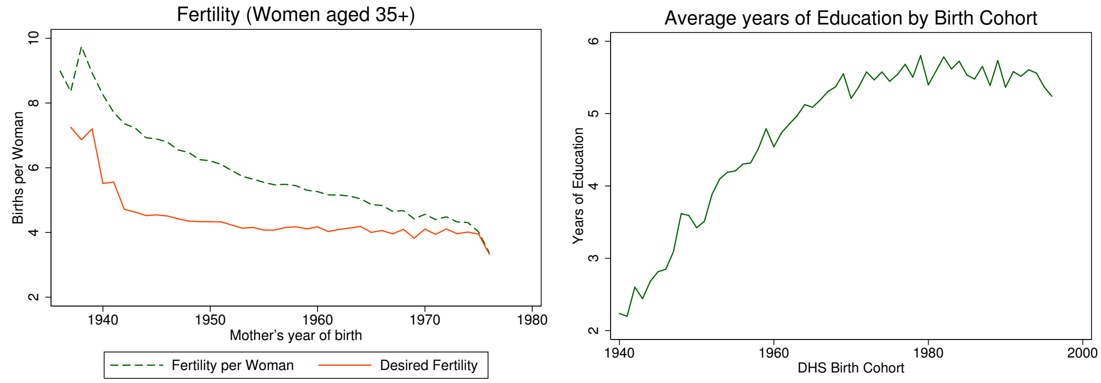
Table 8 provides these bounds at a particular point on this graph. This point corresponds to an assumption of $\gamma \in [0, 2\hat{\delta}]$ and $\gamma \sim U(0, 2\hat{\delta})$ for the UCI and the LTZ approaches respectively, where $\hat{\delta}$ is estimated from equation 9 with DHS data. These results allow us to make informative statements about the Q-Q trade-off, as the entire bounds for $\beta_1 < 0$ for 2+, 3+ and 4+ estimates. Further, these results suggest that the magnitude of this trade-off may be large: and indeed larger than those estimated in many existing studies. For example, the effect of an additional child at birth order three on first- and second-born children (ie the 3+ group) has an effect between -1.5% to as much as -15% of a standard deviation in standardised educational attainment if the prior (based on DHS estimates) is true. While it is not the case that informative statements can always be made given the wide bound on estimates in some cases (particularly the five plus group), it is important to point out that in each case as the twin-exogeneity assumption is loosened, the point estimate of the Q-Q tradeoff becomes more (and considerably) negative.

6 Conclusion

Short.

Figures

Figure 1: Education and Fertility



(a) Trends in Fertility

(b) Trend in Education

Note to figure 1: Cohorts are made up of all individuals from the DHS who are over 35 years (for fertility), and over 15 years (for education). In each case the sample is restricted to those who have approximately completed fertility and education respectively.

Figure 2: Proportion of Twins by Birth Order

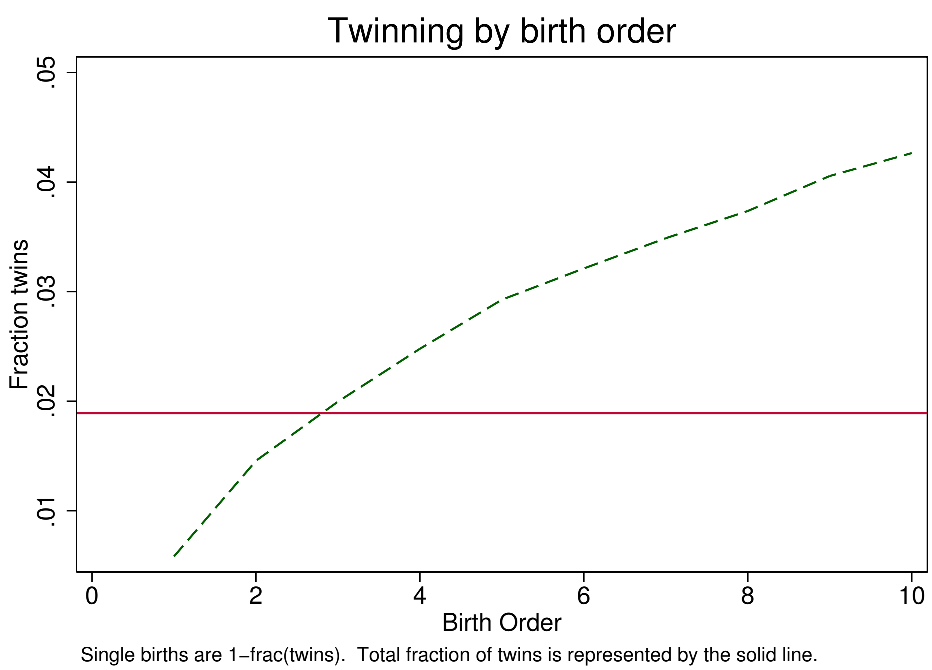


Figure 3: Twin Births and Total Fertility

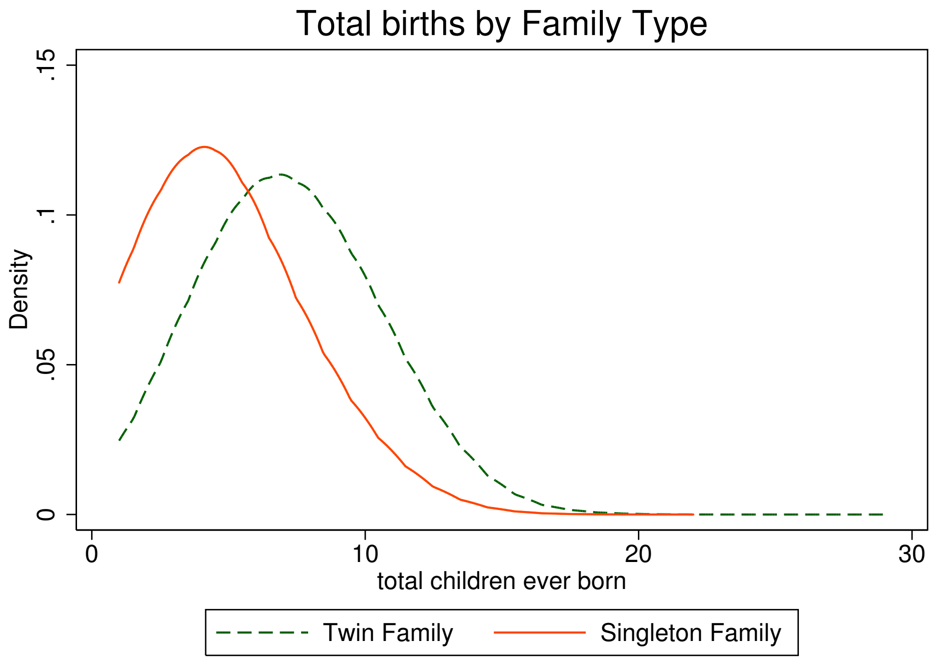


Figure 4: Distribution of Ideal Family Size

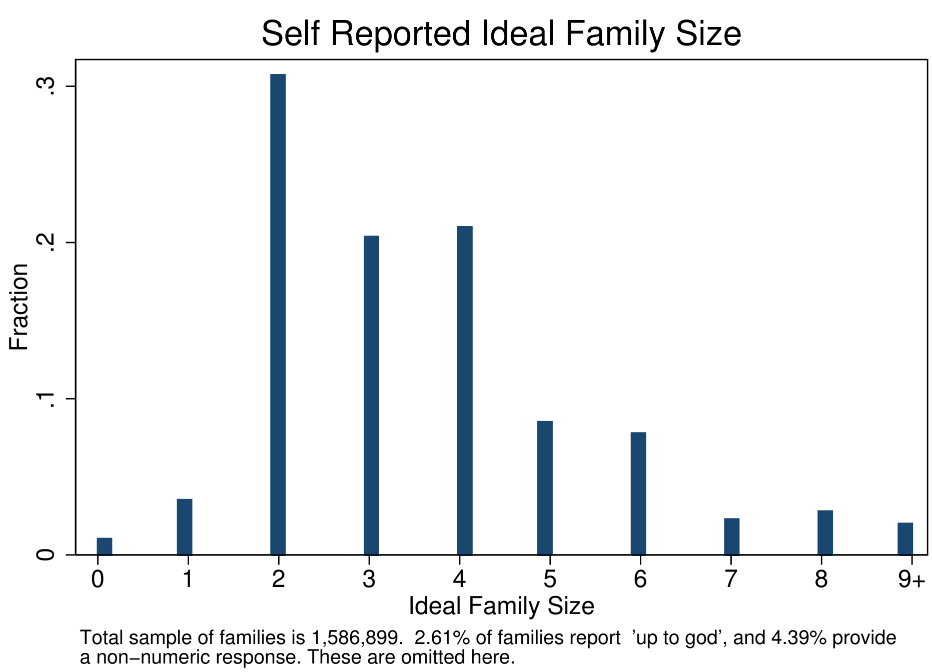


Figure 5: Ideal and Actual Fertility

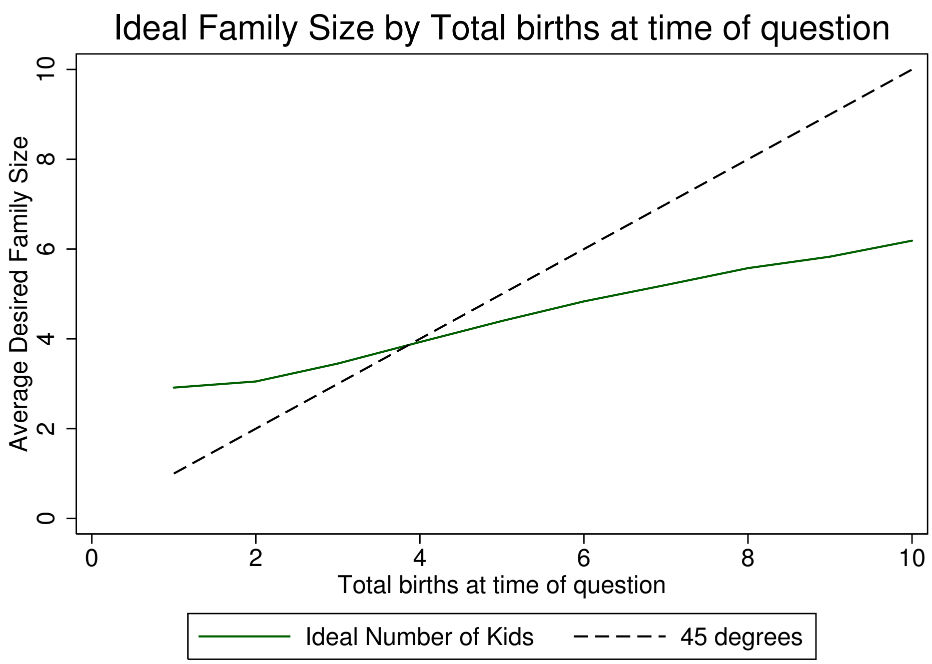
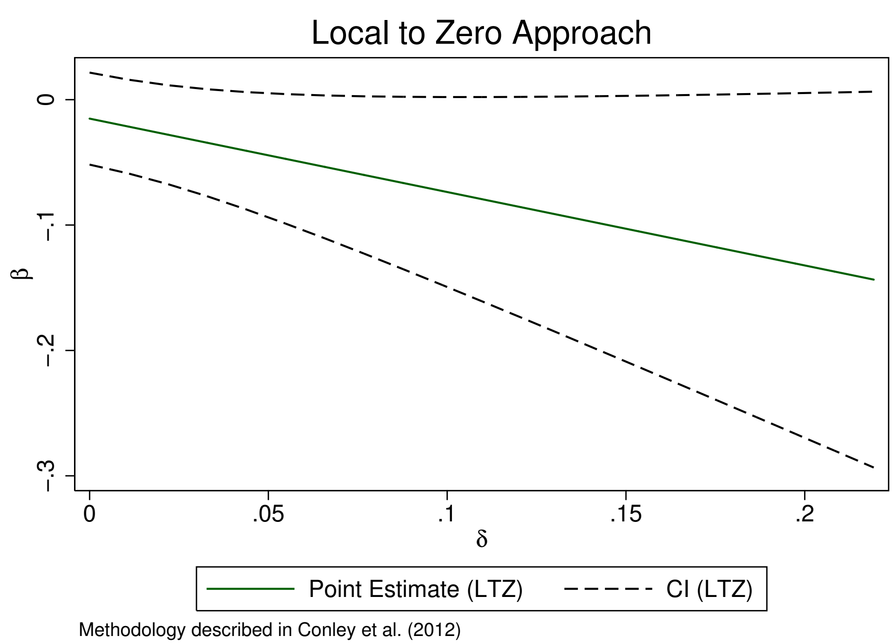
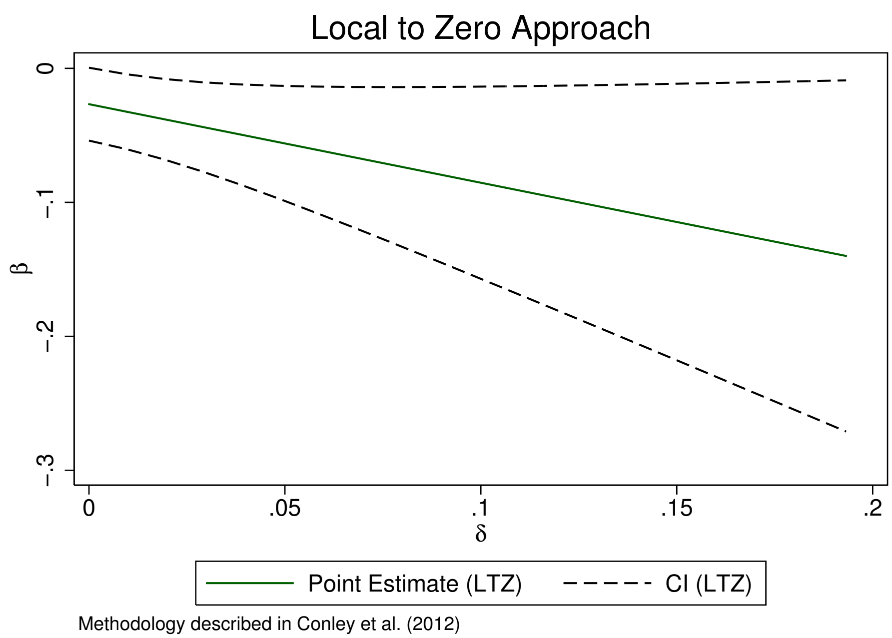


Figure 6: Relaxing Strict Exogeneity (two plus)



Note to figure 6: See note to Figure 7

Figure 7: Relaxing Strict Exogeneity (three plus)



Note to figure 7: Confidence intervals and point estimates are calculated according to Conley et al. (2012). Estimates reflect a range of priors regarding the validity of the exclusion restriction required to consistently estimate $\hat{\beta}_{fert}$ using twinning in a 2SLS framework. The local to zero (LTZ) approach applied here assumes that γ, the sign on the instrument when included in the first stage, is distributed $\gamma \sim U(0, \delta)$. Further discussion is provided in the body of the text and table 8.

Tables

Table 1: Estimation Samples

Sample	N Children	N Mothers	Mean Age	Min Age	Max Age
All	2,721,619	837,836	8.0318	0.0	18.0
Household	1,930,600	622,401	7.0671	0.0	18.0
2+	251,752	82,260	11.2544	6.0	18.0
3+	378,139	95,469	11.4676	6.0	18.0
4+	387,191	78,599	11.7289	6.0	18.0
5+	342,838	58,217	11.9868	6.0	18.0

NOTES: Full summary statistics are provided in table 2.

Table 2: Summary Statistics

	Low Income		Middle Income		
	Single	Twins	Single	Twins	All
FERTILITY					
Fertility	3.670	6.093	3.348	5.425	3.609
	(2.365)	(2.582)	(2.272)	(2.609)	(2.372)
Desired Family Size	4.182	5.296	3.340	4.128	3.892
	(2.500)	(2.832)	(2.083)	(2.498)	(2.403)
Fraction Twin	0.0194		0.0173		0.0185
	(0.1379)		(0.1306)		(0.1348)
Birth Order Twin	4.680		4.017		4.420
	(2.463)		(2.370)		(2.448)
MOTHER'S CHARACTERISTICS					
Age	30.92	34.18	31.99	34.94	31.40
	(7.980)	(7.166)	(8.105)	(7.156)	(8.038)
Education	3.984	3.337	6.817	6.115	5.030
	(4.370)	(4.033)	(4.794)	(5.048)	(4.735)
Height	155.6	157.7	155.6	157.2	155.7
	(7.084)	(6.987)	(6.956)	(6.957)	(7.042)
BMI	21.89	22.47	25.83	26.50	23.38
	(3.983)	(4.098)	(5.066)	(5.437)	(4.822)
Pr(BMI)<18.5	0.172	0.123	0.0344	0.0276	0.119
	(0.377)	(0.328)	(0.182)	(0.164)	(0.324)
Actual Births>Desired	0.297	0.513	0.319	0.567	0.311
	(0.457)	(0.500)	(0.466)	(0.495)	(0.463)
CHILDREN'S OUTCOMES					
Education (Years)	3.695	3.212	5.438	4.999	4.465
	(3.581)	(3.270)	(3.859)	(3.734)	(3.805)
Education (Z-Score)	-0.00869	-0.0130	0.0121	-0.0366	0.000177
	(1.001)	(0.961)	(0.998)	(0.987)	(1.000)
No Education (Percent)	0.200	0.213	0.0634	0.0766	0.140
	(0.400)	(0.409)	(0.244)	(0.266)	(0.346)
Infant Mortality	0.0860	0.165	0.0489	0.113	0.0758
	(0.280)	(0.371)	(0.216)	(0.316)	(0.265)
Child Mortality	0.122	0.208	0.0616	0.131	0.102
	(0.327)	(0.406)	(0.240)	(0.338)	(0.303)
Number of Countries	42	42	34	34	68
Number of Mothers	491,905	7,457	297,413	4,317	850,032
Number of Children (Education)	1,176,513	25,003	714,751	14,333	1,930,600
Number of Children (Ever Born)	1,716,247	43,866	940,204	21,302	2,721,619

NOTES: Summary statistics are presented for the full estimation sample consisting of all children 18 years of age and under born to the 850,032 mothers responding to any publicly available DHS survey. Group means are presented with standard deviation below in parenthesis. Education is reported as total years attained, and Z-score presents educational attainment relative to country and cohort (mean 0, std deviation 1). Infant mortality refers to the proportion of children who die before 1 year of age, while child mortality refers to the proportion who die before 5 years. Maternal height is reported in centimetres, and BMI is weight in kilograms over height in metres squared. For a full list of country and years of survey, see appendix table 10.

Table 3: Test of Balance of Observables: Twins versus Non-twins

	Non-Twin Family	Twin Family	Diff. (Diff. SE)
Total Fertilty	4.457	6.786	-2.328***
			(0.0252)
Desired Fertility	4.452	5.563	-1.112***
			(0.0309)
Age First Birth	19.16	18.98	0.179***
			(0.0421)
Mother's Education	4.295	3.379	0.916***
			(0.0519)
Father's Education	5.534	4.732	0.801***
			(0.0579)
Mother's Height	156.1	157.9	-1.741***
			(0.0840)
Pr(BMI < 18.5)	0.119	0.0931	0.0262***
			(0.00377)
Prenatal care (doctor)	0.314	0.212	0.102***
			(0.00539)
Prenatal care (nurse)	0.454	0.524	-0.0702***
			(0.00582)
Prenatal care (none)	0.190	0.179	0.0114*
			(0.00458)
Mother's Age	25.56	27.70	-2.134***
			(0.0559)
Child Mortality	0.0128	0.0303	-0.0175***
			(0.00111)
Infant Mortality	0.00690	0.0187	-0.0118***
			(0.000703)
Wealth Quintile 1	0.258	0.272	-0.0138**
			(0.00512)
Wealth Quintile 2	0.219	0.226	-0.00752
			(0.00484)
Wealth Quintile 3	0.198	0.201	-0.00327
			(0.00466)
Wealth Quintile 4	0.176	0.177	-0.000711
			(0.00446)
Wealth Quintile 5	0.150	0.124	0.0253***
			(0.00416)

NOTES: All variables are at the level of the mother. Education is measured in years, mother's height in centimetres, and BMI is weight in kilograms over height in metres squared. Wealth quintiles are determined by DHS methodology and are based on presence/absence of particular goods in the household. Diff. SE is calculated using a two-tailed t-test. Sample is identical to that in table 2.*p<0.1; **p<0.05; ***p<0.01

Table 4: Probability of Giving Birth to Twins

	(1)	(2)	(3)	(4)	(5)	(6)
Twin*100	All	Income		Time		Prenatal
		Low inc	Middle inc	1990-2013	1972-1989	
Age	0.594***	0.613***	0.554***	0.646***	0.314***	0.632***
	(0.029)	(0.036)	(0.050)	(0.033)	(0.075)	(0.040)
Age Squared	-0.008***	-0.008***	-0.007***	-0.009***	-0.003*	-0.009***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age First Birth	-0.053***	-0.093***	0.005	-0.052***	-0.055***	-0.041***
	(0.009)	(0.012)	(0.014)	(0.010)	(0.019)	(0.013)
Education (years)	0.040**	0.086***	-0.005	0.046**	0.021	-0.070**
	(0.017)	(0.022)	(0.029)	(0.020)	(0.034)	(0.028)
Education squared	-0.002	-0.006***	0.001	-0.002	0.001	0.003
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Height	0.058***	0.057***	0.059***	0.062***	0.043***	0.059***
	(0.004)	(0.005)	(0.007)	(0.005)	(0.008)	(0.007)
BMI	0.048***	0.063***	0.039***	0.045***	0.054***	0.045***
	(0.006)	(0.009)	(0.009)	(0.007)	(0.011)	(0.011)
Prenatal (Doctor)						0.913***
						(0.128)
Prenatal (Nurse)						0.073
						(0.108)
Prenatal (None)						-0.484***
						(0.132)
R-squared	0.01	0.01	0.01	0.01	0.01	0.01
Observations	1930600	1201516	729084	1524894	405706	615908

NOTES: All specifications include a full set of year of birth and country dummies, and are estimated as linear probability models. Twin is multiplied by 100 for presentation. Height is measured in cm and BMI is weight in kg divided by height in metres squared. l Prenatal care variables are only recoreded for recent births. As such, column (6) is estimated only for that subset of births where these observations are made. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 5: Test of hypothesis that women who bear twins have better prior health

Infant Mortality (per 100 births)	Base	+S&H	Observations
Treated (2+)	-2.065***	-2.110***	503785
	(0.212)	(0.213)	
Treated (3+)	-4.619***	-4.632***	686931
	(0.201)	(0.201)	
Treated (4+)	-4.257***	-4.243***	676303
	(0.183)	(0.183)	
Treated (5+)	-3.353***	-3.324***	587919
	(0.183)	(0.183)	

Notes: The sample for these regressions consist of all children who have been entirely exposed to the risk of infant mortality (ie those over 1 year of age). Subsamples 2+, 3+, 4+ and 5+ are generated to allow comparison of children born at similar birth orders. For a full description of these groups see the the body of the paper or notes to table 7. Treated=1 refers to children who are born before a twin while Treated=0 refers to children of similar birth orders not born before a twin. Base and S+H controls are described in table 7.*p<0.1; **p<0.05; ***p<0.01

Table 6: OLS Estimates of the Q-Q Trade-off

	Base Controls	+ Health	+Health &Socioec	Bord Controls	Desired	Altonji Ratio 1	Altonji Ratio 2
PANEL A: ALL COUNTRIES							
Fertility	-0.119***	-0.111***	-0.0781***	-0.0850***	-0.0745***	13.875	1.91
	(0.000919)	(0.000909)	(0.000875)	(0.00104)	(0.000935)		
Fertility×desire					-0.00643***		
					(0.000552)		
Observations	1,059,351	1,059,351	1,059,351	1,059,351	1,059,351		
R^2	0.091	0.106	0.159	0.160	0.159		
PANEL B: LOW INCOME							
Fertility	-0.117***	-0.108***	-0.0767***	-0.0840***	-0.0722***	12.0	1.903
	(0.00120)	(0.00117)	(0.00112)	(0.00130)	(0.00121)		
Fertility×desire					-0.00745***		
					(0.000684)		
Observations	651,145	651,145	651,145	651,145	651,145		
R^2	0.092	0.118	0.177	0.178	0.177		
PANEL C: MIDDLE INCOME							
Fertility	-0.125***	-0.119***	-0.0850***	-0.0910***	-0.0833***	19.833	2.125
	(0.00143)	(0.00143)	(0.00140)	(0.00170)	(0.00146)		
Fertility×desire					-0.00361***		
					(0.000931)		
Observations	408,206	408,206	408,206	408,206	408,206		
R^2	0.095	0.102	0.141	0.144	0.141		

NOTES: Base controls consist of child gender, mother's age and age squared mother's age at first birth, child age, country, and year of birth dummies. Socioeconomic augments 'Base' to include mother's education and education squared, and Health includes mother's height and BMI. "Desire" takes 1 if the child is born before the family reaches it's desired size, and 0 if the child is born after the desired size is reached. The Altonji et al. (2005) ratio determines how important unobservable factors must be compared with included observables to imply that the true effect of fertilty on educational attainment is equal to zero. Ratio 1 compares no controls to socioeconomic controls, while ratio 2 compares no controls to socioeconomic and health controls. Standard errors are clustered at the level of the mother. *p<0.1; **p<0.05; ***p<0.01

Table 7: Principal IV Results

	2+			3+			4+		
School Z-Score	Base	+H	+S&H	Base	+H	+S&H	Base	+H	+S&H
All									
Fertility	0.006	-0.026	-0.026	-0.004	-0.036	-0.038*	-0.017	-0.036	-0.035*
	(0.029)	(0.027)	(0.026)	(0.024)	(0.022)	(0.021)	(0.025)	(0.023)	(0.021)
Observations	249536	249536	249536	375987	375987	375987	385389	385389	385389
Low-Income									
Fertility	0.035	0.008	0.012	0.016	-0.016	-0.027	-0.011	-0.031	-0.024
	(0.034)	(0.032)	(0.031)	(0.030)	(0.028)	(0.026)	(0.029)	(0.027)	(0.025)
Observations	149602	149602	149602	232371	232371	232371	246622	246622	246622
Middle-Income									
Fertility	-0.065	-0.087*	-0.093**	-0.046	-0.079**	-0.067*	-0.027	-0.048	-0.054
	(0.053)	(0.049)	(0.047)	(0.040)	(0.036)	(0.035)	(0.043)	(0.040)	(0.037)
Observations	99934	99934	99934	143616	143616	143616	138767	138767	138767
Adjusted Fertility									
Fertility	0.017	-0.052	-0.055	-0.013	-0.073*	-0.077*	-0.033	-0.068	-0.066*
	(0.065)	(0.056)	(0.054)	(0.047)	(0.043)	(0.040)	(0.045)	(0.042)	(0.039)
Observations	249505	249505	249505	375957	375957	375957	385363	385363	385363
Twins and Pre-Twins									
Fertility	-0.021	-0.073***	-0.078***	-0.019	-0.062***	-0.067***	-0.018	-0.039**	-0.046**
	(0.024)	(0.021)	(0.020)	(0.020)	(0.018)	(0.018)	(0.021)	(0.019)	(0.018)
Observations	488815	488815	488815	563177	563177	563177	523197	523197	523197

NOTES: The two plus subsample refers to all first born children in families with at least two births. Three plus refers to first- and second-borns in families with at least three births, and four plus refers to first- to third-borns in families with at least four births. Each cell presents the coefficient of a 2SLS regression where fertility is instrumented by twinning at birth order two, three or four (for 2+, 3+ and 4+ respectively). Different rows of the table correspond to different sub-groups or specifications. In order these correspond to: all children, grouped by country income status, adjusting fertility to correct exclude children who did not survive to one year, and including both pre-twins *and* twins in the regression. Base controls include child age, mother's age, and mother's age at birth fixed effects plus country and year-of-birth FEs. In each case the sample is made up of all children aged between 6-18 years from families in the DHS who fulfill 2+ to 4+ requirements. First-stage results in the final panel correspond to the second stage in row 1. Full first stage results for each row are available in table 12. Standard errors are clustered by mother.*p<0.1; **p<0.05; ***p<0.01

Table 8: 'Plausibly Exogenous' Bounds

	UCI: $\gamma \in [0, \delta]$		LTZ: $\gamma \sim U(0, \delta)$	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Two Plus	-0.1859	0.0175	-0.1613	-0.0009
Three Plus	-0.1710	0.0007	-0.1527	-0.0134
Four Plus	-0.1538	-0.0072	-0.1380	-0.0199
Five Plus	-0.1372	0.0273	-0.1205	0.0139

NOTES: This table presents upper and lower bounds of a 95% confidence interval for the effects of family size on (standardised) children's education attainment. These are estimated by the methodology of Conley et al. (2012) under various priors about the direct effect that being from a twin family has on educational outcomes (γ). In the UCI (union of confidence interval) approach, it is assumed the true $\gamma \in [0, \delta]$, while in the LTZ (local to zero) approach it is assumed that $\gamma \sim U(0, \delta)$. In each case δ is estimated by including twinning in the first stage equation and observing the effect size $\hat{\gamma}$. Estimated $\hat{\gamma}$'s are (respectively for two plus to five plus): 0.1070, 0.0968, 0.0822, 0.0926.

Table 9: Q-Q IV Estimates by Gender

	Females				Males			
	Base	Socioec	Health	Obs.	Base	Socioec	Health	Obs.
Two Plus	0.005	-0.039	-0.037	122,414	0.010	-0.010	-0.015	127,122
	(0.043)	(0.039)	(0.038)		(0.040)	(0.038)	(0.036)	
Three Plus	-0.024	-0.056*	-0.052*	187,098	0.016	-0.015	-0.022	188,889
	(0.033)	(0.030)	(0.029)		(0.030)	(0.028)	(0.027)	
Four Plus	-0.029	-0.052*	-0.053**	192,714	-0.005	-0.020	-0.018	192,675
	(0.032)	(0.029)	(0.027)		(0.030)	(0.028)	(0.027)	

NOTES: Female or male refers to the gender of the index child of the regression. All regressions include full controls including socioeconomic and maternal health variables. The full lis of controls are available in the notes to table 7. Full IV results for male and female children are presented in table 11. Standard errors are clustered by mother.*p<0.1; **p<0.05; ***p<0.01

References

D. Almond, K. Y. Chay, and D. S. Lee. The costs of low birth weight. *The Quarterly Journal of Economics*, 120(3):1031–1083, August 2005.

J. G. Altonji, T. E. Elder, and C. R. Taber. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151–184, February 2005.

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

J. D. Angrist and W. N. Evans. Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, 88(3):450–77, June 1998.

G. S. Becker and H. G. Lewis. On the interaction between the quantity and quality of children. *Journal of Political Economy*, 81(2):S279–88, Part II, 1973.

G. S. Becker and N. Tomes. Child endowments and the quantity and quality of children. *Journal of Political Economy*, 84(4):S143–62, August 1976.

S. E. Black, P. J. Devereux, and K. G. Salvanes. 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, 2005.

H. Bleakley and F. Lange. Chronic Disease Burden and the Interaction of Education, Fertility, and Growth. *The Review of Economics and Statistics*, 91(1):52–65, February 2009.

N. Braakmann and J. Wildman. Fertility treatments and the use of twin births as an instrument for fertility. MPRA Paper 54106, University Library of Munich, Germany, Mar. 2014. URL `http://ideas.repec.org/p/pra/mprapa/54106.html`.

S. G. Bronars and J. Grogger. The economic consequences of unwed motherhood: Using twin births as a natural experiment. *The American Economic Review*, 84(5):1141–1156, 1994.

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

T. G. Conley, C. B. Hansen, and P. E. Rossi. Plausibly Exogenous. *The Review of Economics and Statistics*, 94(1):260–272, February 2012.

E. Fitzsimons and B. Malde. Empirically probing the quantity-quality model. IFS Working Papers W10/20, Institute for Fiscal Studies, Sep 2010.

J. G. Hall. Twinning. *The Lancet*, 362(6736):735–43, August 2003.

J. P. Jacobsen, J. W. P. III, and J. L. Rosenbloom. The effects of childbearing on married women's labor supply and earnings: Using twin births as a natural experiment. *Journal of Human Resources*, 34(3):449–474, 1999.

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

Z. Li, J. Gindler, and H. Wang. Folic acid supplements during early pregnancy and likelihood of multiple births: a population-based cohort study. *The Lancet*, 361:380–84, 2003.

N. Qian. Quantity-quality and the one child policy: The only-child disadvantage in school enrollment in rural China. NBER Working Papers 14973, National Bureau of Economic Research, Inc, May 2009.

M. R. Rosenzweig and K. I. Wolpin. Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica*, 48(1):227–40, January 1980a.

M. R. Rosenzweig and K. I. Wolpin. Life-cycle labor supply and fertility: Causal inferences from household models. *Journal of Political Economy*, 88(2):pp. 328–348, 1980b.

M. R. Rosenzweig and J. Zhang. Do population control policies induce more human capital investment? twins, birth weight and China's one-child policy. *Review of Economic Studies*, 76 (3):1149–1174, 07 2009.

C. Sanhueza. Tamaño de familia y orden de nacimiento en Chile: Usando gemelos como experimento natural. *Revista de Analisis Económico Economic Analysis Review*, 24(2):133–152, Dec 2009.

Appendices

A Appendix Tables

Table 10: Full Survey Countries and Years

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

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

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

Table 11: Instrumental Variables Estimates: Female and Male Children

	Females			Males		
	2+	3+	4+	2+	3+	4+
All						
Fertility	-0.037	-0.052*	-0.053**	-0.015	-0.022	-0.018
	(0.038)	(0.029)	(0.027)	(0.036)	(0.027)	(0.027)
Low-Income Countries						
Fertility	0.017	-0.020	-0.027	0.013	-0.034	-0.013
	(0.041)	(0.035)	(0.030)	(0.045)	(0.033)	(0.033)
Middle-Income Countries						
Fertility	-0.119	-0.120**	-0.100**	-0.064	-0.018	-0.025
	(0.074)	(0.049)	(0.050)	(0.058)	(0.046)	(0.046)
Adjusted Fertility						
Fertility	-0.072	-0.105*	-0.101**	-0.034	-0.046	-0.034
	(0.078)	(0.055)	(0.050)	(0.076)	(0.052)	(0.049)
Twins and Pre-Twins						
Fertility	-0.073***	-0.088***	-0.062***	-0.081***	-0.051**	-0.027
	(0.027)	(0.022)	(0.023)	(0.027)	(0.024)	(0.024)
First Stage						
All						
Twin	0.846***	0.824***	0.839***	0.794***	0.828***	0.882***
	(0.041)	(0.033)	(0.032)	(0.039)	(0.033)	(0.033)
Adjusted Fertility						
Twin	0.410***	0.430***	0.448***	0.375***	0.424***	0.487***
	(0.039)	(0.033)	(0.034)	(0.040)	(0.034)	(0.033)

Notes: Each cell presents the coefficient from a 2SLS regression of standardised educational attainment on fertility. 2+, 3+ and 4+ refer to the birth orders of children included in the regression. For a full description of these groups see table 7. Each regression includes full controls including maternal health and socioeconomic variables. The sample is made up of all children aged between 6-18 years from families in the DHS who fulfill birth order and gender requirements indicated in the header. Standard errors are clustered by mother.*p<0.1; **p<0.05; ***p<0.01

Table 12: First Stage Results

Fertility	2+			3+			4+		
	Base	+H	+S&H	Base	+H	+S&H	Base	+H	+S&H
All									
Twin	0.776***	0.821***	0.822***	0.794***	0.827***	0.826***	0.840***	0.859***	0.861***
	(0.031)	(0.029)	(0.028)	(0.027)	(0.027)	(0.026)	(0.027)	(0.027)	(0.026)
Observations	249536	249536	249536	249536	249536	249536	249536	249536	249536
Low-Income									
Twin	0.826***	0.853***	0.848***	0.810***	0.828***	0.834***	0.867***	0.873***	0.869***
	(0.038)	(0.038)	(0.037)	(0.033)	(0.033)	(0.032)	(0.033)	(0.033)	(0.033)
Observations	149602	149602	149602	149602	149602	149602	149602	149602	149602
Middle-Income									
Twin	0.718***	0.774***	0.784***	0.757***	0.817***	0.801***	0.783***	0.831***	0.839***
	(0.050)	(0.045)	(0.043)	(0.046)	(0.045)	(0.043)	(0.047)	(0.044)	(0.042)
Observations	99934	99934	99934	99934	99934	99934	99934	99934	99934
Adjusted Fertility									
Twin	0.354***	0.393***	0.395***	0.403***	0.428***	0.427***	0.453***	0.467***	0.468***
	(0.028)	(0.028)	(0.028)	(0.026)	(0.026)	(0.026)	(0.027)	(0.027)	(0.027)
Observations	249505	249505	249505	249505	249505	249505	249505	249505	249505
Twins and Pre-Twins									
Twin	0.727***	0.782***	0.788***	0.809***	0.828***	0.832***	0.853***	0.855***	0.859***
	(0.027)	(0.025)	(0.025)	(0.027)	(0.026)	(0.026)	(0.027)	(0.025)	(0.025)
Observations	488815	488815	488815	488815	488815	488815	488815	488815	488815

Notes: Each cell represents the coefficient from the first-stage of a two-stage regression. The first-stage represents the effect of twinning at parity N on total fertility where N is 2, 3 or 4 for 2+, 3+ and 4+ groups respectively. The 2+ group includes all first borns in families with at least 2 births, the 3+ group includes first and second borns in families with at least 3 births, and the 4+ group includes all first to third borns in families with at least four births. In each regressions the sample is made up of all children aged between 6-18 years from families in the DHS who fulfill these birth order conditions. Controls in each case are identical to those described in table 7. Standard errors are clustered at the level of the mother.*p<0.1; **p<0.05; ***p<0.01