The Twin Instrument*

Sonia Bhalotra† Damian Clarke‡

April 7, 2015

Preliminary Draft

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. 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. The neglect of this fact in the existing literature will tend to lead to under-estimation of the quantity-quality (Q–Q) trade-off and so could contribute to explaining the negative results in the literature. Our contention that women who produce twin births are positively selected is demonstrated using data from richer and poorer countries. Using large samples of microdata from developing countries and from the USA which include indicators of maternal characteristics including health, 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. These results have important implications for twin studies in all contexts examined here.

JEL codes: J12,J13,C13,D13,I12.

*We are grateful to Paul Devereux, James Fenske, Cheti Nicoletti, Carol Propper, Atheen Venkataramani, Marcos Vera-Hernandez and Frank Windmeijer, along with seminar audiences and discussants at CMPO Bristol, ESPE, NEUDC, CSAE, The University of Essex, The University of Oxford, the Barcelona GSE summer forum, PacDev, and the RES conference for helpful comments. We thank Emilia Del Bono, Climent Quintana-Domeque, Pedro Ródenas, Libertad González, Anna Aevarsdottir, Martin Foureaux Koppensteiner and Ryan Palmer who have very kindly shared data and/or source code from their work.

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

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

1 Introduction

Since the pioneering work of Rosenzweig and Wolpin (1980a), economists have attempted to leverage the occurrence of twin births to estimate the effect of family size on child outcomes. If twin births occur at random, their occurrence constitutes a fertility shock that is uncorrelated with family characteristics, including parental preferences, and other unobservables which may be related to child quality. This provides the exogenous variation (in quantity) required to estimate the quantity-quality model of Becker (1960); Becker and Lewis (1973); Becker and Tomes (1976). The essential idea is that the shadow price of child quality is increasing in child quality and *vice versa*. By comparing those families who unexpectedly produced additional children to those who always produced one child per birth, it is argued that we can isolate the direct effect of an individual's sibling size on her human capital attainment.

However, the consistent estimation of this effect is based upon the untestable assumption that twinning is exogenous. This requires not only that twin conceptions are randomly assigned to families, but also that taking a twin conception to term does not depend upon a woman's behaviours during pregnancy or on her endowments prior to pregnancy. This is at odds with the evidence.

We show that endowments and behaviours affect the chances of twins being born. In data from a large sample of developing countries, we show that taller women and women with a higher body-mass index are significantly more likely to give birth to twins.[1] Maternal health is likely to be an especially significant determinant of birth outcomes in poorer countries where many women are chronically under-nourished (an indicator of which is their final stature), exhibit anemia and low-BMI and are prone to infections. In these conditions only relatively healthy women will have the resources to support a successful twin pregnancy. However, our critique applies to richer countries too. Using administrative data from the Scotland, the USA, Sweden and Spain, and survey data from the UK and Chile, we find that women who are taller and less likely to engage in risky behaviours such as smoking, drug taking or alchol consumption during pregnancy are significantly more likely to have a (live) twin birth.[2]

Overall, our argument is that women who give birth to twins are positively selected and that the common tendency to ignore this will result in under-estimation of the Q–Q trade-off. We focus upon maternal health because no previous work has highlighted it as a determinant

[1] Height is an index of the stock of health of a woman and is a function of investments over her growth period and, especially, the early years of her life (see references in Bhalotra and Rawlings (2013)).

[2] Simiarly, it is well known that fertility treatments increase the likelihood of twinning. While we do not discuss this extensively here, we note that this is likely to be less of an identification challenge given that IVF treatment is an entirely observable behaviour.

of twinning and because it is inherently impossible to fully control for. Even if we had data that included all of the indicators of health we mention above, we may not observe whether women skip breakfast (Mazumder and Seeskin, 2014), whether they are stressed (Black et al. (2014) and references therein), whether they seek and adhere to antenatal care and so on. Our contention adds a novel twist to a recent literature which suggests that mothers' health and fetal environment matter and may alter the birth weight of children, the sex ratio at birth and a range of future human capital outcomes (Almond et al., 2011; Bhalotra and Rawlings, 2013; Barker, 1995). Like birth weight and the probability of a boy relative to a girl birth, the probability of a twin birth is increasing in the health of the mother/the fetal environment.

The emerging consensus in the literature using twin births to test the Q–Q model is that there is in fact no significant or substantial trade-off between fertility and investments in children. In this paper we suggest that this result may, in principle, follow from the bias created by twin-mothers being positively selected. We re-examine the validity of these results, accounting for the innovation discussed previously. If twin births are not truly exogenous, but instead depend upon maternal health stocks and behaviours, there is an identification challenge. Specifically, if healthier mothers are both more likely to give birth to twins, and more likely to invest in child human capital in later life, existing estimates may significantly underestimate the size of the Q–Q trade-off. When ignoring these considerations we find that the effect of an additional birth on human capital attainment (education) is minor, or even weakly positive. However, when taking into account this innovation we find that a significant trade-off does exist, and that an additional birth reduces standardised schooling behaviour by at least 4% of a standard deviation.

The estimated effect of $\sim 4\%$ of an s.d. is at most the lower bound for the true size of the Q–Q trade-off. Given that we suggest that maternal health predicts twinning, and given that maternal health is multidimensional in nature and difficult to observe fully, we will only ever be able to include a partial set of controls to account for the inconsistency in IV estimates. As such, we examine a number of methods to estimate plausible bounds on the Q–Q trade-off. First, we argue that the true estimate is bounded by the OLS and IV estimtes. We also follow Conley et al. (2012) in conceptualizing twins as a plausibly exogenous event, and derive estimates by assuming that the traditional exclusion restriction is 'close to' holding.

Empirically, work which further probes the Q–Q trade-off is important, especially in the developing country setting on which we place considerable focus in this paper. In order to conduct empirical tests, we construct a large microdata set from 68 developing countries with observations on more than 2.5 million children and nearly 1 million mothers. The macro level trends in this data suggest that educational attainment has risen considerably while completed and desired fertility has fallen sharply over the past 50 years (see figures 1a and 1b). Similar effects have been described extensively in the economic and demographic literature (*eg* Hanushek (1992)). It is of considerable relevance to researchers and to policy makers to determine whether such a trend is (at least partially) causal. In practical terms, a significant number of government bodies report that family planning is considered an important concern,[3] and in some cases these concerns have resulted in aggressive, and at times group-specific, fertility control policies.

This paper unfolds as follows. In the next section we discuss the existing literature which estimates the Q-Q model using twins. We then discuss the methodology that we will use to examine twinning, and bound the Q–Q trade-off. Section 4 discusses our data sources and estimation samples, and section 5 presents results. We briefly conclude in the final section.

2 Twins

The ocurrence of twins has fascinated people, not least of all social scientists, for as long as recorded history exists. Stories of Romulus and Remus, the mythological founders of Rome, date from at least the fourth century BC. However, the first use of twins as an exogenous increase in family size in the economic literature came much later, in Rosenzweig and Wolpin (1980a).

Rosenzweig and Wolpin (1980a) proposed to incorporate twins into an estimation strategy in order to circumnavigate problems of the joint determination of child quantity and quality first raised by Becker (1960); Becker and Lewis (1973); Becker and Tomes (1976). Under a number of assumptions, they show that twins as an increase to family size will be sufficient to directly identify the interaction between quality and quantity. As well as the occurrence of twins pushing at least some families above their desired total fertility, this requires that twin births are random. Empirically then, Rosenzweig and Wolpin (1980a) estimate the Q–Q trade-off, accounting for the increase in rates of twinning by parity (a biological relationship)[4], and by total number of births (a mechanical relationship). Beyond these variables, twinning must be random to produce consistent estimates.[5]

The twin instrument has been widely used since Rosenzweig and Wolpin's (1980a) initial work. As well as its use in the estimation of the quantity-quality trade-off (Black et al. (2005);

[3] A recent survey of national governments suggests that fertility was perceived as too high in 50% of developing countries, with this figure rising to 86% among the least developed countries United Nations (2010).

[4] Such a relationship is an empirical regularity in all data examined. Rosenzweig and Wolpin (1980a) report rates which increase by parity in USA. In DHS data a similar pattern is observed, as presented in figure 2.

[5] It is important to note that in their original paper this *is* tested formally. The authors report that a joint test of the effect of farm size, non-farm earnings and parental education on the rate of twins per births did *not* allow for the rejection of no statistical effect.

Cáceres-Delpiano (2006); Li et al. (2008); Angrist et al. (2010), among others), it has been used to estimate the effect of childbearing on female labour force participation (Rosenzweig and Wolpin, 1980b; Jacobsen et al., 1999; Angrist and Evans, 1998), and the effect of unwed childbearing on marriage market outcomes, poverty and welfare receipt (Bronars and Grogger, 1994). Typically, estimation is based on two-stage least squares, with the set of controls included in the first- and second-stage varying slightly over time.[6]

Table 1 documents the principal studies in the Q–Q trade-off literature where twins are employed. Along with the data sample and time period under study, we list the set of controls included in each case. The more recent wave of these studies make a similar conditional randomness assumption, however refurbish the controls in IV estimates to include variables such as a mother's race and her educational attainment. 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 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.

The most comprehensive controls considered in the economic literature (often by necessity due to data restrictions), include maternal age, parental education and measures of income and/or goods. However, recent evidence from the medical literature points to the fact that twinning may depend more deeply upon a mother's health behaviours or endowments. 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))

Unrelated to health measures *per se*, recent studies seek to control for the fact that multiple births are correlated with fertility treatments. Typically, such an analysis 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

[6]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.

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.

Finally, the twin instrument is not without critique for other reasons. Existing 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[7]—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 (2014). We turn to bounds estimation in section 5.4 of this paper.

3 Methodology

Empirical analyses of the quality-quantity trade-off focus on producing consistent estimates of β_1 in the following equation:

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

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 with 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).

[7]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. In our data samples similar endowment differences are observed. For example, appendix figure A1 documents the much larger average reported birth size of twins versus singletons in DHS data. Birth weight figures show similar patterns.

3.1 Quantity-Quality with Twins

One proposed solution has been to employ 2SLS estimation, where fertility is is instrumented using twin births.[8] The corresponding first stage is:

$$fert_j = \pi_0 + \pi_1 twins_j + \boldsymbol{X\pi} + \nu_j, \tag{2}$$

where $twins_j$ is an indicator for whether the n^{th} birth in a family is a twin birth. As described in section 4.2, 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.[9] The logic, in quasi-experimental terms, is that existing children (the subjects) are randomly assigned either one (control group) or two (treatment group) siblings.

Consistent estimation of β_1 can thus proceed provided (among other things) that instrumental validity holds:

$$\text{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} twin_j u_{ij} = 0 \tag{3}$$

The typical challenge in IV estimates arises when considering (3); as the error term u_{ij} consists, by nature, of unobservable components, whether or not the equality holds cannot be tested formally. There is however nothing which stops us from partially testing (3) by removing a subset of observable components from u_{ij} and testing whether their (conditional) correlation with $twin_j$ is significantly different from zero. The error term u in (1) is a function of a large number of elements:

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

While many of the relevant elements are either completely or partially unobservable, some of these variables, such as maternal education and partial measures of health stocks and behaviours, can be observed. Thus a partial test of the twin methodology consists of estimating the following regression.

$$twin_j = \alpha_0 + \boldsymbol{X\alpha}_1 + \boldsymbol{S\alpha}_2 + \boldsymbol{H\alpha}_3 + \varepsilon_j. \tag{5}$$

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.

[8] 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)

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

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

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

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. 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 in their children during 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) $\boldsymbol{\alpha_2} > 0$.

An alternative test of whether twins appear to be as good as random consists of comparing women who give birth to twins with those who give birth to singletons *before* these children are born. If twin births occur randomly in the population, the two groups of mothers should appear identical before these births occur. In order to compare health stocks before twins, we run tests comparing the rate of infant mortality—a completely predetermined variable—of children in each of the $n+$ groups described above. If, as we contend, healthier mothers are more likely to give birth to twins, this should be captured in lower infant mortality rates (IMR) in early births.

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

$$u_{ij} = u^S_{ij} + u^H_{ij} + u^*_{ij}.$$

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

$$\hat{\beta}^{IV}_1 = \beta_1 + \text{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} twin_j(u^S_{ij} + u^H_{ij} + u^*_{ij}) \tag{7}$$

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

$$\hat{\beta}_1^{IV,S+H} = \beta_1 + \text{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} twin_j u_{ij}^*, \tag{8}$$

where the superscript $S+H$ signifies that socioeconomic and health variables have been included as additional controls and, correspondingly, have been removed from the stochastic error term. What's more, if both the likelihood that a woman takes twins to term and a family's subsequent 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:

$$\hat{\beta}_1^{IV} > \hat{\beta}_1^{IV,H} > \hat{\beta}_1^{IV,S+H} > \beta_1. \tag{9}$$

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.

3.2 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:

$$educ_{ij} = \beta_1 fert_j + \gamma twin_j + \boldsymbol{X\beta} + u_{ij}. \tag{10}$$

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. Specifically, we are concerned that healthier mothers are more likely to give birth to twins, and, all else constant, healthier mothers are more likely to be able to invest more in children post-pregnancy. Thus, γ, the coefficient on twins in (10), reflects the interaction between the partial correlation of a mother's health and her likelihood of giving birth to twins, which we denote ϕ_t, and the partial correlation between her health and child quality, denoted ϕ_q.

In this paper we estimate β_1 under a range of assumptions regarding the true nature of γ. Firstly we estimate β_1 by simply assuming a support assumption for γ: namely that γ falls between zero (implying instrumental validity) and some (positive) number δ:

$$\gamma \in [0, \delta]. \tag{11}$$

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_1 over the entire support of γ. This UCI, then, provides bounds for β_1 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:

$$\gamma \sim U(0, \delta). \tag{12}$$

This stronger assumption allows for a tighter estimate of the bounds on β_1. 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 (11) and (12) depend upon the values of δ that we believe hold in the case of twinning and the Q-Q equation. In order to form a prior for γ, we turn to a specific case which allows us to causally estimate each of ϕ_t and ϕ_q, and hence the coefficient γ. By using a well documented shock to mother's health—the arrival of sulfonamide antibiotics in the USA in 1937 (Bhalotra and Venkataramani, 2014; Jayachandran et al., 2010)—we can observe the causal effect of this positive health shock on the quality of her children and on the likelihood of giving birth to twins during her fertile life. By taking the ratio of these effects ϕ_t/ϕ_q, we isolate the direct correlation between twinning and child quality. We provide a more comprehensive discussion and derivation in of this method in appendix B, and return to estimate of γ and bounds on our estimand of interest β_1 in section 5.4.

4 Data and Estimation Samples

We consult two data sets for our main IV and OLS analysis, and a large number of auxiliary datasets when only considering the link between twinning and maternal health. In order to estimate the (health augmented) specification 1, we require information on each child's siblings, his or her mother's characteristics, as well as the child's eventual 'quality' measure. We focus our empirical tests on two principal datasets which contain all of these variables. Firstly, the Demographic and Health Surveys, which have been applied over 20 years in a range of developing countries, and secondly, the United States National Health Interview Surveys (NHIS). In what follows, we describe the characteristics of these datasets. A comprehensive description of all data, including that which is used only to illustrate twinning and its relation to maternal characteristics, is provided in data appendix A.

4.1 Data

4.1.1 The 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 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. We merge these two databases (all children who live in the same household as their mother merge without loss). 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 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. We drop from our sample children aged 18 and over, due to concerns that these will *not* be representative of the general population.

We pool all publicly available DHS data resulting in microdata on 3,297,318 children ever-born to women who responded fully to any DHS survey. A full list of the DHS countries and years of surveys which make up this sample is provided in an online appendix (table B1). Of the 3,297,318 offspring reported in survey data, 2,033,510 remain living in the same household as their mother. The majority of these 2,033,510 children are aged 18 and under (92.96%) and hence make up our principal estimation sample (in future we will refer to this as the 'household sample'). The remaining 1,263,808 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 (53.9%) were aged over 18 or had died prior to the date of survey.[10]

4.1.2 The NHIS

The National Health Interview Survey (NHIS) is a yearly survey, conducted from 1957 and ongoing as at 2014, with participants drawn from each of the 50 US States as well as the District of Columbia each year.[11] We pool all survey data from 2000 until 2013, resulting in data on 127,009 mothers and 246,646 children. We focus on this period given that prior to 2000, changes in a number of key variables make it difficult to compare between years, and post-1996 the survey was considerably revised.

Each set of surveys is collected at the level of the household. For our analysis we use all households which consist of a biological mother and her children, whether or not any father is present. For all children who remain in the household, the survey records total fertility. We infer twin status by assuming that all children who share a birth month, birth year and biological mother must be twins. For each child and mother, we have a number of measures of usage of health care along with a self-reported measure for health status, whether or not the mother smokes, and the level of completed education (at the time of the survey) of mothers and children. Once again, we subset to children aged below 18, and for education measures, children who are aged above 6 years old, and hence who are able to be enrolled in school. Descriptive statistics of this and DHS data is provided in section 4.3.

[10] 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 2.7 years older, born to less educated and younger mothers, and are slightly more likely to be males.

[11] The NHIS has a survey design to oversample Hispanic and African American people. We use NHIS-specific probability weights in all analyses.

4.2 Estimation Samples

The quasi-experimental variation exploited in twin studies is to leverage the effect of an unexpected additional child on siblings who were born before the extra birth. Thus, all first-born children in families of at least two births are split into two groups: the treatment group, which consists of first-born children in a family whose second birth results in two children (twins), and a control group consisting of first-born children in families where the second birth results in just one child.

For our main IV specification, we follow the existing literature in defining birth-order specific estimation samples, as laid out in the preceding paragraph. These samples are referred to as the 2+, 3+, and 4+ samples. These samples are defined $\forall\ n \in \{2, 3, 4\}$ such that they include first-born to $n-1$ born children in families with at least n births.[12] As an example, the 2+ sample (described in the previous paragraph) 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 (see figure 2). 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.

4.3 Descriptive Statistics

Table 2 provides summary statistics for DHS data, and table 3 describes NHIS 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 number of observations at each level is provided at the bottom of the table.

For DHS data, survey countries are classified according to country income level in order to allow for a disaggregation of Q–Q results by income group.[13] We present summary statistics by

[12] 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 a birth-order group, 4+.

[13] 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 B1.

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 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.

Similar patterns are observed when turning to NHIS data. Twin mothers have (unconditionally) higher education, and greater health stocks as measured by BMI, percent underweight (BMI<18.5) and self-reported health. This is despite having higher total fertility, and being somewhat older.

Child 'quality' is measured using each child's educational attainment. Our principal outcome variable in each case 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 cohort or reference. In the case of DHS data, this cohort is made up of all children in the same country and birth cohort, while in NHIS data, it is made up of children with the same month and year of birth. The use of a standardised score rather than just total years of education allows us to express all effect-sizes in terms of a one standard deviation increase in total educational attainment.

5 Results

5.1 Twinning

In table 4 we present results of a regression of a child's twin status (one if a twin, zero if a singleton) on their mother's health, education, and a range of other demographic and family characteristics.[14] These results suggest that twin births are not random, even after conditioning on maternal age and child birth order as is typical in the recent twin literature summarised in table 1. 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 3 that higher investments (for example in maternal health) required to maintain multiple healthy fetuses in utero may result in non-random twin births. We return to the mechanism by which twin selection may occur in the following section.

Initially, results from the pooled DHS data are presented as this provides a particularly large sample with which to test the hypothesis of twin exogeneity. This is presented in table 4 column (1) and provides considerable evidence that live twin births are related to family choice variables such as education (tests for the joint significance of socioeconomic variables and health variables are rejected with p-values of <0.00). Regressions displayed here are estimated by OLS, however are robust to alternative functional forms and estimation methods.[15]

Columns (2)–(5) suggest that these results are not due only to the most low income countries, or the post-IVF time period, especially when considering maternal health. Given that the frequency of multiple births increases in cases where the mother undergoes fertility treatment, column (5) presents regression results for births in a period not potentially affected by IVF.[16] 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 BMI: measures of health stocks, are positively correlated with twinning regardless of the sample. Similarly, this result is not driven by a particular country or region. Figure 4 provides evidence that healthy women (as proxied by height) are significantly more likely to have twins in nearly all of the 68 countries included in DHS surveys. Along with higher average rates of twinning in countries with taller women, a positive within-country gradient exists, with taller women in a given country more likely to have twins than their shorter counterparts. The size of DHS estimates are considerable. Increasing a woman's height by 1 standard deviation increases the probability of twinning by 0.44% (as compared to a mean rate of twins of 1.85%).

[14] 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.

[15] 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. Alternatively, rather than running a regression we can run (unconditional) balance of characteristics tests by twin status. These are available in the online appendix (table B2). The findings are similar.

[16] 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.

Results for the infant mortality test described in the previous section are presented in table 5. In row 1, we regress IMR for first-borns on the twin status of second-borns.[17] Mothers who have second-born twins have much lower rates of infant mortality *before* the twins than women who had second-born singletons. This suggests that mothers of twins are healthier when considering pre-determined measures of health. Similarly, the rates of infant mortality among first- and second-borns are much lower in families of women who have third-born twins than in those who have third-born singletons. This holds for all parity levels examined.

Much of the existing twin literature focuses on the USA, or other developed countries. In table 6, we provide similar regressions for women in the USA based on the full set of NHIS surveys. These results show that twins are not as good as random, even in the context of a country with a more developed healthcare system and social safety nets. Taller mothers, heavier mothers, and mothers who don't smoke prior to conception (a positive health behaviour) are significantly more likely to have twins.

The dependence of twinning on positive maternal health stocks and behaviours is a consistent and quantitatively important phenomenon in all data sets we have examined. We have compiled data and run similar regressions using vital statistics data from the USA, Brazil, Spain, Scotland and Sweden, and additional survey data from Chile and the United Kingdom (see tables A1–A4 for results). In each case, the probability of twinning increases as mothers become more healthy and are less likely to engage in risky health behaviours before and during pregnancy. Along with the results described in tables 4 and 6, these additional sources of data show that mothers who consume alcohol, tobacco or other drugs, who suffer from chronic disease, stress during the second or third trimester of pregnancy, or who have less access to prenatal care are significantly less likely to give birth to twins.

Finally, if twinning is related to positive health stocks and behaviours of prospective mothers and families, we can examine how rates of twinning respond to time-series variations in (female) health outcomes. While only suggestive, as many other environmental variables may explain changes in twinning, time-series evidence from the USA leads to similar conclusions. Figure 5 plots the rate of twinning from vital statistics data since birth type (single or multiple) was first recorded. Interestingly, the rate of twins has increased steadily over time, even before the advent of IVF and other fertilisation treatments. This is in line with increasing trends in female health over this period, which is proxied by female life expectancy and plotted in the same figure.

[17] IMR is defined as 1 for children who die before their first birthday. We remove from the sample any children who were not yet 1 at the time of the following birth, as these children have not yet been entirely exposed to the risk of infant mortality.

5.2 Selection into twinning: mechanisms

In a wide variety of contexts, healthier women are more likely to give birth to twins. There are a number of competing hypotheses which may explain why this is the case. Firstly, it may simply be the case that healthier mothers are more likely to conceive twins. This may reflect some underlying biological process, such as that mediated by follicle stimulating hormone as discussed in Hall (2003). Secondly, conditional on conceiving twins, healthier mothers may be more likely to take both fetuses to term. Finally, it may be the case that (conditional on conceiving twins and taking them to term), healthier mothers may be more likely to survive the birth, and hence appear in survey or vital statistics data. In broad terms we will refer to these as the conception mechanism, the gestation mechanism and the birth (survival) mechanism.

When considering IV estimates with twins, any of these processes is sufficient to invalidate causal inference insofar as observing twins depends upon hard-to-measure maternal behaviours and characteristics. Nonetheless, we may be interested in determing which of these are the relevant channels in explaining the results from the previous section. Particularly, the mechanism may be relevant when considering the use of the instrument. For example, if twins are less likely *only* due to selective maternal death, then as mothers become more likely to survive childbirth (ie moving from high maternal mortality countries to low maternal mortality countries), threats to instrumental validity become less relevant.

We test these mechanisms below. In order to determine whether twin selection could be entirely explained by selective maternal survival, we follow Alderman et al. (2011) in simulating estimates under the counterfactual scenario that unhealthy women—who are more likely to die in childbirth—were all carrying twins. Using DHS data described in section 4.1, we observe a woman's height, BMI, pregnancy outcomes, and the maternal mortality status of all her sisters. As we do not observe health stocks of women who died in childbirth, we assume that her sister's health (height and BMI) is a reasonable proxy for the health of the woman who died within 42 days of giving birth (which is classed as a maternal death). Appendix figure A2 shows that maternal mortality is much higher among more unhealthy women. Women shorter than the mean height of 155.5 cm are considerably more likely to suffer maternal death, with this being particularly so below heights of 145cm.

To test the potential importance of maternal survival in explaining twin selection, we simulate observations for the number of women who, according to DHS data, would exist in the sample if it were not for the fact that they died in childbirth. We then examine the coefficients of interest in our twin regression (5), if all unhealthy women who died were pregnant with twins, while all healthy women who died were not. As this relies on a binary 'healthy vs unhealthy' distinction, we define this in various ways, based on height and BMI. These results are presented in table 7. The first column shows the estimated coefficients on height and BMI in the unaltered sample of women from DHS countries where maternal mortality data is available. In this sample, a BMI increase of 1 point is associated with a 0.046% increase in the probability of twinning. The remaining columns add in observations based on maternal mortality rates among sisters of surveyed women. For example, in the second column, we examine the effect of adding to the sample unhealthy and healthy women based on the maternal mortality rate in each group, and then assuming that all unhealthy women would give birth to twins, and all healthy mothers would not. As expected, this reduces the importance of positive maternal health in predicting twinning, with the coefficient on BMI falling from 0.0460 to 0.0437. The other columns continue in this manner, however using continually less conservative assumptions in assigning members to the unhealthy group who are defined as giving birth to twins. Even in the final column, where the entire bottom half of the anthropometric distribution is assumed as being unhealthy, the coefficient on both height and BMI remains positive and significant.[18]

These results suggest that selective maternal death alone is not enough to explain why healthier mothers are more likely to have twins. Turning to the gestation mechanism, we are able to test whether less healthy women who are pregnant with twins are more likely to miscarry than healthier women who are also pregnant with twins. In one DHS survey (Nepal), data on miscarriages as well as the type of miscarriage (single or multiple fetuses), is recorded. We thus run a series of regressions where miscarriage is the dependent variable, and the independent variables are measures of poor maternal health, whether the pregnancy is single or multiple, and interactions between pregnancy type and poor maternal health. We would expect that both poor maternal health and a non-singleton pregnancy increase the likelihood of miscarriage, however we are interested in determing if more unhealthy mothers are *more* likely to miscarry twins than healthy mothers carrying twins. Thus, we are interested in testing if the coefficients on the interaction terms are significantly larger than zero.

These regression results are reported in table **??**. As expected, columns (1) and (2) suggest that more unhealthy and less educated women are more likely to report ever miscarrying.[19] Maternal health stocks are proxied by height and BMI, where (negative) outcomes for these

[18]Examining selection in this way (as per Alderman et al. (2011)) is only one way to examine the effect of selection on estimated coefficients. An alternative measure as proposed by Lee (2009) involves trimming the control and treatment group (in our case unhealthy and healthy mothers), to account for differential selection by treatment status. This results in bounds estimates of the effect of treatment (good health) on the outcome variable (twinning). We report Lee bounds in appendix table A6, however note that these bounds are based on the assumption that treatment is random, which here it is not. Nonetheless, Lee (2009) bounds agree with the simulated estimates in table 7, providing further evidence that selective maternal survival is not enough to explain the correlation between maternal health and twinning.

[19]These results hold conditional and unconditional on total fertility.

variables such as underweight and very underweight are based on ICD-10 definitions. Turning to the interaction terms, although standard errors are reasonably large due to the low frequency of twinning, women who are unhealthy (as proxied by a very low BMI), and with no education are significantly more likely to miscarry with twins. Using a richer set of variables from administrative data in the USA and Spain, similar regressions are run. The results in appendix tables 8 and A7 suggest that, depending on the context, less educated women and women who report consuming alcohol during pregnancy are more likely to miscarry twins than mothers with higher education and who do not consume alcohol. These results provide evidence in favour of the gestation mechanism, as conditional on conceiving twins, it is shown that unhealthy women are more likely to miscarry these fetuses before birth.

5.3 The twin instrument and the Q-Q trade-off

Results from section 5.1 show that twinning is not as good as random, even when conditioning on race, maternal age, and even parental education. Healthier mothers are more likely to have twins. If these mothers also invest more after birth, the Q-Q trade-off will be under-estimated (we can see this by considering the term $twin_j u_{ij}^H$ in equation (7). However, progressively including additional health controls in our first and second stage equations should drive IV estimates in the direction of the true Q-Q trade-off from below, as we partially correct for the bias due to the $twin_j u_{ij}^H$ term.

Conversely, OLS estimates are typically thought to over-estimate the true magnitude of the Q-Q trade-off. If unobserved parental behaviours favour both lower fertility and higher child human capital,[20] OLS estimates of β_1 will be negatively biased, and hence *more* negative than the true trade-off. As items are removed from the error term and included in the principal equation to be estimated by OLS, we thus expect that these estimates should approach the true parameter from above.

We examine this intuition by estimating β_1 from (1) by OLS and IV. Our principal data is the large DHS sample, where quality is measured by school Z-score, a child's educational attainment compared to his or her country birth cohort. After considering DHS estimates we turn to NHIS data from USA.

Table 9 reports pooled OLS estimates from all DHS data, and in low and middle-income

[20]As a (highly stylised) example, consider a prospective mother's eventual labour market plans. A mother who plans to join the labour market may prefer fewer children, facilitating more immediate labour force participation, but have more resources to invest in child quality.

country groups. OLS is estimated separately by the 2+, 3+, and 4+ fertility groups, and included in appendix table A8. As is typically found in empirical studies of the Q-Q tradeoff, conditional correlations between family size and child outcome variables are negative, and strongly significant. The results in table 9 suggest that an additional sibling is associated with an approximatelty 0.1 s.d. decrease in standardised schooling outcomes. The magnitude of these estimates decreases as additional controls for maternal health and family socioeconomic variables are included in the regressions. This is in line with the hypothesised effect of these variables. To the degree that components from the error term are removed which are positively related to high desired quality and to low desired family size, the bias in OLS estimates should be reduced, and the estimated magnitude of the trade-off should move in the direction of zero. Of course, as long as any such variables remain unobserved and as part of the stochastic error term, OLS estimates *will not* converge to unbiased values. The following sections examine alternative estimation methods to bound the Q–Q trade-off.

5.3.1 The twin instrument: estimates of the Q–Q trade-off

In table 10, we turn to IV estimates using twins. As we outline in section 3.1, 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 β_1 using twin births.

However, it is likely that the β_1 which 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 (8). 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, thus resulting in larger family sizes. At the same time, we expect healthier mothers to invest more in their children after birth, and hence have 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.

The main specification is displayed in the top row of table 10, with separate columns for the 2+, 3+ and 4+ sample groups. For each parity group, 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 from the twin Q–Q literature, 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 (or close to statistically significant in the case of the 2+ sample) Q–Q trade-off of at least 3 or 4%.

Thus, in low- and middle-income country data, the inclusion of health indicators in the twin instrument does have an important effect on IV estimates, moving as hypothesised in (9). In table 11, we present identical estimates based upon NHIS data from the USA. As this survey focuses on health, as well as standardised educational attainment, we examine the effect of additional siblings on the reported health of children. Results for both variables show similar to those reported based on DHS data. Focusing on the 3+ group, the inclusion of health and socioeconomic controls results in OLS estimates moving closer to zero, and IV results further away from zero. In the case of self reported health status, the inclusion of additional twin predictors is sufficient to result in statistically significant evidence in favour of the Q–Q trade-off, despite the reasonably imprecise standard errors of estimated coefficients. While for the 2+ subgroup (firstborns), the effect of a second-born twin does not result in statistically significant results, it is important to note that the point estimates become more negative, moving in favour of a negative β_1, consistent with all other NHIS and DHS results.

5.3.2 Heterogeneity

Theoretical derivations of the Q–Q model are based on the assumption that all children in a family are of the same quality. More recent work (for example the theoretical work of Aizer and Cunha (2012)) has loosened this assumption. Among other things, this allows for reinforcing behaviours by parents in child human capital investment decisions.[21] If this is the case, the coefficient β_1 may vary by children in the family. More generally, β_1 may be context specific, depending upon the returns to human capital in a given time-period or economy.

Empirically, we find that estimates of the Q-Q trade-off are heterogeneous across birth orders, country income level, and the gender of the child affected by the additional birth. 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

[21] An empirical review of early life human capital and reinforcing versus compensating behaviour, (however not explicitly related to the Q–Q hypothesis), is provided by Almond and Mazumder (2013).

three to four births or four to five births. However, in lower fertility environments the effect is, as expected, concentrated on lower birth orders. The third row of table 10 suggests that in middle-income countries 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.

Similarly, effects of the Q-Q trade-off vary considerably depending upon a child's gender. In appendix table A10 we present regression results 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 male children. Interestingly, recent empirical of the Q–Q trade-off from other (middle income) contexts finds similar gender-biased results (Ponczek and Souza, 2012) disfavouring girl children.

5.4 Bounding the Q-Q trade-off

The results from the previous subsection provide consistent estimates of β_1 via 2SLS *if* the full set of controls in the first and second stage equations 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 these datasets do not exhaustively capture all aspects of health stocks and behaviours, 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 outlined in section 3.2, this involves the definition of some prior belief over the sign and magnitude that the coefficient on twinning would take in the structural equation 10. Results are displayed in figures 6 and 7. At each point on the horizontal axis of these figures, the bounds for β_1 are displayed, along with the corresponding point estimate under the assumption that γ is distributed $\sim U(0, \delta)$. Dashed lines present the 95% confidence interval, while the solid line represents the point estimate.

While this technique allows us to agnostically estimate bounds over a range of values for γ, we have thus far made no restriction over the true magnitude of δ. While in figures 6 and 7 we have assumed that this is less than 0.2 standard deviations, we can can form far more precise bounds by estimating γ directly. As discussed more extensively in appendix B, this requires us to (causally) estimate the effect of maternal health shocks on twinning, and the effect of maternal health shocks on child quality. By taking the ratio of these, we can calculate the partial correlation between twinning (which occurs with higher frequency for healthier mothers), and child quality, and then we can plug our estimate of γ (where a non-zero value for γ implies instrumental invalidity) in Conley et al.'s bound estimator.

We follow the specification outlined in Bhalotra and Venkataramani (2014) (and appendix B of this paper) to estimate the effect of the positive health shock associated with Sulfanide drugs on $\Pr(twin)$ and child quality (school Z-score). These estimates are presented in appendix table A11. We take the ratio of these estimates to isolate the effect of twinning on child quality. These are displayed at the end of table A11. For what remains of this section, we use the estimate from column 1, where $\hat{\gamma} = 0.00306/0.0535 = 0.0571$.

In table 12 we provide estimates of Conley bounds using our more precisely estimated γ. These are our preferred bounds estimates for β_1. As described in appendix B, for the UCI approach this implies an assumption that $\gamma \in [0, 2\hat{\gamma}]$ (or $\gamma \in [0, 2 \times 0.0571]$). For the LTZ approach, we assume that $\gamma \sim N(\mu_{\hat{\gamma}}, \sigma_{\hat{\gamma}})$. The assumption of normality is driven by resampling (bootstrap) estimates of $\hat{\gamma}$, which allows us to construct a distribution for γ. This is then tested for equality against a normal distribution, and the null hypothesis of different distributions is not rejected with a p-value of 0.1805. The empirical and analytical distributions for γ that are applied in the LTZ method are displayed in appendix figure A3b-A3a.

These bounds estimates are informative for the three-plus and four-plus groups. In each case, both techniques (UCI and LTZ), have their bounds (at a 90% significance level), entirely below zero. This is the case, despite the fact that confidence intervals are quite wide. Although the bounds themselves are wide, the point estimates at the centre of the bounds in all cases fall between the downward-biased OLS estimates and the upwards-biased IV estimates which we have discussed earlier in this paper. For each of the three parity groups examined, the LTZ approach suggests that the Q-Q trade-off is approximately -0.08 of a standard deviation when considering children's educational attainment. While for the 2+ sample this is not significantly different to zero, for all other samples these estimates are significant, with a p-value< 0.1.

6 Conclusion

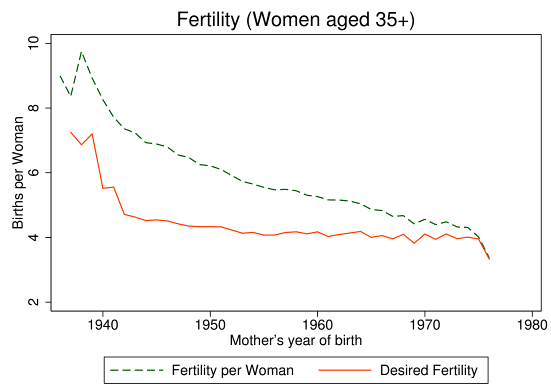
Twin births are not random. Rather, they appear to be far from it, in a wide variety of environments, time periods and contexts. Based on a considerable body of evidence compiled from vital statistics and survey data from low- and high-income countries, we demonstrate that mothers with greater health stocks, those who engage in positive health-related behaviours, and those living in healther environments are much more likely to take twins to term. It is demonstrated that these mothers are healthier *prior* to twinning, and this results in a greater likelihood of taking twins to term, conditional upon conceiving two fetuses.

These results have important implications for empirical work which aims to identify the causal effect of child quantity (more siblings) on child quality (higher human capial). The existing evidence from the Q–Q literature is mixed. A range of studies which estimate the effect of quantity on quality find that the effect is small, or frequently not statistically different from zero. By assembling large datasets linking a child's human capital outcomes with her mother's health—in both the developing world and the USA—we show that partially correcting for twin endogeneity is sufficient to push estimates of the trade-off up by about 3%-4% of a standard deviation, potentially explaining the lack of significant results in the existing literature. Using partial identification to bound the effect of child quantity on child quality suggests that the *true* effect size, once accounting for the entire health differential in favour of twin families, may even be as high as 8% of a standard deviation.

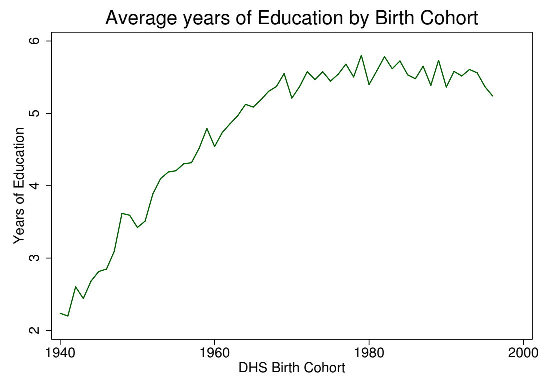
We are able to conclude that additional unexpected births do have quantitatively important effects on their siblings' educational outcomes. An 8% of a standard deviation increase is equivalent to an additional 0.3 years in the classroom. While the true effect of these 0.3 years of course depend on the quality of education imparted, time in school is at the very least a necessary condition for more general increases in learning outcomes. The implications of these findings are wide-reaching, both in terms of the vindication of Beckerian theory, and particularly, in an applied sense, when considering human well-being in developing countries which are yet to fully pass through the demographic transition.

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

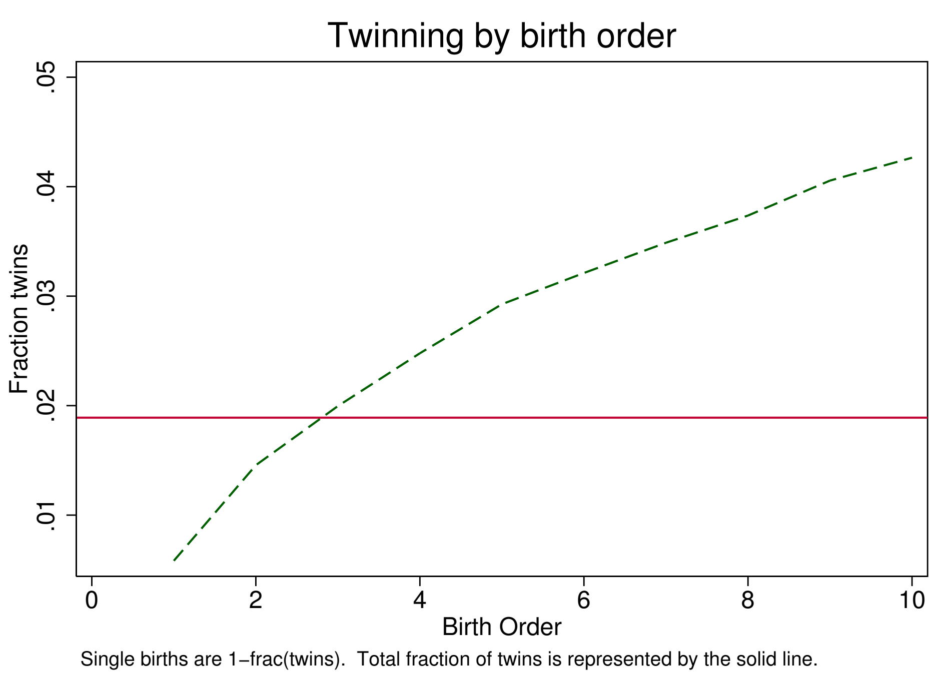


Figure 3: Twin Births and Total Fertility

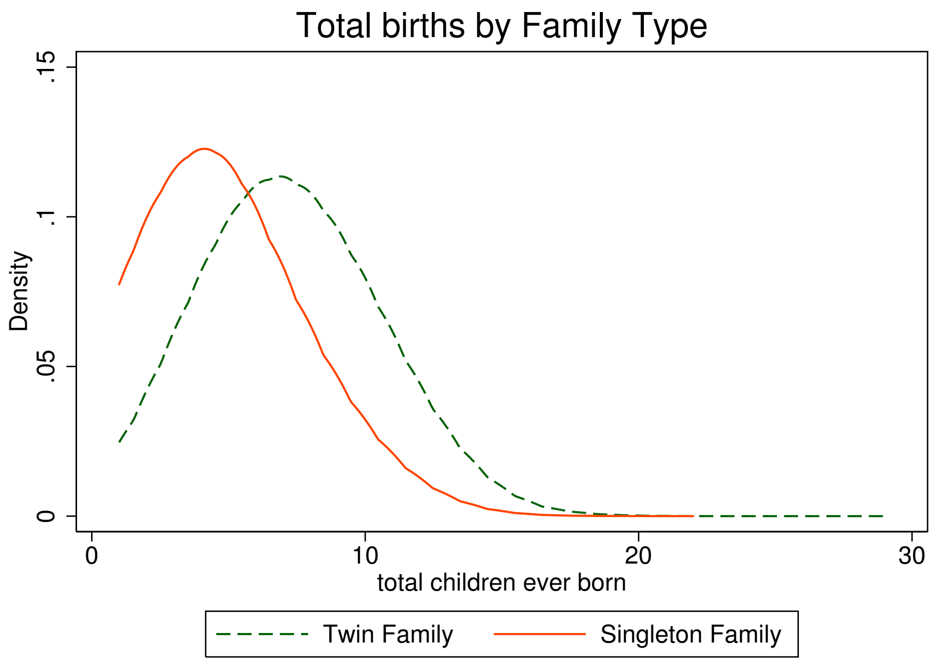


Figure 4: Intra- and Inter-country trends: height and twinning

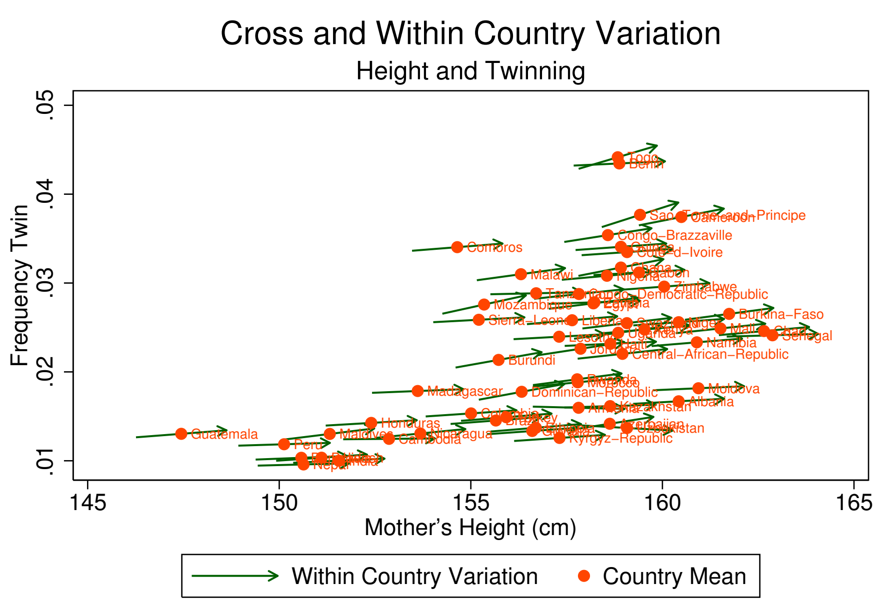
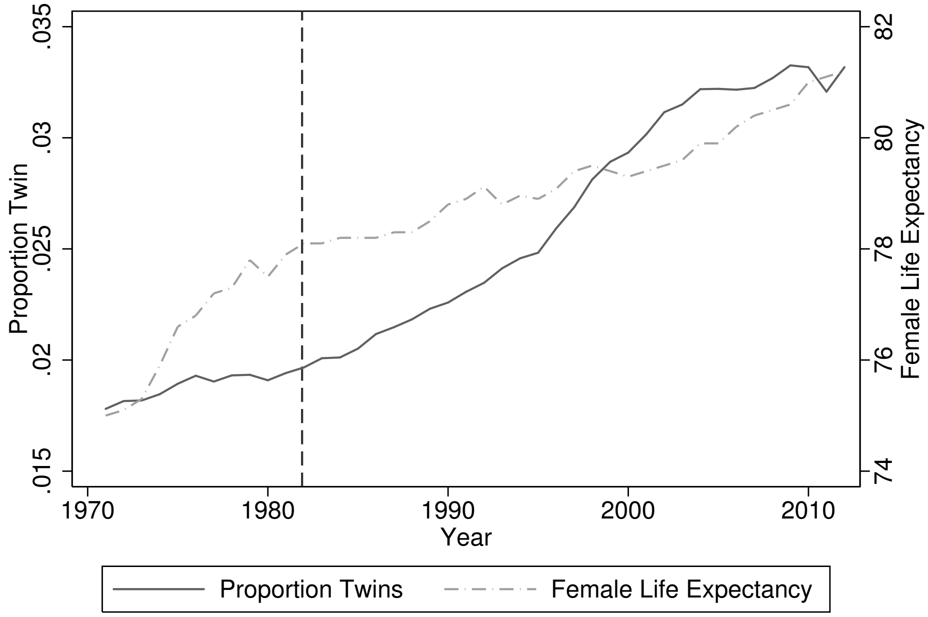
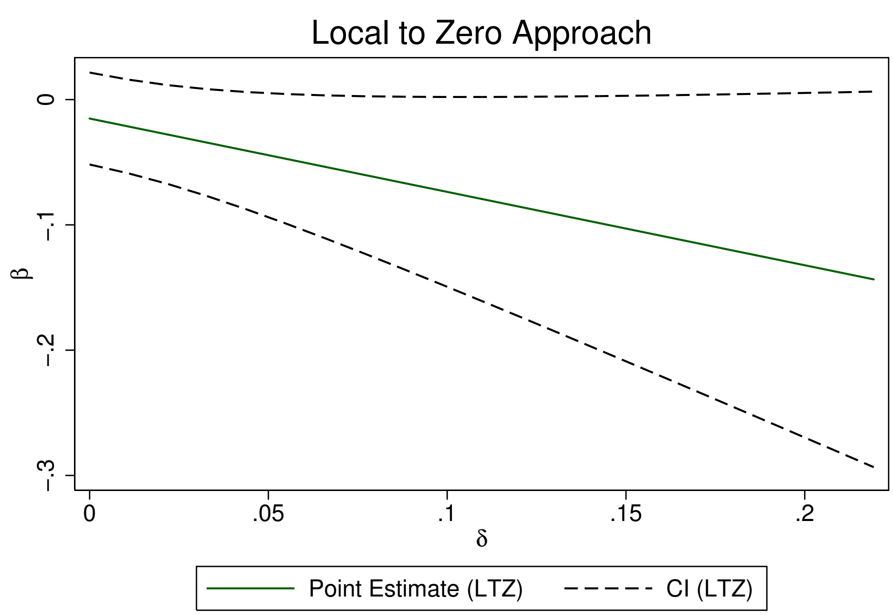


Figure 5: Proportion of Twins of All Births (USA)



Twin data from NVSS Birth Certificate Data. Female Life Expectancy data from World Bank.
Dotted line represents first ever IVF birth in USA.

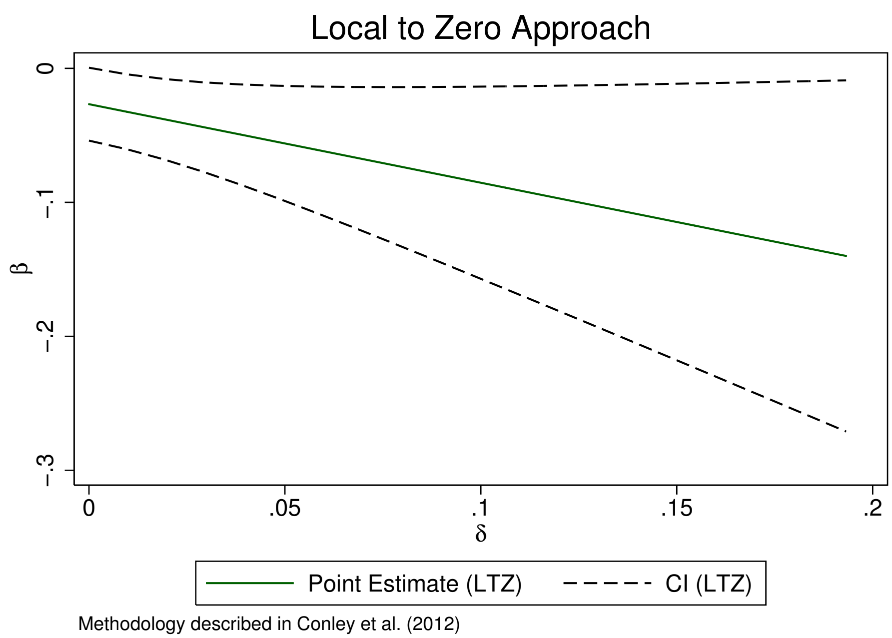
Figure 6: Relaxing Strict Exogeneity (two plus)



Methodology described in Conley et al. (2012)

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 12.

Tables

Begin overleaf.

Table 1: Fertility and the Twin Instrument: Literature

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

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

Notes: Individual sources discussed further in the body of the text. Estimates reported in each study are presented along with their standard errors in parenthesis. Parentheses marked as † contain the t-statistic rather than the standard error.

Table 2: Summary Statistics

	Low Income		Middle Income		
	Single	Twins	Single	Twins	All
Fertility					
Fertility	3.670	6.092	3.348	5.426	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)	(2.830)	(2.083)	(2.497)	(2.403)
Fraction Twin	0.0194		0.0174		0.0185
	(0.1378)		(0.1306)		(0.1348)
Birth Order Twin	4.680		4.017		4.420
	(2.463)		(2.370)		(2.448)
Mother's Characteristics					
Age	30.91	34.19	31.99	34.94	31.40
	(7.980)	(7.163)	(8.105)	(7.158)	(8.038)
Education	3.983	3.340	6.817	6.119	5.030
	(4.369)	(4.033)	(4.794)	(5.047)	(4.735)
Height	155.6	157.7	155.6	157.2	155.7
	(7.084)	(6.988)	(6.957)	(6.956)	(7.042)
BMI	21.89	22.46	25.83	26.50	23.38
	(3.983)	(4.097)	(5.066)	(5.436)	(4.822)
Pr(BMI)<18.5	0.172	0.123	0.0344	0.0277	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	3.211	5.438	4.999	4.465
	(3.581)	(3.269)	(3.859)	(3.733)	(3.805)
Education (Z-Score)	-0.00868	-0.0137	0.0121	-0.0356	0.000178
	(1.001)	(0.960)	(0.998)	(0.986)	(1.000)
No Education (Percent)	0.200	0.212	0.0634	0.0765	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	492,057	7,525	297,184	4,387	850,036
Number of Children (Education)	1,176,563	24,992	714,764	14,334	1,930,653
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 850,036 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 B1.

Table 3: Summary Statistics (NHIS)

	Single	Twins	All
FERTILITY			
Fertility	1.925	3.107	1.955
	(1.001)	(1.176)	(1.022)
Fraction Twin		0.0251	
		(0.0156)	
Birth Order Twin		2.207	
		(1.063)	
MOTHER'S CHARACTERISTICS			
Age	36.05	36.88	36.07
	(8.396)	(7.997)	(8.387)
Education	12.54	12.70	12.54
	(2.326)	(2.232)	(2.323)
BMI	27.45	28.01	27.47
	(6.628)	(7.247)	(6.645)
Pr(BMI)<18.5	0.0206	0.0168	0.0205
	(0.142)	(0.128)	(0.142)
Excellent Health	0.320	0.325	0.320
	(0.466)	(0.468)	(0.466)
CHILDREN'S OUTCOMES			
Education (Years)	5.135	4.633	5.123
	(3.835)	(3.736)	(3.833)
Education (Z-Score)	0.00236	-0.0980	0.0000
	(1.001)	(0.949)	(1.000)
Pr(Excellent Health)	0.526	0.536	0.526
	(0.499)	(0.499)	(0.499)
Number of Mothers	121,912	5,097	127,009
Number of Children	237,152	9,494	246,646

NOTES: Summary statistics are presented for the full estimation sample consisting of all children 18 years of age and under included in NHIS surveys from 2000-2013. 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 month-of-birth cohort (mean 0, std deviation 1). Excellent health is self reported by the mother of the child, and BMI is weight in kilograms over height in metres squared.

Table 4: Probability of Giving Birth to Twins

Twin*100	(1) All	(2) Income	(3)	(4) Time	(5)	(6) Prenatal
		Low inc	Middle inc	1990-2013	1972-1989	
Age	0.596***	0.615***	0.554***	0.647***	0.326***	0.631***
	(0.029)	(0.036)	(0.050)	(0.033)	(0.075)	(0.040)
Age Squared	-0.008***	-0.008***	-0.008***	-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***	0.004	-0.052***	-0.056***	-0.041***
	(0.009)	(0.012)	(0.014)	(0.010)	(0.019)	(0.013)
Education (years)	0.042**	0.089***	-0.005	0.048**	0.022	-0.068**
	(0.017)	(0.022)	(0.029)	(0.020)	(0.034)	(0.028)
Education squared	-0.002	-0.006***	0.001	-0.002	0.000	0.003
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
Height	0.058***	0.057***	0.059***	0.062***	0.042***	0.058***
	(0.004)	(0.005)	(0.007)	(0.005)	(0.008)	(0.007)
BMI	0.048***	0.063***	0.039***	0.046***	0.055***	0.044***
	(0.006)	(0.009)	(0.009)	(0.007)	(0.011)	(0.011)
Prenatal (Doctor)						0.906***
						(0.128)
Prenatal (Nurse)						0.067
						(0.108)
Prenatal (None)						-0.497***
						(0.132)
R-squared	0.01	0.01	0.01	0.01	0.01	0.01
Observations	1930653	1201555	729098	1524947	405706	615935

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 10. 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 10.*p<0.1; **p<0.05; ***p<0.01

Table 6: Probability of Giving Birth to Twins USA (NHIS)

Twin×100	All	Time	
		1982-1989	1990-2013
Age	0.0198	-0.569**	0.0306
	(0.0432)	(0.225)	(0.0462)
Age Squared	-0.000903	0.00513**	-0.000883
	(0.000601)	(0.00248)	(0.000649)
Age First Birth	0.153***	0.200*	0.143***
	(0.0131)	(0.106)	(0.0139)
Education (years)	0.0157	0.0793*	0.0107
	(0.0154)	(0.0453)	(0.0163)
Height	0.0341*	-0.0163	0.0386*
	(0.0201)	(0.0597)	(0.0213)
BMI	0.00852***	0.0158*	0.00770**
	(0.00304)	(0.00849)	(0.00324)
Smokes (pre-birth)	-0.186*	0.171	-0.206*
	(0.112)	(0.312)	(0.119)
Observations	114,037	10,114	103,923
R^2	0.003	0.006	0.003

All specifications include a full set of survey year, region of birth, and mother's race 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. Standard errors clustered by mother are included in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Can Selective Maternal Survival Explain Twinning Rates?

Twin×100	MMR Sample	<140cm or BMI <16	<145cm or BMI <16.5	<150cm or BMI <17	<155cm or BMI <17.5
Height	0.0657***	0.0635***	0.0590***	0.0514***	0.0417***
	(0.00414)	(0.00414)	(0.00418)	(0.00420)	(0.00425)
BMI	0.0460***	0.0437***	0.0427***	0.0409***	0.0405***
	(0.00637)	(0.00636)	(0.00637)	(0.00643)	(0.00650)
Observations	844,638	848,642	848,686	848,557	848,667
R^2	0.024	0.024	0.024	0.023	0.022

Each column represents a separate regression of maternal characteristics on twinning. For a full list of variables included see table 4. Only health variable are included in regression output. Column 1 includes the full sample of women surveyed in countries where the DHS maternal mortality module is applied. Columns 2-5 inflate samples in line with maternal mortality rates, where ‘unhealthy’, is defined as described in the column title. Full details are available in the body of the text. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Are Twins More Likely to Miscarry to Unhealthy Mothers? (USA Vital Stats)

Miscarriage× 100	All (1)	Alcohol (2)	Health (3)
African American	2.545***	6.205***	4.692***
	(0.0253)	(0.0469)	(0.0430)
Primary Education	1.413***	0.173**	1.000***
	(0.0359)	(0.0794)	(0.0726)
Secondary Education	0.719***	1.220***	1.042***
	(0.0227)	(0.0398)	(0.0364)
Consumed tobacco (pre-birth)	1.426***	3.702***	3.040***
	(0.0353)	(0.0620)	(0.0567)
Twin	8.334***	11.02***	15.07***
	(0.677)	(1.159)	(1.074)
Twin × Tobacco	1.176***	1.362***	1.590***
	(0.209)	(0.365)	(0.334)
Twin × No Education	4.578***	5.144**	3.865*
	(1.703)	(2.363)	(2.160)
Twin × Primary Education	-2.593***	2.108***	0.728
	(0.220)	(0.527)	(0.482)
Twin × Secondary	-0.681***	1.071***	0.858***
	(0.122)	(0.216)	(0.198)
Twin × African American	-0.0244	-0.402	0.961***
	(0.134)	(0.248)	(0.226)
Twin × Alcohol		4.630***	5.435***
		(1.339)	(1.224)
Pregnancy related hypertension			0.179**
			(0.0716)
Eclampsia			6.885***
			(0.0728)
Twin × Hypertension			3.690***
			(0.279)
Twin × Eclampsia			-5.543***
			(0.289)
Consumed alcohol (pre-birth)		3.788***	4.046***
		(0.215)	(0.196)
Constant	87.97***	78.70***	53.14***
	(0.103)	(0.171)	(0.162)
Observations	3,955,099	1,753,396	1,753,396
R^2	0.464	0.519	0.598

Data is a 10% sample of NVSS birth and fetal death data from 2003–2012. Each regression includes month and year of birth fixed effects, and mother age fixed effects (interacted with a binary variable for twins) and is estimated by OLS. Miscarriage is a binary variable, and is multiplied by 100 for presentation. The omitted education variable is tertiary. Standard errors are included in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 9: OLS Estimates of the Q-Q Trade-off

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

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. 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 10: 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

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). 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. Full first stage results for each row are available in table A9. Standard errors are clustered by mother.*p<0.1; **p<0.05; ***p<0.01

Table 11: NHIS Estimates (USA): Education and Health

	2+			3+			4+		
	Base	+H	+S&H	Base	+H	+S&H	Base	+H	+S&H
OLS									
School Z-Score	-0.043***	-0.033***	-0.027***	-0.036***	-0.028***	-0.020**	-0.010	-0.004	0.002
	(0.005)	(0.006)	(0.006)	(0.008)	(0.009)	(0.009)	(0.018)	(0.018)	(0.018)-
Excellent Health	-0.012***	-0.005***	-0.004*	-0.018***	-0.010***	-0.008***	-0.028***	-0.019***	-0.017***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)	(0.006)	(0.005)	(0.005)
IV									
School Z-Score	-0.064	-0.074	-0.077	-0.027	-0.037	-0.036	-0.094	-0.099	-0.113
	(0.063)	(0.059)	(0.058)	(0.065)	(0.064)	(0.065)	(0.152)	(0.155)	(0.150)
Excellent Health	0.013	0.022	0.019	-0.016	-0.054*	-0.055*	0.057	0.017	0.006
	(0.025)	(0.020)	(0.020)	(0.038)	(0.031)	(0.031)	(0.058)	(0.054)	(0.052)
Descriptives									
School Z-Score	-0.0233			-0.0363			-0.0628		
	(0.993)			(1.041)			(1.086)		
Excellent Health	0.533			0.511			0.490		
	(0.499)			(0.500)			(0.500)		
Observations	75,902	75,902	75,902	57,413	57,413	57,413	26,128	26,128	26,128
Joint F-test Educ		164.5	64.7		101.3	39.6		38.0	7.7
Joint F-test Health		34469.6	163.9		15335.6	28.4		5276.4	17.1

NOTES: Each cell presents the coefficient of interest from a regression using NHIS survey data (2002-2014). Base controls include child age FE (in months), mother's age, and mother's age at first birth plus race dummies for child and mother. First stage omitted for clarity (generally all around 0.7). Standard errors are clustered by mother.*p<0.1; **p<0.05; ***p<0.01.

Table 12: 'Plausibly Exogenous' Bounds

	UCI: $\gamma \in [0, \delta]$		LTZ: $\gamma \sim N(\mu_\delta, \sigma_\delta)$	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Two Plus	-0.1560	0.0244	-0.1855	0.0139
Three Plus	-0.1605	-0.0013	-0.1978	-0.0024
Four Plus	-0.1581	-0.0048	-0.1962	-0.0067

NOTES: This table presents upper and lower bounds of a 90% 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 $\gamma \sim N(\mu_\delta, \sigma_\delta)$. In each case δ (or μ_δ and σ_δ) is estimated by the method discussed in appendix B and section 3.2.

References

A. Aizer and F. Cunha. The Production of Human Capital: Endowments, Investments and Fertility. NBER Working Papers 18429, National Bureau of Economic Research, Inc, Sept. 2012.

H. Alderman, M. Lokshin, and S. Radyakin. Tall claims: Mortality selection and the height of children. Policy Research Working Paper 5846, The World Bank, Oct 2011.

D. Almond and B. Mazumder. Fetal Origins and Parental Responses. *Annual Review of Economics*, 5(1):37–56, 05 2013.

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.

D. Almond, B. Mazumder, and R. van Ewijk. Fasting During Pregnancy and Children's Academic Performance. NBER Working Papers 17713, National Bureau of Economic Research, Inc, Dec. 2011.

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.

D. J. P. Barker. Fetal origins of coronary heart disease. *BMJ*, 311(6998):171–74, 1995.

G. S. Becker. An Economic Analysis of Fertility. In *Demographic and Economic Change in Developed Countries*, NBER Chapters, pages 209–240. National Bureau of Economic Research, Inc, June 1960.

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. Bhalotra and S. Rawlings. Gradients of the Intergenerational Transmission of Health in Developing Countries. *The Review of Economics and Statistics*, 95(2):660–672, May 2013.

S. Bhalotra and A. Venkataramani. Shadows of the Captain of the Men of Death: Early Life Health Interventions, Human Capital Investments, and Institutions. Mimeo, University of Essex, 2014.

S. Black, P. J. Devereux, and K. Salvanes. Does grief transfer across generations? In-utero deaths and child outcomes. NBER Working Papers 19979, National Bureau of Economic Research, Inc, Mar. 2014.

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.

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. *Journal of Population Economics*, 27(1):33–68, Jan 2014.

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

E. A. Hanushek. The trade-off between child quantity and quality. *Journal of Political Economy*, 100(1):84–117, February 1992.

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.

S. Jayachandran, A. Lleras-Muney, and K. V. Smith. Modern Medicine and the 20th-Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs. *American Economic Journal: Applied Economics*, 2(2):118–46, 2010.

D. S. Lee. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, 76(3):1071–1102, 07 2009.

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.

B. Mazumder and Z. Seeskin. Skipping breakfast and the sex ratio at birth. Mimeo, 2014.

V. Ponczek and A. P. Souza. New evidence of the causal effect of family size on child quality in a developing country. *Journal of Human Resources*, 47(1):64–106, 2012.

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.

C. Quintana-Domeque and P. Ródenas-Serrano. Terrorism and Human Capital at Birth: Bomb Casualties and Birth Outcomes in Spain. IZA Discussion Papers 8671, Institute for the Study of Labor (IZA), Nov. 2014.

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.

United Nations. World population policies 2009. Technical report, Department of Economic and Social Affairs: Population Division, 2010.

Appendices

A Data Appendix

Main IV and OLS results for this paper are based on DHS and NHIS data described in section 4. These data are downloaded directly off the web and merged to form the estimation samples of interest. For DHS data, we use two surveys: the Individual (woman) Recode (IR), and the Household Recode (HR) providing education for each household member. For NHIS data, we merge three of the datafiles made available by the CDC: familyxx, household, and person. In each case, full generating code for this process is made available on the authors' websites. This code downloads, merges and cleans DHS and NHIS data to produce the datasets (one line per child) used in analysis.

In auxiliary regressions examining the characterstics of mothers and the relationship these characteristics and twin births and miscarriage, we consult a large number of other datasets. These are the following:

- United States National Vital Statistics Birth Data
- United States National Vital Statistics Fetal Death Data
- Spanish Vital Statistics (INE)
- The Swedish Medical Birth Register
- Scottish Vital Statistics
- Longitudinal Early Life Survey, Chile (ELPI)

In the case of the first 5 datasets (administrative records of births and/or fetal deaths), we use all recorded instances, focusing on twins as our outcome variable of interest. Depending upon the data source, we use all available measures of pre-determined maternal health stocks or family socioeconomic indicators. The ELPI survey from Chile focuses on child early life, and records mother's behaviours before, during and after pregnancy, along with child birth outcomes. We use all children from the first wave of this survey to run the twin regression included in the appendix tables. Further notes regarding each dataset and the particular years and number of births can be found in the notes to each table.

B Plausibly Exogenous Bounds and Estimating γ

From (10), we are interested in forming a consistent estimate of $\gamma = \frac{\partial educ}{\partial twin}|_X$. From Bhalotra and Venkataramani (2014) we have:

$$Y_{stc} = \alpha + \phi(Post_t \times basePneumonia_s) + \theta_{rs} + \eta_{rt} + \varphi \mathbf{X}_{st} + \lambda_{rc} + \theta_s \times \eta_t + \varepsilon_{stc} \tag{13}$$

where ϕ is the effect of access to sulfanide drugs on the outcome variable Y_{stc}.[22] Let's consider estimating the effect of this exogenous shock to maternal health on two outcomes: a mother's twinning rates and her child quality:

$$Pr(Twin)_{stc} = \alpha^t + \phi^t(Post_t \times basePneumonia_s) + \cdots + \varepsilon^t_{stc} \tag{14a}$$

$$educ_{stc} = \alpha^q + \phi^q(Post_t \times basePneumonia_s) + \cdots + \varepsilon^q_{stc} \tag{14b}$$

where superscript t refers to coefficients in the twin regression while q refers to quality, and the remainder of (14a) and (14b) are identical to (13). We can thus causally estimate $\phi^t = \frac{\partial twin}{\partial bP}|_X$ and $\phi^q = \frac{\partial educ}{\partial bP}|_X$ by OLS (where bP is the $Post_t \times basePneumonia_s$ variable) as long as the aforementioned parallel trends assumption holds. Using (14a) and (14b), we can (seperately) estimate the effect of positive maternal health shocks on twinning, and the effect of positive maternal health shocks on child quality. However, for γ—the violation of the exclusion restriction—we need to estimate the twin-mediated effect of health shocks on child quality. Thus, to estimate γ, we can take the ratio of estimates in each of the above two equations:

$$\phi^t/\phi^q = \left.\frac{\partial educ/\partial bP}{\partial twin/\partial bP}\right|_X = \left.\frac{\partial educ}{\partial twin}\right|_X = \gamma. \tag{15}$$

which is our estimand of interest, and which we can then plug in to our estimates of the bounds on β_1 using Conley et al.'s method.

In order to estimate (15) we turn to US census data described in Bhalotra and Venkataramani (2014), firstly estimate ϕ^t and ϕ^q, and then calculate ratio of these values to form an estimate of γ, which we denote $\hat{\gamma}$. In the case of the UCI approach, this is sufficient to estimate the bounds of β_1, assuming that: $\gamma \in [0, 2\hat{\gamma}]$.[23] In the case of the more precise LTZ approach (our preferred bounds estimates), the logic is similar, however now we must form a prior over the entire distribution of γ. Calculating the variance of γ is not as straightforward as using the variance-covariance matrix corresponding to each of the estimates $\hat{\phi}^t$ and $\hat{\phi}^q$. In this case however we can

[22] This is an unbiased effect of sulfanide drugs on health if parallel trends are satisfied between high- and low-intensity states. Evidence of this is provided by Bhalotra and Venkataramani (2014).

[23] We scale γ by the aribtrary factor of 2 in order to be conservative with the maximum upper bound, Conley et al. (2012) provide a similar example to calculate the returns to education using the UCI approach.

use bootstrapping to calculate J replications of $\hat{\phi}^t/\hat{\phi}^q$, and from these estimates construct an estimated distribution of $\hat{\gamma}$ estimates, which allows us to determine our prior for the distribution of γ. From this empirical distribution, we observe the estimated mean and standard deviation, and finally test whether the distribution is normal using Kolmogorov-Smirnov test of equality of distributions.[24]

[24] In order to do this, we first estimate the empirical distribution as described previously. We then observe the mean $\hat{\mu}$ and the standard deviation $\hat{\sigma}$, and run a one-sample test do determine whether the observed empirical distribution is is significantly different to a normal distribution $\sim N(\hat{\mu}, \hat{\sigma}^2)$.

C Appendix Tables

Table A1: Probability of Giving Birth to Twins USA (Vital Statistics)

	Time		Alcohol	Health
Twin×100	1980-1989	2003-2012	Controls	Controls
African American	0.629***	0.585***	0.477***	0.373***
	(0.0228)	(0.0288)	(0.0451)	(0.0455)
Other Race	-0.379***	0.0185	0.140***	-0.578***
	(0.0481)	(0.0280)	(0.0470)	(0.0634)
Secondary Education	0.0608***	-0.0126	0.0190	0.734***
	(0.0193)	(0.0278)	(0.0504)	(0.0645)
Tertiary Education	-0.0470**	1.115***	1.100***	1.963***
	(0.0217)	(0.0294)	(0.0499)	(0.0662)
Consumed tobacco (pre-birth)		-0.334***	-0.445***	-0.440***
		(0.0397)	(0.0576)	(0.0580)
Consumed alcohol (pre-birth)			-1.394***	-1.332***
			(0.204)	(0.205)
Mother Anemic				-1.362***
				(0.106)
Mother Cardiac Disease				-0.550**
				(0.244)
Mother Chronic Lung Disease				-0.639***
				(0.142)
Mother Diabetic				-0.341***
				(0.0886)
Mother Chronic Hypertension				-0.795***
				(0.167)
Pregnancy Associated Hypertension				-4.393***
				(0.0846)
Eclampsia				-5.959***
				(0.280)
Married	-0.173***			
	(0.0215)			
Observations	3,566,621	3,891,882	1,660,669	1,222,212
R^2	0.001	0.008	0.008	0.012

Each regression includes full maternal age and child birth year and birth order fixed effects and is estimated as a linear probability model. The outcome variable (a twin birth) is multiplied by 100 for presentation. Heteroscedasticity robust standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A2: Probability of Giving Birth to Twins (Chile)

	(1)		
Twin×100			
PRE-PREGNANCY		PREGNANCY	
Income p.c.	-0.006	Smoked	-0.573
	(0.011)		(0.416)
Income p.c. squared	0.000	Drugs (infrequent)	-0.119
	(0.000)		(1.646)
Secondary Education	0.142	Drugs (frequent)	-1.872***
	(0.300)		(0.344)
Tertiary Education	1.507***	Alcohol (infrequent)	-0.002
	(0.583)		(0.570)
Low Weight	-0.589	Alcohol (frequent)	-1.891***
	(0.471)		(0.290)
Obese	-1.997***	No Check-ups	-1.031
	(0.766)		(0.966)
Mother's Age	0.410***	Hospital Birth	0.939***
	(0.133)		(0.344)
Mother's Age Squared	-0.007***	Diabetes	-0.255
	(0.002)		(0.505)
Indigenous	-1.027***	Depression	0.031
	(0.395)		(0.416)
Observations	14268	R^2	0.01

NOTES: Data comes from the Encuesta Longitudinal de Primera Infancia (ELPI) from Chile. Education at each level are dummy variables, primary education is the omitted base. Regional controls and child age fixed effects are omitted for clarity. Heteroscedasticity robust standard errors are presented in parenthesis.*p<0.1; **p<0.05; ***p<0.01

Table A3: Probability of Giving Birth to Twins (Scotland)

	(1)		
Twin×100			
PRE-PREGNANCY		PREGNANCY	
Deprivation Index (Quintile 2)	-1.628**	Smoker	0.001
	(0.958)		(0.669)
Deprivation Index (Quintile 3)	-0.188	Previous Smoker	1.717**
	(0.967)		(0.877)
Deprivation Index (Quintile 4)	-0.421	Alcohol (1-2 per week)	-4.498*
	(0.934)		(1.935)
Deprivation Index (Quintile 5)	-1.132	Alcohol (3+ per week)	-3.030*
	(0.920)		(1.543)
Height	0.306***	Overweight	-0.092
	(0.044)		(0.643)
Married	3.272***	Obese	1.350**
	(0.878)		(0.746)
Age	-0.337	Diabetes	-0.188
	(0.400)		(0.967)
Age Squared	0.020***		
	(0.007)		
Observations	193,254	R-squared	0.01

NOTES: Data comes from Scottish birth records. All births occurring after 24 weeks for the period 1997-2012 are included. Twin births account for 1.51% of the sample. *p<0.1; **p<0.05; ***p<0.01

Table A4: Probability of Giving Birth to Twins (Sweden Birth Registry)

Twin×100	(1) All	(2) Time 1991-2010	(3) Time 1983-1990
Mother's Age	0.257***	0.483***	0.277***
	(0.031)	(0.047)	(0.040)
Mother's Age Squared	-0.002***	-0.005***	-0.003***
	(0.001)	(0.001)	(0.001)
Mother's Education (years)	0.014	0.011	-0.013
	(0.009)	(0.014)	(0.012)
Height	0.038***	0.047***	0.019***
	(0.003)	(0.003)	(0.004)
Smoked	0.252***	0.237***	0.208***
	(0.045)	(0.069)	(0.056)
Consumed alcohol	0.160	-2.381***	0.137
	(1.176)	(0.414)	(1.192)
R-squared	0.01	0.01	0.01
Observations	1,628,737	1,098,076	530,661

NOTES: Conditional on birth cohorts and marriage status FEs. Linear probability estimates. Standard errors in parenthesis are clustered at the level of the mother. *p<0.1; **p<0.05; ***p<0.01

Table A5: Twinning and Stress *in Utero* (Quintana-Domeque and Ródenas-Serrano, 2014)

	Twins (per 1,000)	Twins (per 1,000)	Twins (per 1,000)
ETA Bomb casualities 1^{st} trimester of pregnancy	0.023	-0.015	-0.015
	(0.058)	(0.057)	(0.043)
ETA Bomb casualities 2^{nd} trimester of pregnancy	-0.103***	-0.096***	-0.099***
	(0.036)	(0.037)	(0.036)
ETA Bomb casualities 3^{rd} trimester of pregnancy	-0.124*	-0.130*	-0.129**
	(0.065)	(0.075)	(0.058)
Observations	6,793,890	6,759,120	6,759,120
Year×month and province FE	Y	Y	Y
Socio-demographic controls		Y	Y
Provice-specific linear year-month trends			Y

NOTES: Data consists of live births conceived between January 1980 and February 2003. Treatment is defined as number of ETA bomb casualties in the province of conception. Full details are provide in Quintana-Domeque and Ródenas-Serrano (2014). Standard errors are clustered at the level of the province (50 provinces). *p<0.1; **p<0.05; ***p<0.01.

Table A6: Can Selective Maternal Survival Explain Twinning Rates? (Lee Bounds)

Twin×100	>140cm & BMI >16	>145cm & BMI >16.5	>150cm & BMI >17	>155cm & BMI >17.5
Upper Bound	0.371	2.339	0.723	-0.253
	(2.172)	(0.977)	(0.663)	(0.517)
Lower Bound	6.566	6.652	6.680	6.786
	(0.199)	(0.203)	(0.211)	(0.233)

Estimates of Lee (2009) bounds of the effect of treatment (positive health) on twinning. Selection is due to death during pregnancy, proxied by any sister of the index women suffering a maternal death. In each column, "healthy" is defined in the column title.

Table A7: Twins, Miscarriage and Maternal Health (Spain Vital Statistics)

Miscarriage×100	(1) All		
Primary	-0.602***	Twin×Primary	-0.662***
	(0.014)		(0.209)
Secondary	-0.720*	Twin×Secondary	-0.559***
	(0.015)		(0.209)
Tertiary	-0.800***	Twin×Tertiary	-0.651***
	(0.015)		(0.209)
Immigration	-0.072***	Twin×Immigration	0.228
	(0.0171)		(0.296)
Married	-0.074***	Twin×Married	-0.090
	(0.008)		(0.119)
No Father	0.686***	Twin×No Father	3.252
	(0.238)		(4.093)
Observations	2,869,329	R^2	0.01

Note: Spanish vital statistics data, 2007-2012. Outcome is a binary variable for miscarriage (late term fetal death) multiplied by 100 for presentation. Standard errors are presented in parenthesis.*p<0.1; **p<0.05; ***p<0.01

Table A8: OLS Estimates of the Q-Q Trade-off (n+ groups)

	Base Controls	+ Health	+Health &Socioec	Bord Controls
Two Plus				
Fertility	-0.152***	-0.0993***	-0.0990***	-0.0989***
	(0.00181)	(0.00178)	(0.00178)	(0.00177)
Observations	261,434	261,434	261,434	261,434
R^2	0.109	0.177	0.177	0.179
Three Plus				
Fertility	-0.142***	-0.0942***	-0.0939***	-0.0936***
	(0.00168)	(0.00162)	(0.00162)	(0.00161)
Observations	397,750	397,750	397,750	397,750
$R^2$0.093	0.167	0.168	0.170	
Four Plus				
Fertility	-0.123***	-0.0831***	-0.0829***	-0.0824***
	(0.00180)	(0.00170)	(0.00170)	(0.00170)
Observations	411,690	411,690	411,690	411,690
R^2	0.081	0.158	0.159	0.161

Notes: Two plus refers to all first borns in families with at least two births. Three plus refers to all first and second borns in families with at least three births. Four plus refers to first to third borns in families with at least four births. For additional notes, refer to table 9.

Table A9: First Stage Results

	2+			3+			4+		
FERTILITY	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	249,536	249,536	249,536	375,987	375,987	375,987	385,389	385,389	385,389
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	149,602	149,602	149,602	232,371	232,371	232,371	246,622	246,622	246,622
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	99,934	99,934	99,934	143,616	143,616	143,616	138,767	138,767	138,767

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 10. Standard errors are clustered at the level of the mother.*p<0.1; **p<0.05; ***p<0.01

Table A10: 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 list of controls are available in the notes to table 10. Standard errors are clustered by mother. *p<0.1; **p<0.05; ***p<0.01

Table A11: Estimates of a Maternal Health Shock on Quality and Twinning

	(1)	(2)	(3)
PANEL A: CHILD QUALITY			
Post × *basePneumonia*	0.0535***	0.0415**	0.0320
	(0.0191)	(0.0168)	(0.0238)
Observations	943,038	926,712	926,712
R-squared	0.009	0.009	0.009
PANEL B: PR(TWIN)			
Post × *basePneumonia*	0.00306	0.00277	0.00183
	(0.00224)	(0.00233)	(0.00302)
Observations	1,018,457	1,000,324	1,000,324
R-squared	0.001	0.001	0.001
PANEL C: ESTIMATE OF γ FOR CONLEY ET AL BOUNDS			
	0.0572	0.0667	0.0572
State, year FE	Y	Y	Y
State medical controls		Y	Y
State trends			Y

NOTES: Regression results for specifications (14a) and (14b) using the 5% sample of 1980 census data. Specifications and samples are identical to those described in Bhalotra and Venkataramani (2014). The estimate of γ is formed by taking the ratio of panel B and panel A estimates. A full description of this process, along with the estimation of the generation of the variance of γ is provided in appendix B, and figure **??**

D Appendix Figures

Figure A1: Birth Size of Twins versus Singeltons

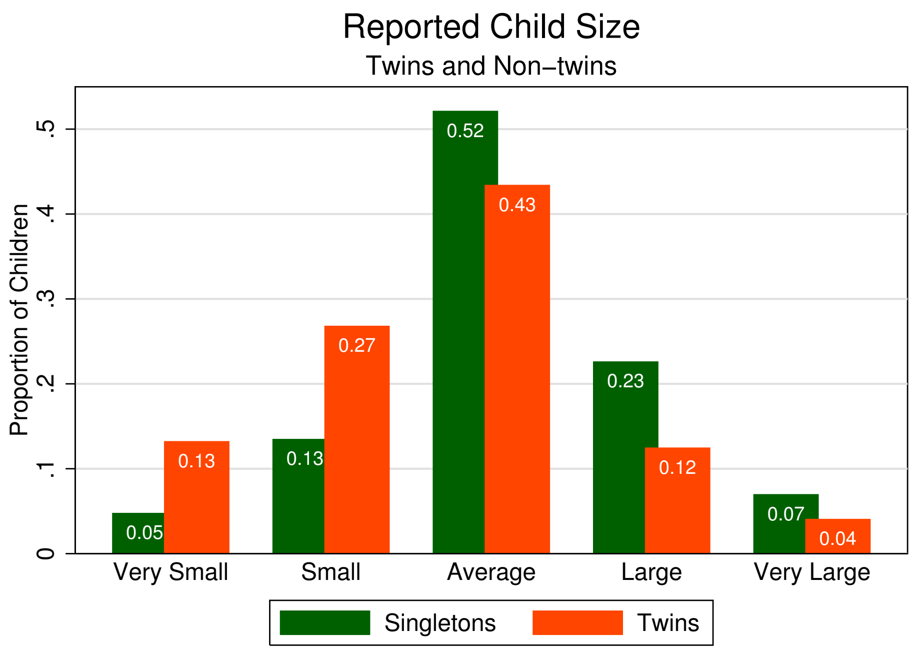


Figure A2: Height and Selective Survival

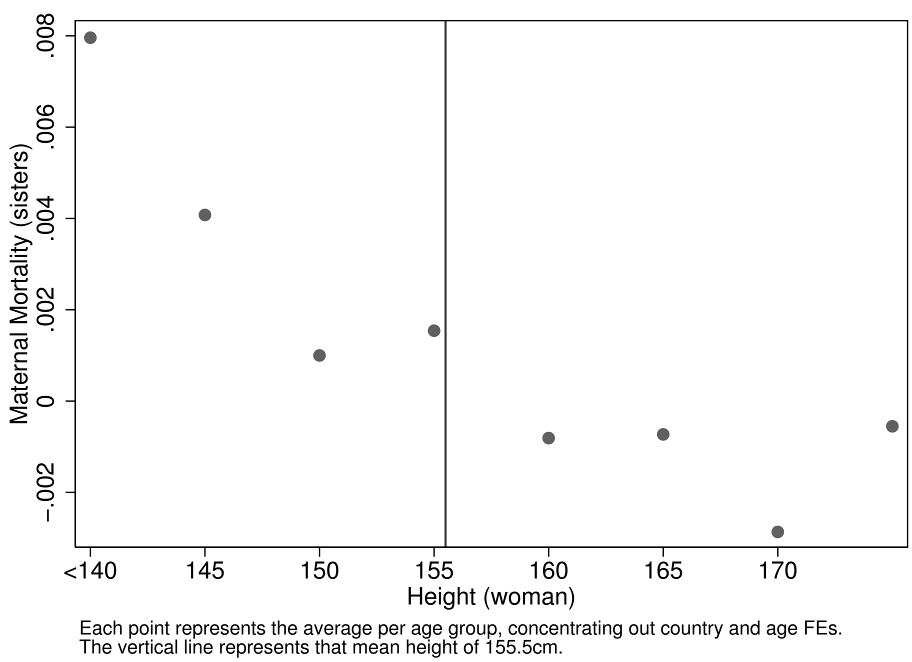


Figure A3a: Bootstrap Estimates of $\hat{\gamma}$

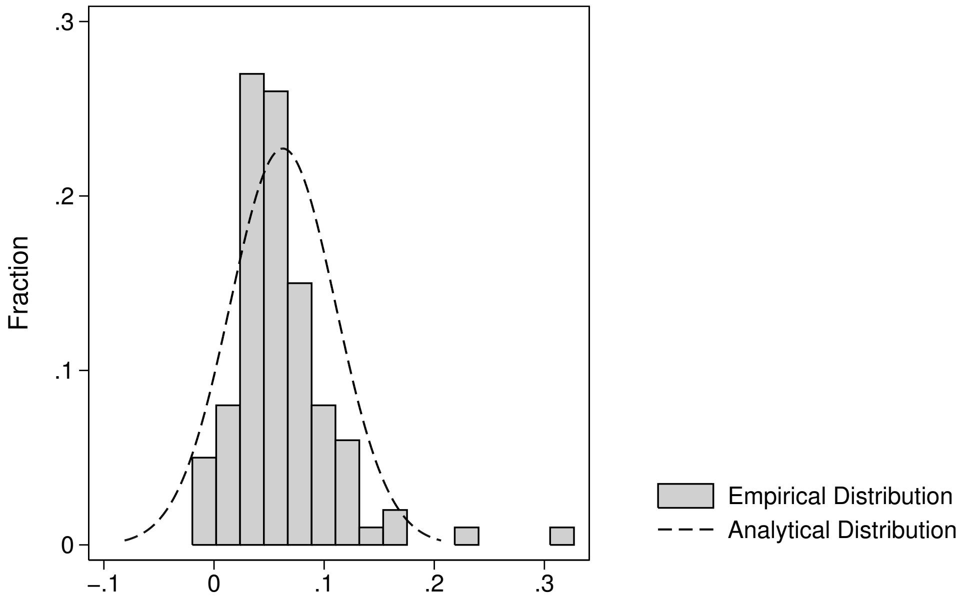
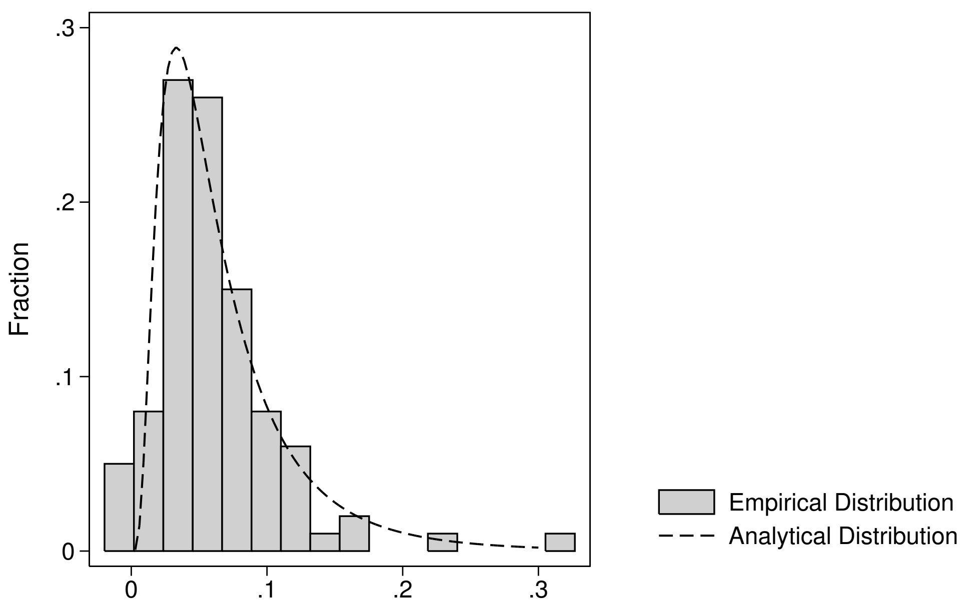


Figure A3b: Bootstrap Estimates of $\hat{\gamma}$



Note to figures A3b-A3a: The empirical distributionis generated by performing J=100 bootstrap replications to estimate ϕ_t and ϕ_q (see discussion in appendix B). The analytical distribution is a normal $\sim N(\mu_{\hat{\gamma}}, \sigma_{\hat{\gamma}})$, or log Normal with the same mean and standard deviation. The values estimated in the bootstrap replications are $\mu_{\hat{\gamma}} = 0.0625$ and $\sigma_{\hat{\gamma}} = 0.0479$.

Online Appendices

A Tables

Table B1: 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: Each year listed represents a DHS survey. 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 B2: Test of Balance of Observables: Twins versus Non-twins

	Non-Twin Family	Twin Family	Diff. (Diff. SE)
PANEL A: TWO PLUS			
Mother's Education	5.082	5.114	-0.0316
			(0.0963)
Mother's Height (cm)	155.6	157.2	-1.561***
			(0.142)
Prenatal care available	0.942	0.949	-0.00707
			(0.00469)
Mother's Age in Years	20.27	20.77	-0.492***
			(0.0774)
Age First Birth	19.85	20.34	-0.489***
			(0.0769)
Total Fertilty	3.627	4.524	-0.897***
			(0.0307)
Infant Mortality (pre-twin)	0.00370	0.00358	0.000120
			(0.00122)
School Z-score (pre-twin)	0.0995	0.107	-0.00790
			(0.0193)
Percent male child (pre-twin)	0.510	0.508	0.00268
			(0.0100)
PANEL B: THREE PLUS			
Mother's Education	4.057	4.174	-0.116
			(0.0767)
Mother's Height (cm)	155.7	157.1	-1.436***
			(0.125)
Prenatal care available	0.935	0.949	-0.0133**
			(0.00428)
Mother's Age in Years	21.35	21.82	-0.471***
			(0.0669)
Age First Birth	19.29	19.68	-0.388***
			(0.0620)
Total Fertilty	4.421	5.270	-0.848***
			(0.0263)
Infant Mortality (pre-twin)	0.00622	0.00508	0.00114
			(0.00131)
School Z-score (pre-twin)	0.00496	0.0167	-0.0117
			(0.0162)
Percent male child (pre-twin)	0.504	0.512	-0.00765
			(0.00752)

NOTES: All variables are at the level of the mother from full DHS data described in table 2. Panel A contains mothers of children who have had at least two births, where twin is defined as a twin at the second birth. Panel B contains mothers of children who have had at least three births, where twin is defined as a twin at the third birth. Education is measured in years, mother's height in centimetres, and prenatal care is binary, taking 1 if available in the mother's region. Diff. SE is calculated using a two-tailed t-test. *p<0.1; **p<0.05; ***p<0.01