Analysis of Twins

Sonia Bhalotra & Damian Clarke

(This paper also appears as CAGE Discussion paper 638)

September 2022 No: 1428

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

Analysis of Twins*

Sonia Bhalotra[†] Damian Clarke[‡]

September 29, 2022

Abstract

The occurrence of twin births has been widely used as a natural experiment. With a focus upon the use of twin births for identification of causal effects in economics, this chapter provides a critical review of methods and results.

JEL Codes: D10, I26, J13, J24, .

Keywords: Twins, identification, fertility, birth spacing, child development, women's labour market outcomes, child penalty.

*This paper was invited for the Handbook of Labor, Human Resources and Population Economics, section "Family Economics", and is forthcoming in the Handbook. We are grateful to the editor Klaus F. Zimmermann and to an anonymous referee for feedback. We thank Cristina Persico Blanco and Enzo Faulbaum Solari for outstanding research assistance. Sonia acknowledges part funding for her research time from Warwick University. Damian acknowledges financial support from the Millennium Science Initiative (Millennium Institute for Market Imperfections and Public Policy). This is a living reference work and will be updated to reflect additional material in the future. Please refer to the Handbook for any future updates.

[†]Department of Economics, University of Warwick, CAGE, CEPR, and IZA. Email sonia.bhalotra@warwick.ac.uk

[‡]Department of Economics, University of Chile, MIPP, CAGE and IZA. Email: damian.clarke@uchile.cl.

Contents

1 Introduction

The twin population is large and, more recently, is growing. On a global scale, approximately 1.6 million twins are born each year, with one in every 42 children born a twin. The global twin birth rate has risen by an astounding third over the past forty years (Monden et al., 2021). One reason for this is that families are increasingly postponing parenthood. While the risk of conception falls with age, the probability of twin conception increases because older women have higher levels of the follicle stimulating hormone (Beemsterboer et al., 2006; Pison and D'Addato, 2006). OECD figures indicate that the average age at first birth has risen by about 7 years since the 1970s, and 3.4 years since 2000. The other is the steadily increasing use of fertility treatments, some of which reflects the decaying chances of conception associated with delayed parenthood, but some of which reflects a rise in infertility even among younger parents. Couple-level infertility, defined as failure to achieve pregnancy after 12 months or more of regular unprotected sexual intercourse, is currently estimated at between 7% and 15% of all couples (Geyter, 2021; Ekechi, 2021). A meta-analysis of studies of the sperm count of men (unselected on fertility) in western countries concludes that it has halved in the last 40 years (Levine et al., 2017).

The phenomenon of twin birth has been leveraged across disciplines as a tool to advance scientific knowledge. This chapter discusses research in economics that has used twin birth to aid identification of causal effects. One strand of research in economics has leveraged twin birth as a shock to fertility, an unexpected increase in family size, to pursue causal identification of the impact of fertility on investments in children and on women's labour supply (Rosenzweig and Wolpin, 2000; Bronars and Grogger, 1994; Black et al., 2005). Another strand follows a tradition in behavioural genetics, demography and psychology, using comparisons between the individuals in a monozygotic twin pair to purge genetic influences on outcomes so as to assess the importance of nurture relative to nature (Polderman et al., 2015), and to seek to understand the return to human capital measures by studying between twin-differences, thus maintaining fixed genetics and home environments. This is not the first chapter attempting to review research on twin birth but the chapter attempts to highlight some of the challenges that prevailing methods face, and additionally, provide an overview of empirical findings from models based on twins. Existing surveys in economics include Rosenzweig and Wolpin (2000), Behrman (2015) and Kohler et al. (2011). Each of these surveys has a different focus to this current survey. Rosenzweig and Wolpin (2000) provides a broader discussion of 'natural experiments' in economics, where twins are discussed. Behrman (2015), the closest to this survey, focuses principally on describing the key methods which involve twins. Kohler et al. (2011) review the relationship between the economic and behavioral genetic approaches to the analyses of twins.

2 The Twin Instrument

Most research in economics that uses twin birth to instrument fertility studies impacts of fertility either on child outcomes (most often education), or on parental outcomes (most often women's labour supply). Examples of studies using the twin IV to explain child outcomes include Rosenzweig and Wolpin (1980b) and Hanushek (1992), while studies that use the twin IV to study women's labour market outcomes include Rosenzweig and Wolpin (1980a) and Jacobsen et al. (1999). The motivation for attempting to identify causal relationships is that there are empirical regularities in these relationships. In cross sectional and time-series data, the number of children in a family tends to exhibit a negative association with both the educational attainment of children and the labour market outcomes of women. However it is plausible that this relationship is not causal, but reflects preferences- for instance, women who have a preference for large families may have a preference for a career. This is why a source of exogenous variation in the number of children a family has is required. Under certain assumptions, the shock of twin birth provides this variation.

2.1 The Initial Set-up

In their pioneering research, Rosenzweig and Wolpin (1980a) used twin birth to study the impact of fertility on women's labour supply, and Rosenzweig and Wolpin (1980b) studied its impact on investments in children. These papers use the ratio of twins to all births in a reduced form specification, on the assumption of (conditional) exogeneity. They condition upon maternal age and parity, as the biomedical literature has noted associations of twin birth with these variables. The basic insight in the earlier reduced form approach was set out in an instrumental variables (IV) strategy by Black et al. (2005) and this has been widely adopted since.

2.1.1 Twins and Sibling Outcomes

The 2SLS model of Black et al. (2005) is laid out as follows: :

$$
\begin{aligned}
fertility_{ij} &= \pi_0 + \pi_1 twin_j + X'_{ij}\Pi + \nu_{it} & (1)\\
y_{ij} &= \beta_0 + \beta_1 fertility_{ij} + X'_{ij}\Gamma + \varepsilon_{ij} & (2)
\end{aligned}
$$

The first stage consists of instrumenting fertility of mother i in family j with twin births occurring to that mother.y_{ij} is the outcome of interest, for example, the education of a child. Observable controls X_{ij} often include age and birth order. The question of omitted controls potentially correlated with the instrument is discussed in section 2.2, and the outcome variables are discussed further in section 2.3 and 4.1.

Black et al. (2005) seek to address the problem that any fertility choices following the birth of twins are potentially endogenous. Clearly a family is less likely to continue fertility if they have had twins than if, at the same birth order, they have had a singleton birth, and there is possibly endogenous heterogeneity across families with respect to fertility preferences and fertility control. The solution they propose is to analyse outcomes for children born *before* the potential twin birth. Treated siblings are those followed by a twin, while untreated individuals are those followed by a singleton. To ensure that the treated and control samples are balanced on birth order they propose estimating the model by the birth order at which the potential twin birth occurs, and conditioning the sample on having at least as many children as are implied by the particular birth order. For example, investments in first born children in families with at least two births could be analysed (this is referred to as the 2+ sample), using as an instrument for fertility an indicator of whether a twin vs a singleton was born at birth order 2. The outcome is always defined for children born before twins are born, but Table 1 lays out the manner in which the analysis sample varies with the birth order at which twins occur.

If all families only had two birth events, the twin instrument would shift fertility upwards by exactly one child. In practice, as families have heterogeneous preferences over their total number of children, and families do not have full control over actual fertility, twin birth tends to increase fertility by between 0.5 to 0.9 births, depending on the context and the birth order at which twinning occurs. In high fertility environments, families who have twins at a low birth order have ample time to adjust for the twin shock by reducing future fertility. However, where low fertility is the norm, or else when twins occur at a higher birth order, compensating changes in future fertility are smaller and twin births have a larger impact on completed fertility, reflected in a larger first stage coefficient.

Table 1: Samples and instruments for IV based on twin birth

Sample	Definition of Estimation Sample	Instrument
"2+"	1^{st} births in all families with at least 2 births	Twin birth at birth = 2
"3+"	1^{st} and 2^{nd} births in all families with at least 3 births	Twin birth at birth = 3
"4+"	1^{st}, 2^{nd} and 3^{rd} birth in all families with at least 4 births	Twin birth at birth = 4
"K+"	1^{st}, …, (K-1)th birth in all families with at least K births	Twin birth at birth = K

Average causal response functions. As is the case with all IV estimates, estimates using the twin instrument provide a Local Average Treatment Effect (LATE), 'local' to individuals whose behaviour is changed due to the instrument. The implications of this in the case of twins is now considered. Firstly, note that the rate of compliance (the share of individuals who increase fertility owing to a twin birth) is high. As discussed previously, this compliance can be summarised by the parameter π_1 in equation 1. If π_i=1, this implies that all women who have twins go on to have 1 birth more than women who do not, while if $\pi_1 < 1$ this implies that some proportion of women are always takers, and have the same number of future births whether or not they have twins.

As noted by Angrist and Imbens (1995); Angrist et al. (2010), where the endogenous variable of interest is multi-valued, rather than binary, the LATE must be cast in terms of the variation which the instrument generates in the endogenous variable of interest. They refer to these as Average Causal Response (ACR) functions. In the case of twins, this refers to the variation in fertility *at particular margins* that is generated by a twin birth. Thus, where the twin instrument refers to twins at birth order 2, the causal estimate which is recovered is the impact that a twin birth has on future fertility. For some individuals, twin born at birth order 2 may shift the total number of births from 2 to 3 while, for other individuals, they may shift births from 3 to 4. These higher order shifts are likely to be more relevant the lower is a woman's control over fertility. The ACR functions can be estimated by regressing the indicator for births at any particular birth order on the twin instrument. A visual presentation of ACR functions provides a simple way to understand the variation in fertility generated by the twin instrument. ACR estimates are provided, for example, in Angrist et al. (2010); Bhalotra and Clarke (2020). The ACR describes the weighting functions over parities induced by instrumental variation and this is useful if, as in Bhalotra and Clarke (2020), the researcher wants to observe how the weighting function changes conditional upon a particular set of controls.

2.1.2 Twins and Parental Outcomes

Equations 1-2 describe the canonical model used to estimate the impact of sibship size on child outcomes. When the outcome of interest is a parental outcome such as the mother's earnings, or divorce, the analysis proceeds with a single observation for each family. The analysis is now implemented at the level of the parent, so there are no sample restrictions but the approach still allows the relationship to vary with the parity at which twins occur. This model has been discussed by Angrist and Evans (1998) who cite, among earlier applications (often implementing reduced form models), Rosenzweig and Wolpin (1980a); Gangadharan et al. (1996) and Bronars and Grogger (1994). Angrist and Evans (1998) analyse the impact of fertility on women's labour supply, using the occurrence of twins at second birth as an instrument for having at least 3 births. This model is estimated by 2SLS as follows:

$$\text{Three Births}_i = \pi_0 + \pi_1 \text{Twins At Second Birth}_i + \boldsymbol{x}_i'\boldsymbol{\gamma} + u_i \tag{3}$$

$$\text{Labour Supply}_i = \beta_0 + \beta_1 \text{Three Births}_i + \boldsymbol{x}_i'\boldsymbol{\delta} + \eta_i \tag{4}$$

where equation 3 captures the first stage impact of twins at second birth, which implies at least three births, and 4 captures the impact of a third birth on labour supply measures for the mother (or father) i. In general, covariates $\boldsymbol{x}_i$ include maternal age, age at birth, and child gender. These models are estimated with the sample of families that have at least two births. Applications, as well as extensions to other birth orders include, for example, Vere (2011) (marginal changes at the second and third births), or Silles (2016); Zhang (2017) (marginal changes at the first, second and third births). They can be extended to study impacts of the marginal birth at higher parity, but note that these models do not allow us to study impacts of extensive margin fertility i.e. impacts of first birth on parental outcomes.

2.2 Identification Challenges

Identification of the impact of fertility on parent or child outcomes using the twin instrument requires that, conditional on observable characteristics, twinning is as good as random. Moreover, twin birth must not affect the outcome (investment in children, child human capital or parental labour supply, for instance) beyond its impact on fertility. This is the exclusion restriction. Early papers in the twin IV literature including Rosenzweig and Wolpin (1980a,b) allow that twinning depends upon maternal age and parity, in view of biological evidence of this (Hall, 2003; Bulmer, 1970). As these are typically observable, they are controlled for in the 2SLS procedure. Another widely recognized determinant of twinning is assisted reproductive technology Braakmann and Wildman (2014, 2016); Farbmacher et al. (2018). As discussed later in this paper, the risk of twin birth has stood in the region of 25–30% during the double embryo transfer regime of in vitro fertilization (IVF), compared with a global rate of 2–3%. Again, where use of IVF is observable, this can be included as a control. Alternatively, identical (Monozygotic, or MZ) rather than non-identical (Dyzygotic, or DZ) twins can be used to instrument fertility given that fertility treatments only increase the likelihood of DZ twinning (Farbmacher et al., 2018).

However, more recently, twinning has been shown to depend upon maternal health. This violates the assumption that twin birth is quasi-random. Using data on almost 17 million births in 72 countries, of which 462,246 (2.73%) are twins, Bhalotra and Clarke (2019) show that the mothers of twins are selectively healthy. They demonstrate that this is the case using sixteen different markers of maternal condition, including health stocks and health conditions prior to pregnancy (height, obesity, diabetes, hypertension, asthma, kidney disease, smoking), exposure to unexpected stress in pregnancy and measures of the availability of medical professionals and prenatal care. The underlying mechanism, which the authors test, is that twins are more likely than singletons to suffer miscarriage and more so when maternal health is challenged. So, even if twin conceptions are random, healthier mothers are more likely to be able to carry twin conceptions to term. The effects are sizable, with a 1 standard deviation improvement in the indicator tending to increase the likelihood of twinning by 6–12%. This is important because it is virtually impossible to observe all relevant aspects of maternal health. To take a few examples, fetal health is potentially a function of whether pregnant women skip breakfast (Mazumder and Seeskin, 2015), they suffer bereavement in pregnancy (Black et al., 2016), or they are exposed to air pollution (Chay and Greenstone, 2003). The authors further show that maternal health is potentially correlated with the outcomes of potential interest. As a result, estimates obtained using the twin instrument are biased. Notably, the twin IV bias is positive, while the OLS bias that motivates the search for an instrument is negative. Bhalotra and Clarke (2020) show that, given the power of the twin instrument, credible bounding exercises can be conducted, and the direction of bias determined.

While Bhalotra and Clarke (2019, 2020) highlight that mothers of twins are selectively healthy, Rosenzweig and Zhang (2009) highlight that twins have weaker health endowments at birth. Although this is widely known, Bhalotra et al. (2022) study population-level impacts of a sharp policy-driven drop in the share of twin births in Sweden to profile the fact of lower endowments of twins using a range of measures of health at birth. They refer to evidence from other studies that parents often respond to differences in endowments across their children by acting to reinforce or compensate. They find reinforcing behaviour in China. Bhalotra and Clarke (2020) find that parents in developing countries and in the US reinforce endowments in their breastfeeding behaviour. Using information on the extent to which parents read with their child, which is only available in the US (in the NLSY), they find that the reading investments of college-educated mothers tend to compensate endowments. Rosenzweig and Zhang (2009) argue that any impact of twin birth in the family on children born before the twins might then reflect not the added child (fertility) but re-allocation of resources across children in response to having children with weak endowments. Research using the twin instrument should investigate controls for maternal health and child health endowments but as these some part of these factors will always be unobservable, partial identification methods provide a way forward Bhalotra and Clarke (2020); Rosenzweig and Zhang (2009). As twinning is a relatively infrequent outcome, and IV procedures have wide confidence intervals generally, even when using survey data, surveys need be large to achieve reasonable power in IV models.

2.3 Extensions and Limitations

Power limitations. This approach has been extended in more recent work. A common weakness of the IV approach is that twin births are relatively infrequent at 2–3 percent of all births, and there are well known challenges in conducting inference in 2SLS models with large standard errors. Thus making progress with the parity-specific approach relies upon access to large data sets. In an attempt to address this limitation, Angrist et al. (2010) demonstrate how the researcher can pool the parity-specific samples to achieve the power to detect effects. They document that if one assumes a common causal (ie linear) parameter on fertility, then a residualised version of the twin-instrument, conditioning on parity-specific controls, will be independent of fertility provided the same assumptions are met as in parity specific twin-IV models, even if now pooling across various parity-specific groups. A different issue is raised in Mogstad and Wiswall (2012), who note that this pooling process proposed by Angrist et al. (2010) will only be valid if the relationship between twinning and included controls is linear. They offer a non-parametric refinement to the Angrist et al. (2010) pooling process which loosens the linearity assumptions, and hence is robust to this problem. Precise proposed estimation procedures are laid out in Angrist et al. (2010); Mogstad and Wiswall (2012).

Allow non-linearity in birth order. In a further innovation Mogstad and Wiswall (2016) relax the assumption of linearity in the effects of additional births on family size to allow that the marginal birth at different birth orders may have different implications for the outcomes. For instance, scale economies or scale diseconomies are unlikely to evolve linearly with the number of children a family has. In practice, moving from two births to three births may imply very different family responses then moving from three births to four births, in which case linearity would be an unreasonable assumption. Their estimator replaces the single endogenous fertility measure instrumented by twins at a particular birth order with a range of measures of family size exceeding specified thresholds, allowing twins at different birth orders to move family size differently across the different thresholds.

The twin instrument does not identify impacts of extensive margin fertility A limitation of the twin IV approach is that it cannot deliver estimates of extensive margin impacts of fertility, that is, the impacts of a first child. A twin birth at birth order 1 increases fertility from one to two children while a twin birth at birth order 2 increases fertility from two to three children. Twin births cannot illuminate the consequences of having a child vs no child. Yet the extensive margin is important. A number of studies reveal that women's earnings and employment rates start to deviate from those of men (of similar age and education) after their first birth. Moreover, women's earnings do not fully recover, registering a child penalty as long as ten years after the first birth. Additional births may modify this but their impact appears to be relatively small. In fact another commonly used instrument for fertility is similarly restricted to providing estimates for intensive margin impacts. This is the sex mix instrument, defined on the gender of the first two (or three) children. It acts as an instrument for fertility on the premise that most families have a preference for one son and one daughter. Thus, a family with two children of the same gender is more likely to have a third child than a family with two children of opposing gender. The first application was to US data by Angrist and Evans (1996), but it has been born out in a few settings, with the first stage for the sex mix instrument tending to be strong.

Recent approaches to identifying impacts of extensive margin fertility are now discussed. In a study using the entire population of Danes rather than the IVF population, Kleven et al. (2019b) propose that impacts of first child birth on parental earnings and employment can be identified from idiosyncratic variation in birth timing across women, conditional upon cohort and year fixed effects. The counterfactual for a woman giving birth at, for example, age 22, is a woman of the same birth cohort who gives birth at a later age such as, for instance, age 25. A number of studies, in a range of countries, have replicated their main finding, which is of a large and highly persistent child penalty on the earnings and other labour market outcomes of mothers but not fathers, see for instance, Berniell et al. (2021); Andresen and Nix (2022), although see Bhalotra, Clarke and Nazarova (2022) for a discussion of the assumptions underlying the model of Kleven et al. (2019b).

With a view to identifying impacts of the birth of a first child, Lundborg et al. (2017) rely upon the chances of success with IVF procedures being idiosyncratic, or uncorrelated with family characteristics. Using Danish administrative data and restricting the sample to women who undertake IVF, they use the binary IVF success indicator as an instrument for whether the woman has one child or no child. They find a significant impact of having a child on women's earnings, larger than the median intensive margin impact in the literature. They argue that their results generalize to the wider population. However, 25% of women who give birth with IVF give birth to twins (and the average increment in fertility in the sample is thus 1.25, not 1). Thus their estimate of extensive margin fertility impacts is combined with the impact of having twins at birth.

Using Swedish administrative data Bhalotra et al. (2022) also focus on women undertaking IVF. They leverage a policy reform effective from 2003 that mandated single embryo transfer for IVF, against the default of double embryo transfer. The reform led to a precipitous drop in the share of twin births from about 30% to 5% and, by 2018, it had fallen to 2.54%, which is one of the world's lowest twin birth rates among IVF users. They identify impacts on a number of outcomes including women's earnings. Their objective is not particularly to identify impacts of extensive margin fertility but they do provide estimates for women at first parity, for whom the estimated increase in earnings reflects the impact of having one child rather than two. In other words, they estimate impacts of a decrement in fertility by one child. However, now the impact of having one less child is not separable from the impact of twins (or of two children with no birth spacing).

The twin instrument cannot identify the impact of fertility on younger siblings. The fact that the twin instrument does not allow for the consideration of impacts of high fertility in a family on the later born children (in cases where the later born are born after twins) is a limitation. In practice, it may be that later-born children are most affected, be it by time or financial constraints imposed by having a larger number of siblings, or by learning from older siblings. This is discussed by Rosenzweig and Zhang (2009), who note also that twin births themselves are excluded from the estimation sample. A child may suffer lower parental investment on account of being born with a twin, and this may influence their eventual educational outcomes, but the twin instrument has nothing to say on this.

Twin instrument estimates are sensitive to the birth interval preceding twins and the age at which sibling outcomes are measured. As discussed earlier, when twins are used as an instrument the estimation samples are constructed to isolate potentially exogenous variation in fertility. Specifically, twinning at birth order K is seen as a shock, and the outcomes potentially impacted by the shock are of children born before twins, at birth orders $1, \ldots, K-1$. Thus, the twin instrument allows us to identify the impact of higher order (later) fertility on earlier born children. However, what is often not made explicit but is relevant is that the exposure of an earlier-born child to the marginal birth (the fertility shock) will be decreasing in the birth interval between the two siblings. For instance, if a family has twins when their first born child is age two, this child may experience a different (possibly larger) decline in parental inputs than if the child was age six when the twins were born. The estimated impact of a marginal birth, stemming from having twins, may further depend upon the age at which the outcome of the older sibling is observed. Investments in a newborn such as breastfeeding are often quite different from investments in a toddler, but later life investments, being more similar, may be more complementary. For example parents can read jointly to children age 3 and 5.

Turning to the case of fertility impacting parental labour market outcomes, many mothers are entitled to maternal leave, and many take further time out when the child is young. The impact of fertility on women's earnings is likely to be attenuated after this period, as formal childcare and school provision become available. When studying impacts of fertility on parents, Angrist and Evans (1998); Aaronson et al. (2020) focus exclusively on women of relatively young children, who can be linked to their parents through census microdata. Often papers stratify by the time since childbirth, see for example Vere (2011); Zhang (2017); Silles (2016) (among others). More recently, Kleven et al. (2019b) study dynamic impacts of first birth on women's earnings, profiling both short and long-term impacts of fertility. The evidence indicates that women's earnings never recover from motherhood, with the penalty 10 years after birth being as large as 20–40% in many countries (Kleven et al., 2019b,a).

The studies of impacts of fertility on sibling outcomes vary in the age at which sibling outcomes are measured, and their results must be understood in terms of the sample studied. Black et al. (2005); Angrist et al. (2010) observe children at the end of their formal schooling period, and as such estimate impacts on human capital achievement at adulthood. Other papers, such as Bhalotra and Clarke (2020) consider impacts of fertility on school age children, and as such estimate impacts of fertility on a flow measure of human capital. Some studies, for instance, (Baranowska et al., 2017), consider other outcomes like survival and this can be measured to the end of an individual's life or, as is common in poor countries where child mortality is high, as survival to age five.

3 Between-Twin Models

Twin births have also served a different purpose in research by economists, which is to hold constant genetic and family endowments in order to isolate the role of factors such as education or birth weight that vary within twin pairs. In their review Rosenzweig and Wolpin (2000) suggest that the earliest presentation of earnings differentials based on twins dates to Behrman and Taubman (1976). Early modelling considerations based on within-twin and within-family designs are discussed in Goldberger (1979). Early studies including Behrman and Taubman (1976, 1989); Ashenfelter and Krueger (1994); Behrman et al. (1996); Ashenfelter and Rouse (1998); Rouse (1999) study the impact of education on labour market success. More recent papers have sought to estimate returns to education in other dimensions, such as health (Amin et al., 2013; Behrman et al., 2015; Lundborg, 2013), or children's education (Behrman and Rosenzweig, 2002), or returns to health, often health at birth (Behrman and Rosenzweig, 2004; Behrman et al., 2011; Royer, 2009).

3.1 The Basic Set-up

Consider the relationship between an outcome y_{ij} and an independent variable of interest x_{ij} which varies between children i in family j. Pooling observations over children and families, one could estimate a linear model of the form:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \boldsymbol{w}'_{ij}\gamma + \eta_{ij} \tag{5}$$

where η_{ij} refers to an unobserved error term, and $\boldsymbol{w}$ a vector of individual or family-level controls. Estimating β_1 by OLS on pooled data would lead to unbiased estimates only in the quite restrictive case where—conditional on included controls $\boldsymbol{w}_{ij}$—the dependent variable x_{ij} is independent of unobserved factors η_{ij}. In the case where x_{ij} is an endogenous variable expected to depend on both family and genetic factors, twins can potentially offer a way forward. To see this, write the unobserved error term from equation 5 as a function of family or household characteristics (h_i), genetic factors (g_{ij}), and an additional unobserved component (u_i):

$$\eta_{ij} = f(h_i, g_{ij}) + u_{ij}. \tag{6}$$

In equation 5, the composite error term will result in estimation bias if any unobserved factors are correlated with both the dependent and independent variables. For example, in the case where x_{ij} refers to educational attainment, and y_{ij} to labour market outcomes, family-level measures such as parental investment in children are likely to be positively correlated with both educational and labour market outcomes, generating a positive bias in the estimated coefficients.

Data on twins can help address this bias. Provided that twins live in the same household and have identical genetic endowments, h_i is a common term, and $g_{i1} = g_{i2}$, where $j \in \{1, 2\}$ refers to the two children with a twin pair. A within-twin transformation of equation 5 can be written as:

$$y_{i2} - y_{i1} = \beta(x_{i2} - x_{i1}) + (\boldsymbol{w}_{i2} - \boldsymbol{w}_{i1})'\gamma + (u_{i2} - u_{i1}), \tag{7}$$

where consistent estimation now no longer requires that x_{ij} be independent of family-level and genetic endowments, but only that within twin pair differences in x_{ij} are conditionally independent of within-twin pair differences in unobservables which are not captured by genetic or household fixed effects.

The assumption of identical genetic endowments is only reasonable for monozygotic (MZ) twins. Dizygotic (DZ) twins share only 50% of their genetic material, see Taubman (1976b) for a nice early discussion of this. Along this spectrum, differencing across siblings who are not twins will remove some common family-level factors, such as access to a clinic or parental mental health, but their genetic endowments will of course not be identical. Thus, where controlling for genetic endowments is important, for example, where they represent elevated risk of a chronic disease, within-twin models offer cleaner identification than within-sibling models. Differencing the relationship of interest between the two individuals in a twin pair purges common genetic determinants. This is, of course, only useful if genetics are a relevant omitted variable, that is, if genes are correlated with both the outcome and the independent variable of interest

3.2 Identification Challenges

The twin fixed effects approach encounters a number of identification challenges. These include concerns relating to power, measurement error, potential confounders which remain after taking within twin transformations, potential family responses to twinning that complicate unbiased estimation, and the generalizability of the estimates.

Power limitations. A potential weakness of the between-twin strategy is that it may be under-powered. Identification relies upon differences between the individuals in a twin pair, which are often relatively small. For example, researchers have used the between-twin estimator to estimate causal impacts of birth weight on human capital outcomes (Bharadwaj et al., 2018; Royer, 2009), but it is uncommon that twins have substantially different birth weights. Similarly, the observed difference in education between twins is not large, estimated at 1 year by Savelyev et al. (2022), using the Minnesota Twins Registry.

When differencing within twin pairs, estimation will be driven off the (generally) small difference between twin pairs, and this will magnify the importance of measurement error, leading to a potentially substantial attenuation bias and the bias is increasing in the within-pair correlation (Griliches, 1979; Rosenzweig and Wolpin, 2000). This is a concern given that within-twin correlations in outcomes are generally very high. For example, Behrman and Rosenzweig (1999) report correlations of schooling at 0.74 within identical twin pairs, similar values are reported in Ashenfelter and Rouse (1998) and birth weight is also highly correlated within twin pairs Royer (2009); Almond et al. (2009). Ashenfelter and Krueger (1994) proposed a way to avoid much of the burden of this problem, which is now widely adopted. They proposed using reported differences in the independent variable of interest reported *by each twin*, and using a 2SLS procedure where the within-twin difference reported by one twin pair is instrumented using the within-twin difference reported by the other twin pair. This method recovers consistent estimates even if measurement error is correlated between twin pairs (Ashenfelter and Krueger, 1994). This proposal has influenced the collection of large twin datasets, and an instrumental variables approach based on a similar logic, such as reports from the children of twins (Behrman et al., 1994).

Twin differences may be correlated with the outcome. Another challenge to identification is that confounders often remain after twin-differencing. Even if twins have a different realization of the independent variable (education or birth weight for example), this difference must truly be random. While this may be the case, for example if one twin is randomly assigned to better teachers (Lundborg, 2013), or a differential position in the womb leading to different nutritional intake (Black et al., 2007; Almond et al., 2005), there are reasons to think that this may not necessarily be the case. For example, congenital malformations may be correlated with both birth weight and later life outcomes (Royer, 2009), or accidents during childhood may lead to lower schooling and lower labour market returns. Such biases may be considerable (Bound and Solon, 1999; Neumark, 1999). For example, Sandewall et al. (2014) note that if one considers measures of IQ during adolescence, twins with more education typically have higher IQ, and controlling for this reduces estimated returns to education by around 15%. If such factors can be observed and controlled for, unbiased estimates can be recovered, but if factors are inherently unobservable (for example ability differences within twins), in the best case only partial identification can be achieved (Neumark, 1999).

Behavioural responses of parents. Another concern, somewhat related to the previous point, is that twins themselves or their families may respond to twin-specific realizations of the independent variable. For instance, the twin with a weaker endowment may work harder to compensate, or the parent may seek to implement intra-family transfers to either compensate or reinforce the individual twin endowment. These behavioural responses may invalidate a causal interpretation of parameters. A biological mechanism which for within-household feedback between twins is noted in Bütikofer et al. (2019), though, more generally, there is scope for behavioural mechanisms. There is evidence of parental investments being compensatory (Bharadwaj et al., 2018; Savelyev et al., 2022), which will attenuate estimates of any returns to the endowment being studied, as parents will endogenously cushion the blow of a weaker endowment, at the cost of those with a stronger endowment. However, other studies identify reinforcing behaviour, and whether parents compensate or reinforce tends to vary with whether the outcome is heath or education (Yi et al., 2015). It makes sense that compensating low birth weight by for instance breastfeeding or nourishing the child could achieve the basic parental desire to ensure child survival. At the same time, parents might reinforce innate ability either because the cost of effort is lower, or potential returns higher. This behaviour may, plausibly, vary with the mother's education (Bhalotra and Clarke, 2019). The relevant literature is surveyed in Almond and Mazumder (2013) and Currie and Almond (2011). Overall, the evidence suggests that while parents may compensate or reinforce between non-twin siblings, they treat twins more equally essentially because many investments in twins are naturally shared (such as reading a book together) (Royer, 2009; Bharadwaj et al., 2018; Currie and Almond, 2011). This appears to be the case even in studies where the birth endowment is held constant using gene-realizations that are fixed at conception (Sanz-de Galdeano and Terskaya, 2019). In general, understanding how endowment differentials emerge, and if these result in unobserved responses within families even conditional on shared genes, is a key question which must be addressed should causal effects be argued.

Representativeness of between-twin estimates. It was discussed above that differences in the independent variable of interest between twins are often small. In addition to raising the risk that the estimates are imprecise, this has implications for interpretation of the estimates. In particular, if twins are more likely to have low or high realizations of the independent variable of interest (for example, twins do have lower birth weight than average), the between-twin estimator is not informative of the impacts of potential changes in the areas of the distribution in which little variation is observed (eg high birth weight). This is, of course, only a concern if the relationship of interest is non-linear. However, this is often the case-returns to a marginal increase in education will tend to depend upon the baseline level of education, and returns to an increment in birth weight may be life-saving if they occur in the lower regions of the baseline distribution of birth weight, and of no particular consequence if they occur in the region of normal birth weight.

Another issue to note is that twins are not representative of the population, either in their own observable characteristics (see for example Almond et al. (2005); Bhalotra et al. (2022)), or in terms of their mother's endowments (Bhalotra and Clarke, 2019). As fertility treatment rates are on the increase, twins will increasingly be over-represented among families that have undergone IVF or other assisted reproductive techniques (Beral et al., 1990; Braakmann and Wildman, 2016; Farbmacher et al., 2018), families in which parents are often older and more educated than in the general population. This does not challenge internal validity but it does mean that the onus falls upon the researcher to demonstrate external validity. However, concerns over external validity of the twin-differencing approach are only relevant if the *return* to characteristics is different between twins and others (Lundborg, 2013). Discussion of these and related considerations is in Boardman and Fletcher (2014); Amin et al. (2014a,b), and Blanchflower and Elias (1999) who suggest considerable caution in interpretation of within twin results.

4 The Evidence

Without attempting to be comprehensive this section reviews the evidence in the literature using twins to aid identification.

4.1 Family Trade-offs

Family size and parental labour market outcomes The evidence emerging from twin-IV estimators seeking to estimate the impact of child birth on labour supply are now considered. Early work in this domain used US census microdata. Bronars and Grogger (1994) and Jacobsen et al. (1999) use samples from the 1970 and 1980 US census, while Angrist and Evans (1998) use the 1980 and 1990 census. Jacobsen et al. (1999) primarily report reduced form estimates of twin birth on labour supply decisions, hours worked per week, weeks worked per year and earned income, finding large impacts up to around 4 years after birth, however insignificant impacts thereafter. Bronars and Grogger (1994) compare unmarried women who first gave birth to twins with unmarried women whose first birth was a singleton, and interpret the difference as the impact of unplanned birth. They find adverse impacts in the short term on the mother's labour-force participation, poverty, and welfare dependency (and not so among married mothers). Moreover, while most of the adverse economic effects of unplanned motherhood dissipate over time for white women, they are larger and more persistent among black women. Bhalotra et al. (2022) similarly consider the dynamic impacts of twin vs singleton birth, adapting the estimator of Kleven et al. (2019b). They identify short run penalties on mothers' careers that do not persist beyond two years, alongside positive impacts on father's careers in Sweden.

IV results reported by Angrist and Evans (1998) focusing on women aged 21-35 suggest that a third birth reduces labour supply at both the intensive and extensive margins (by 8% in terms of participation decisions), as well as total labour income, with no significant declines observed among husbands. Other later papers similarly work with these data – Cáceres-Delpiano (2006) documents adverse impacts of child birth on mother's LFP and divorce (related to child investment decisions discussed below), and Vere (2011) extends the analysis to examine effects at higher birth orders, on fathers, allowing rich dynamic effects over time since birth, and calendar time.

Further studies have analysed data from other countries. Cáceres-Delpiano (2012) provide results for 40 developing country samples covering Africa, Asia, and the Americas, broadly finding that child birth results in declines in women's labour market participation, and the types of work done by women, noting heterogeneity by birth order. Silles (2016) studies child penalties in the UK, using pooled survey data from 1986-2009 and again finding large effects of childbearing on women's labour supply and earnings when children are younger (aged 13 and below), which are rendered insignificant when children are older, across birth orders 2-4. In the UK sample, in a methodological advance Braakmann and Wildman (2016) note that these results broadly hold if controlling for fertility treatments, once again noting declining impacts of marginal births over time. Zhang (2017), study labour force participation in Taiwan, showing broadly similar patterns of dynamic effects over time, but noting that fertility has no additional impact on participation at higher birth orders.

A study of particular note as it provides results over a considerable time frame and geographic range is Aaronson et al. (2020). Using data from 441 census waves and surveys from 103 countries between the years 1787-2015, they study the impact of fertility on the labour supply of mothers aged 21-35. Based on a twin IV strategy, they document (among other results) trade-offs from having a third birth which are largest in high income settings, while small and often indistinguishable from zero in low income settings.

While the discussion here focuses on the many studies of labour supply and earnings of mothers and fathers, there are studies of marriage decisions, welfare receipt and future educational outcomes (Bronars and Grogger, 1994), while Cáceres-Delpiano (2006) notes impacts on divorce. A quite different focus is that of the effects of family size on siblings, which are turned to now.

Family size and investment in children Both cross sectional and time series data have, historically, suggested a negative relationship between family size and child outcomes. Yet more credible evidence based on large microdata bases points to small or insignificant results when estimating the impact of family size on indicators of children's human capital. For instance, using the twin instrument, Black et al. (2005) find no impact of fertility on educational attainment in Norway, and Angrist et al. (2010) similarly find no evidence of a trade off for educational and labour market outcomes in Israel. Interestingly, Black et al. (2005) find large *birth order* effects. In particular, second-borns have lower attainment than first borns, and so forth, without there being an additional impact of the total number of children born . A broadly similar pattern of results has been documented in other settings where the twin estimator has been used, for instance, Fitzsimons and Malde (2010) (Mexico), Cáceres-Delpiano (2006) (USA), Li et al. (2008) (China), Dayioglu et al. (2009) (Turkey) and Ponczek and Souza (2012) (Brazil). In certain settings, significant effects emerge in certain groups, such as rural families in China, or girls in Brazil (Li et al., 2008; Ponczek and Souza, 2012), though these are not consistently observed across groups.

Against this backdrop, more recent papers that revisit properties of the empirical specification suggest that a significant trade-off may actually exist. Mogstad and Wiswall (2016) note that zero average effects may actually be hiding significant non-linear trade-offs as fertility levels change. Based on evidence from Norway, they suggest that strong positive effects of fertility may exist at low birth orders (eg when moving from 2 to 3), but that these will be substantially reversed at higher parity (eg when moving from 3 to 4 and above). A second contribution comes from work suggesting how one can account for potential non-randomness of twin birth, to correct the downward bias that appears to be captured in earlier work. Farbmacher et al. (2018), who adjust for use of artificial reproductive technologies. As discussed earlier, Bhalotra and Clarke (2019) show that twin conceptions are more likely to survive when mothers are healthier or in a more health-supportive environment during pregnancy. They show that ignoring this, as most studies have done, results in an upward bias (towards zero) in estimated fertility-investment trade-offs. Implementing partial correction for twin selection and using partial identification (bounds) estimates and large samples of microdata, Bhalotra and Clarke (2020) find significant trade-offs in education measures in linear IV models in both a US and developing country sample, and similarly document non-linear trade-offs similar in nature to those described in Mogstad and Wiswall (2016). Finally, as also indicated earlier, other work considers that trade-offs may be obscured if parents react to children's stocks, potentially compensating or reinforcing based on their endowments, with implications for observed trade-offs if focusing only on certain sub-samples with a family. Rosenzweig and Zhang (2009) note that as twins generally have lower birth-weights than non-twins, and as IV models often generally focus on *pre-twin* births only, this could lead to attenuated estimates of a trade-off if parents reinforce initial endowments. They find evidence suggestive of this in a survey in rural China, with results consistent with a significant trade-off in educational and health measures if focusing on averages over all children in the family.

Depending upon available data and the particular research question, previous studies use either parental investment or child outcomes as a measure of child quality. Estimates based on census data or registry data, such as those provided by Black et al. (2005); Angrist et al. (2010); Mogstad and Wiswall (2016); Åslund and Grönqvist (2010); Li et al. (2008) frequently by necessity focus on outcome measures such as completed education, grade completion, or salaries during adulthood. Studies using principally survey-based data, such as Bhalotra and Clarke (2020); Rosenzweig and Zhang (2009) can often additionally examine measures to capture parental investment behaviour, such as breastfeeding duration, time spent reading, and so forth.

While most papers in this literature focus on educational outcomes or investments as principal measures of children's human capital, there are a number of results which focus on other human capital measures. Children's health stocks have been considered, for example by Bhalotra and Clarke (2020) who observe health during childhood. A more long term measure is that of Baranowska et al. (2017), who find relatively little evidence to suggest that marginal fertility movements owing to twins affect mortality of men or women in Sweden. A quite a different measure is considered by Kolk (2015), who considers *intergenerational* transmission in fertility, studying how twins shape sibling fertility decisions, and finding that larger families, owing to twins, do indeed spillover into childbearing of twin siblings in future generations.

A brief summary of a range of other papers in this line, the context studied, the outcome measures considered, and their data sources is described in Table 2.

Table 2: Twin Instruments and Trade-offs

Authors	Context	Outcome(s)	Main Result
Panel A: Fertility and Parental Labour Market Outcomes			
Bronars and Grogger (1994)	1970-80 US Data.	Fertility and marital decisions, education, labour participation and earnings.	Large short-term effects of unplanned births among unwed mothers, but not among married mothers.
Angrist and Evans (1998)	Data from the US Census	The effect of childbearing on labour supply.	Larger effects for poor and less educated women. No effects on husbands.
Jacobsen et al. (1999)	Public Use Microdata Sample (PUMS) from US Census, between 1970-80.	The impact of an unplanned child on married women's labour supply and earnings.	Declining fertility explains between 6 and 13 percent of the increase in married women's labour supply between 1970 and 1980. Short-term effects become weaker and less persistent.
Cáceres-Delpiano (2006)	Data from the 1980 US Census (5% Public use micro sample).	Attendance at a private school, labour participation of the mother and probability of divorce.	Negative effects are found on private school attendance, the mother's probability of participating in the mother's labour market, and an increased probability of divorce.
Vere (2011)	US Census data from 1980, 1990, and 2000.	Causal effects of second and third children on women's and couples' labour supply.	For single women, the causal effect of fertility has declined significantly over time. Couples, however, have become more specialised along traditional lines, with married men tending to increase labour earnings rather than reduce hours worked.

Continued on next page

Table 2 – continued from previous page

Authors	Context	Outcome(s)	Main Result
Silles (2016)	General Household Survey for Great Britain between 1986 and 2009.	Causal effects of family size on women's labour market outcomes (labour supply and earnings.)	For women with children under 13 years of age, negative effects are found for both variables. Among women with children older than 13 years, there is no evidence of a causal effect.
Braakmann and Wildman (2016)	3 waves of the Millennium Cohort Study (MCS), which tracks a random sample of children (and their families) born in the UK in 2000-2001.	The impact of family size on labour market outcomes using the popular twin-birth instrument	Multiple births might still be a reasonable instrument for family size in labour supply regressions, even in countries and time periods where fertility treatments are common. The effects of the multiple-birth induced variation in family size depend on the time passed since the multiple birth.
Zhang (2017)	Taiwan Population and Housing Census from 2000	Causal effect of fertility on the labour supply of married women.	An additional child reduces female employment by 10.5 p.p. for those who have at least one birth. The effects decrease with increasing number of children, disappearing with 3 or more
Aaronson et al. (2020)	Compiled data set of 429 censuses and surveys, representing 101 countries and 46.9 million mothers.	The evolving impact of childbearing on the work activity of mothers between 1787 and 2014.	The effect of fertility on labour supply is small at low income levels and negative and substantially larger at higher income levels. As income increases, people face a higher cost of time to care for children, but also experience higher income (the former would dominate the latter)

Continued on next page

Table 2 – continued from previous page

Authors	Context	Outcome(s)	Main Result
Panel B: Fertility and Child Outcomes			
Black et al. (2005)	Norway data, all people aged 16-74 during 1986 to 2000 interval.	Family size and birth order over child educational attainment.	Higher birth order has a significant and large negative effect on children's education. No effect of family size.
Angrist et al. (2010)	Data from 1995 and 1983 Israeli censuses	third and higher-parity births on first and second-born schooling, labour status/earnings, marital status and fertility.	No significant evidence of a quantity-quality trade-off.
Cáceres-Delpiano (2006)	Data from the 1980 US Census	Number of children on child investment and well being when twins are born in a large family.	Significant reductions in private schooling in older siblings, mother's labour participation and increases parents divorce.
Marteleto and de Souza (2012)	Brazil 1977–2009: important social and demographic change	Association between family size and children's education.	Positive effect when development and fertility are high; the effect disappears otherwise, and when education expand.
Li et al. (2008)	Data from the 1990 Chinese Population Census	Family size on child educational attainment.	Negative effect which is more evident in rural China (poor public education)
Rosenzweig and Zhang (2009)	Data from the Chinese Child Twins Survey. Collected in late 2002 and early 2003.	Child quality outcomes (Schooling progess, College enrolment, grades in school, assessed health of all children in the family).	At least in one area of China, an exogenous extra child at parity one or at parity two birth-weight effects, decreases the schooling progress, the expected college enrolment in school and the assessed health of all children in the family.

Continued on next page

Table 2 – continued from previous page

Authors	Context	Outcome(s)	Main Result
Dayioglu et al. (2009)	Turkish Demographic and Health Survey (DHS) from 1998.	The effects of sibship size, birth order and sibling sex composition on children's school enrollment in urban Turkey.	They find no causal effects of sibship size on school enrollment. There is evidence of a parabolic impact of birth order where children born in the middle fare worse. The effects are larger for poorer households.
Åslund and Grönqvist (2010)	Swedish population born in 1972–87	Family size over intermediate and long-run outcomes: educational and labour market.	No causal effect on long-term outcomes (years of schooling and earnings). Negative impact on intermediate ones (grades).
Mogstad and Wiswall (2016)	Administrative registers from Statistics Norway. Period between 1986–2000	Relationship between education of children and number of siblings.	Important role of the linearity restriction in masking the family size effects. Using an unrestricted model the authors find nonzero effects of family size both for estimates with OLS and for IV.
Braakmann and Wildman (2014)	British Millennium Cohort Study. Random sample of children (and their families) born during late 2000 and 2001 in the UK.	The rise of fertility treatments as a threat to the commonly used multiple birth instrument for family size	The authors find differences, both in pre-pregnancy characteristics and outcomes, between women with and without fertility treatments. Conditional on having undergone fertility treatment, the birth of twins or triplets appears to be a random event.

Continued on next page

Table 2 – continued from previous page

Authors	Context	Outcome(s)	Main Result
Farbmacher et al. (2018)	Swedish administrative data between 1987–2006 & US Census data from 1980.	The effect of fertility on labour earnings using classical instrument and new approach.	The classic twin instrument underestimates the negative effect of fertility on labour income. People with high incomes are more likely to delay childbearing and therefore have a higher risk of having dizygotic twins.
Bhalotra and Clarke (2020)	US Data and 76 Developing countries	School Z-score, up to 16 years	Significant reductions in schooling for marginal birth at 2 and 3. 0.04-0.15 years in US, 0.06-0.5 grades, developing world.
Panel C: Fertility and Other Outcomes			
Kolk (2015)	Swedish register data	The effect of an exogenous increase of number of siblings on adult fertility.	Fertility is intergenerationally transmitted by shared preferences between parents, not by the size of the upbringing family.
Baranowska et al. (2017)	Swedish population born between 1938-72	Causal effect of growing up in a large family on mortality.	No substantive impact on adulthood mortality, either for men or for women.

4.2 Twin Pair Fixed Effects

Within-twin estimates of the return to health A broad range of studies have examined the impact of differential health stocks at birth on outcomes through childhood, adulthood, and in to the next generation. These studies leverage differentials in some measure of health at birth, while maintaining fixed maternal and home environments to isolate effects of early life health throughout later life. Generally this is captured by birth weight – necessarily owing to changes in fetal growth rate given that gestational length is virtually identical within twin pairs. However, as discussed below, other measures have also been used, such as intra-uterine growth rates, which is birth weight divided by gestational length, though given fixed gestational lengths, these measures capture largely similar underlying differentials.

Early examples of such models based on birth weight are from Almond et al. (2005) and Black et al. (2007), and based on intra-uterine growth are from Behrman and Rosenzweig (2004). Almond et al. (2005), based on the sample of twins born in the US between 1983-2000, find that birth weight has important returns in the short term. From within-twin models, they estimate that a one standard deviation in birth weight results in 8% of a standard deviation decline in hospitalization costs at birth, 5% of a standard deviation decline in birth weight, and 3% of a standard deviation increase in the 5 minute APGAR score. Behrman and Rosenzweig (2004) consider longer term outcomes, using a sample of female twins from the Minnesota Twin Registry to examine the impact of size at birth (birth weight/gestation) on adult outcomes capturing schooling, height, and wages, and for a subset, on impacts on birth weight of their own children. They find large impacts, for example within-twin pairs, a 1 pound increase in birth weight is estimated to increase adult earnings by 7%, and schooling by one third of a year, but do not find substantial impacts on adult BMI, or on transmission of birth weight to the next generation of children. Black et al. (2007) use register data from Norway covering birth cohorts from 1967-1981, and examine both short and long-run outcomes. They find significant impacts on short run (eg infant mortality) and long run (eg height, wages, IQ, and education) outcomes. Black et al. (2007), in their section II.A, additionally provide a nice discussion of why birth weight differs between twins, which is frequently referenced in subsequent literature. Among other noteworthy results, they find that a 10% increase in birth weight results in a 1 percentage point (pp) increase in the likelihood of completing high-school, a 3pp increase in the likelihood of working, and 1% increase in wage income.

More recent studies have documented results in a diverse range of settings, with outcomes which can be broadly defined as (a) early life or later life health outcomes, (b) educational outcomes later in life, (c) labour market or socioeconomic status, and (d) inter-generational impacts. Frequently, studies document impacts of differences in health endowments at birth over various of these outcome groups in a single setting.

Beyond the results of Black et al. (2007); Behrman and Rosenzweig (2004); Almond et al. (2005), all of which examined health outcomes at certain points in life, a number of other studies document impacts on health outcomes. Oreopoulos et al. (2008) documents mortality declines both within 1 year, and until at least 17 years of age using a sample of Canadian administrative data. Conley et al. (2003) document substantial impacts of greater birth weight on reduced infant mortality. Royer (2009), based on data from birth certificates in California, notes impacts on health in a number of dimensions, including infant mortality and complications during pregnancy later in life if the individual becomes a mother. The aforementioned results all come from high income countries, though McGovern (2019) finds that results are if anything larger in developing countries. Pooling data from the Demographic and Health Surveys covering 66 countries, McGovern (2019) finds large effects of marginal increases in birth weight, particularly low in the distribution of birth weight, on infant mortality, stunting, wasting and a number of other measures of health.

On education, results from a broad range of settings point to positive and substantial returns to early life health when considering measures of educational attainment later in life. This includes results from USA (Figlio et al., 2014; Royer, 2009), Chile (Torche and Echevarría, 2011; Bharadwaj et al., 2018), Canada Oreopoulos et al. (2008), Australia Miller et al. (2005) and Norway Black et al. (2007). For example, recent evidence, from Figlio et al. (2014) on a large sample of twins (~15,000) from administrative data covering all births between 1992-2002 in Florida estimates large results of the returns to early life health on schooling outcomes. They find that a 1 kilogram increase in birth weight in twin FE models is estimated to increase scores on standardized tests by 18.7% of a standard deviation. They note that these results are similar in all types of demographic and socioeconomic groups, including when stratifying by immigrant status, educational level, race and ethnicity, child's gender, and mother's age at child birth, among others. They also find quite stable results when estimating returns to marginal changes in birth weight across the birth weight distribution, and when considering impacts by the degree of birth weight discordance within twin pairs.

A small number of studies suggest results in other socioeconomic measures. For example, Oreopoulos et al. (2008) finds that greater birth weight is associated with reduced dependence on the social safety net, both in terms of welfare take-up in Canada, and the number of months for which an individual receives welfare payments. Additionally, positive labour market effects of indicators of health at birth are found in Black et al. (2007) (employment and earnings in Norway), and Rosenzweig and Zhang (2009) (wages in China).

Finally, a small number of papers have touched on transfers of health stocks at birth into outcomes of the next generation, when working with samples of female twins. In early work, Behrman and Rosenzweig (2004) have one set of results examining this, finding a lack of transmission when the focusing on within-twin mother differentials in intra-uterine growth rate. However, more recently, based on a larger sample of births from California which children and mother birth certificates, Royer (2009) finds evidence of substantial inter-generational spillovers in health at birth. Estimating within-twin models, she finds that an additional 1 kg of birth weight increases child birth weight by around 70 grams, compared to an estimate of 180 grams when using standard OLS. She also finds that this result is non-linear, with returns to birth weight in the second generation being slightly larger for children born at less the 2,500 grams in the first generation, when using twins FE models.

Within-twin estimates of the return to education Perhaps the most studied question in within-twin models is that of returns to education, which has a long history, with published studies now spanning 6 decades. Early papers of Taubman (1976b,a); Behrman and Taubman (1976, 1989) all made contributions in defining models, discussing how twins can be used to decompose earnings into genetic and environmental determinants, and in some cases, estimating returns to education and presenting estimates from data on US twins. Below principal messages and empirical lessons which can be taken away from within-twin studies seeking to isolate returns from education are discussed, which is the principal strand taken forward from this literature.

This review focuses on within-twin models which seek to isolate the impacts of some variable in which differences are observed between twin pairs, on some other variable measured across twins. However, twins are also used in another way in a related series of results, which seek to understand the heritability of outcomes, by examining correlations between particular measures across twins, who have shared genes, and across the entire population, where genes vary. Such models have been considered very early in the literature (Goldberger, 1979), and are frequently reported alongside twin fixed effect models Taubman (1976b); Behrman and Taubman (1976); Bjorklund et al. (2003). Such a method has also been taken up recently in understanding the degree to which savings behaviours could owe to genetic pre-disposition (Cronqvist and Siegel, 2015), risk taking (Cesarini et al., 2009a) and over-confidence (Cesarini et al., 2009b). In this section, our focus is on within-twin models, rather than models of genetic determinants of behaviours. A survey based discussion of the divergence between within-twin style models and models which use twins to understand heritability and behavioural genetics is available as Kohler et al. (2011).

The earliest models, discussed above, when examining returns to education, all focus on education and salary. Taubman (1976b), based on a sample of US male twins and non-twins aged around 50 years old estimates that an additional year of schooling in within-twins models increases salaries by around 3% when studying identical, or MZ, twins, (compared to around 5% when studying non-identical, or DZ, twins, and around 8% when simply estimating OLS). Behrman and Taubman (1976) document relatively similar results on returns to earnings, additionally finding positive returns to education when considering a measure of socioeconomic status. It is noteworthy that returns in these early studies are quite low. Following the influential critique of Griliches (1979) that such low returns may owe to measurement error in reported years of education, (Ashenfelter and Krueger, 1994) propose a widely-adopted methodology of instrumenting reported twin educational differentials with differentials calculated by reports from the other twin. Based on these models, they estimate substantially higher returns to education, at between 12-16% per year of education. The question of education, and its impacts on earnings, has been taken up by a large stream of papers, broadly following the method proposed by Ashenfelter and Krueger (1994). Specifically, related questions have been studied in Sweden (Isacsson, 1999, 2004), suggestion returns of around 5%, the UK (Bonjour et al., 2003), suggesting returns of around 8%, the US, including data from the Twinsburg twin festival in particular (Ashenfelter and Rouse, 1998; Rouse, 1999), suggesting returns of 9-10%, and Australia (Miller et al., 2006, 1997, 1995), suggesting 5-7%.

A range of papers use similar designs, however rather than focusing on labour market returns to education, focus on how changes in education impact health behaviours and health stocks later in life. This includes measures such as survival, health investment behaviours, anthropometric measures such as BMI, and self reported health behaviours. Studies on survival include Lundborg et al. (2016) (Sweden), Behrman et al. (2011) (Denmark), Behrman et al. (2015) (China), and Savelyev et al. (2022) (USA). The results of Lundborg et al. (2016) point to strong effects of education on mortality, both for men and women, which is additionally found by Savelyev et al. (2022) in survey data from the USA. Results from China described in Behrman et al. (2015) also point to impacts of education on determinants of mortality such as smoking. However, these results are not observed universally: a study by Behrman et al. (2011) based in Denmark finds no significant results with Danish data. A number of other papers present results based on other health measures. Amin et al. (2013) find mixed evidence of a result using female twins in the UK, noting an impact on BMI, but no significant results observed when examening smoking, alcohol consumption, or exercise frequency. Lundborg (2013), with data from the US however does find that educational attainment – at least when measured as highschool completion, does have impacts on exercise behaviour, self reported measures, and chronic health outcomes. Böckerman and Maczulskij (2016), using Finnish data, find returns to highschool and tertiary education on outcomes such as BMI, medication usage, and exercise frequency. Finally, results fom Amin et al. (2014a, 2015), partially revisted with a larger sample in Savelyev et al. (2022) point to a number of health returns to education, at least when considering self reported health (Savelyev et al., 2022).

Finally, note that there are a number of papers which extend the within-twin design to consider other outcomes. These include Behrman and Rosenzweig (2002); Amin et al. (2011) who find inter-generational links in schooling between parents and their children based on within-twin models, Amin and Behrman (2011), who estimate fertility reductions and marriage market implications of additional education, Behrman et al. (1994) who also discuss marriage markets and endowments. This twin design has also been used when considering returns to variables which are not education (or health endowments at birth), or are quite specific measures of education. This is the case with a study of Behrman et al. (1996) who seeks to understand returns to particular types of university instruction, Webbink et al. (2011) who seeks to isolate impacts of teenage child bearing based on within-twin models, and Webbink et al. (2013) who studies impacts of criminality on educational attainment. Further details, including samples and outcome measures in these studies, along with all those discussed in this subsection are provided in Table 3.

Table 3: Twin Differences and Later-life Returns

Authors	Context	Outcome(s)	Main Result
Panel A: Health at Birth and Later Life Outcomes			
Behrman and Rosenzweig (2004)	Minnesota Twin Registry Data	Intrauterine nutrient intake over adult health and earnings.	Birthweight has a positive effect on adult schooling attainment, labour earnings and height, but no effect on adult body mass.
Almond et al. (2005)	U.S. Administrative data between 1983-97 with information of Birth and infant death & Healthcare Cost.	The hospital costs, health at birth, and infant mortality rates between heavier and lighter infants from all twin pairs born in the United States.	They find substantial lower effects of LBW than previously expected. Existing estimates exaggerate the true costs and consequences of LBW (between 4-20 times). Different approaches to prevent LBW have different health and cost consequences.
Black et al. (2007)	Norwegian births over the period 1967–1997	Adult outcomes (height, BMI, IQ, education, earnings, and birth weight of the first-born child)	Using within-twin variation in birth weight they found that birth weight does matter with short and long run effects.
Miller et al. (2005)	Australian Twin Register Data.	Further analyses of the links between birth weight, schooling and earnings.	Birth weight isn't related to schooling levels, it plays only a minor role in the determination of earnings.

Continued on next page

Table 3 – continued from previous page

Authors	Context	Outcome(s)	Main Result
Conley et al. (2006)	U.S. Live Birth and Fetal Death files for 1995–1997 with the Linked Live Birth/Infant Death Cohort Data Sets for 1995–1997.	Associations between birth weight and infant mortality vary across identical and fraternal twins, gestational age, and time.	Shared aspects of the pregnancy environment (e.g., maternal behavior, exposure to toxins) and/or shared genes explain large portions of the association between birth weight and mortality. Very low birth weight and mortality within twin pairs, however, suggests that prenatal environmental factors and genetic variation may operate in varying ways depending primarily on gestational age.
Oreopoulos et al. (2008)	Adminstrative data from Manitoba, Canada between 1978-85.	The short-, medium-, and long-term consequences of poor infant health.	Infant health is a strong predictor of educational and labour force outcomes. Infant health is found to predict both high school completion and social assistance (welfare) takeup and length.
Royer (2009)	1960-82 California birth and Childhood Longitudinal Study Birth Data.	Correlations between birth weight (proxy for fetal nutrient intake) and adult outcomes.	Weight is related to educational attainment, later pregnancy complications, and the birth weight of the next generation.
Figlio et al. (2014)	All twins born in Florida between 1992 and 2002 and educated in a Florida public school afterward.	The effects of poor neonatal health on child cognitive outcomes.	Twins with higher birth weights enter school with a cognitive advantage that appears to remain stable through the elementary and middle school years.

Continued on next page

Table 3 – continued from previous page

Authors	Context	Outcome(s)	Main Result
McGovern (2019)	Demographic and Health Surveys (DHS). A series of cross-sectional surveys of 90 countries since 1984.	The effects of birth weight on child health and growth.	The author found a consistent effect of birth weight on the risk of mortality, growth retardation, wasting, and cough, with some evidence of fever, diarrhea, and anemia. Boundary analysis indicates that coefficients may be substantially underestimated due to mortality selection
Torche and Echevarría (2011)	Birth registry information with standardized test scores in Chile.	Using a Twin fixed-effects models are used to estimate the causal effect of intra-uterine growth on test scores.	A 400-g increase in birthweight results in a 15% standard deviation increase in Math scores. The effect is larger among (estimated) monozygotic than dizygotic pairs. This effect interacts with family socioeconomic status (SES).
Bharadwaj et al. (2018)	Data on all births from 1992 to 2002 merged with schooling records for the entire education system from Chile.	Relationship between health at birth, academic outcomes and the potential role for parental investments.	They conclude, using a human capital accumulation model, parental investments are compensatory with respect to initial health, but not between twins. Twin fixed-effects models estimate a persistent effect of birth weight on academic performance, while OLS and sibling fixed-effects models find that this relationship decreases over time.
Panel B: Education and Labour Market Outcomes			
Behrman and Taubman (1976)	Twin sample born between 1917 and 1927	Years of education, socioeconomic status of occupation and earnings.	The importance of biases introduced by ignoring genetic and environmental factors.

Continued on next page

Table 3 – continued from previous page

Authors	Context	Outcome(s)	Main Result
Taubman (1976b)	NAS-NRC Twin Sample.	Schooling, genetic endowments and family environment over earnings.	Not controlling for genetics and family environment may cause a bias of 2/3 of the noncontrolled coefficient.
Ashenfelter and Krueger (1994)	Interview on the Twins Festival in Twinsburg.	Contrast the wages of MZ twins with different schooling levels.	An additional year of schooling increases wages by 12-16%.
Ashenfelter and Rouse (1998)	Interviews in 3 Twinsburg Twins Festivals.	Return to schooling for genetically identical individuals.	Average return of 9%, slightly higher returns for less able individuals.
Rouse (1999)	Interviews in 4 Twinsburg Twins Festivals.	Contribution of genetic ability to the observed return to schooling.	The return to schooling among MZ twins is 10% per year of schooling completed.
Bonjour et al. (2003)	U.K. Adult Twin Registry (MZ twins).	Estimate returns to education and ability differences within twin pairs.	Private female return of 7.7. Ability gap within twin are less than between families.
Miller et al. (1997)	Australian Twins Registry Data.	Family background on the schooling and income link, by gender.	Schooling return is greater for females. Family background is greater for males.
Miller et al. (2006)	Australian Twin Registry. Panel of twin born between 1964-1971.	Variations in annual earnings. Returns to schooling.	The economic return to schooling in Australia among a sample of young twins is between 5 and 7 percent when account is taken of genetic and family effects using either fixed-effects models or the selection effects model. Estimates of this order of magnitude are similar to those obtained using conventional models.
Isacsson (1999)	Swedish Twin Registry, between 1886-1967.	Examined whether conventional estimates of the return to schooling in Sweden are biased because ability is omitted from the earnings–schooling relationship.	The measurement error-adjusted estimate of return to education in the MZ twin sample indicated a slight ability bias of approximately 10% in conventional estimates of return to education.

Continued on next page

Table 3 – continued from previous page

Authors	Context	Outcome(s)	Main Result
Isacsson (2004)	Swedish Twin Registry	Earnings premiums associated with different educational levels.	The estimation strategy implies that ability bias can be investigated separately in different parts of the educational distribution.
Panel C: Education and Non-Labour Market Outcomes			
Amin and Behrman (2011)	Data on MZ female twins from the Minesota Twin Registry.	Schooling over completed fertility, chances of being childless and timing of first births.	More schooling causes women to have fewer children and delay marriage and childbearing.
Behrman et al. (2011)	Danish population-level and twin registries.	Associations and causal impact of schooling on health and mortality	Negative link between schooling and hospitalization/mortality. No causal effects.
Behrman and Rosenzweig (2002)	Mail survey of twins from the Minnesota Twin Registry.	Intergenerational effects of increasing the level of maternal and paternal schooling on child schooling.	An increase in the schooling of women would not have beneficial effects in terms of the schooling of children.
Amin et al. (2013)	Department of Twin Research and Genetic Epidemiology, King's College London.	Schooling on health outcomes (obesity and physical health) and health-related behaviors (smoking, alcohol consumption and exercise)	No causal effect for physical health, alcohol consumption and exercise. Schooling reduces body mass index for women
Behrman et al. (2015)	Chinese Adults Twins Survey.	Effects of own schooling on one's own health outcomes and health-related behaviors and spouse health outcomes.	Pro-health effects for own health (mental), health-related behaviors (smoking) and spouse health (overweight).
Amin et al. (2011)	Swedish Twin Registry Data.	Importance of parents schooling for children's schooling.	Parents schooling matters. The effect is entirely driven by the impact of mother's schooling on daughter's schooling.

Continued on next page

Table 3 – continued from previous page

Authors	Context	Outcome(s)	Main Result
Lundborg (2013)	First wave of the MIDUS survey in USA. This data is collected in 1995.	Health returns to schooling, using a twin design	People with greater schooling are significantly healthier, as measured through self-reported health and chronic conditions, and perform exercise more often. These results were based on a twin-differences design, netting out the influence of genetics and family endowments.
Amin et al. (2014a)	3 large US MZ twins datasets from different US states.	Health indicators.	Small or non-significant impacts of schooling on health were found after controlling for unobserved confounders.
Panel D: Other Considerations			
Lundborg et al. (2016)	Swedish male twins born in 1950-79.	Mechanisms behind the relationship between adolescent health and educational attainment.	Teenage health does not have an influence on schooling levels. Raw correlations driven by genes and twin-pair specific environmental factors.
Behrman et al. (1994)	NAS-NRC Twin Sample - Minnesota Twin Registry.	Effect of individual endowments in the household and in the marriage market.	Individual-specific earnings endowments of men and their wives' schooling are negatively associated.
Webbink et al. (2011)	2 cohorts of twins and their relatives of the Australian Twin Registry.	Difference in educational attainment between teen mothers and their identical twin sisters.	The negative effect of teenage childbearing on educational attainment appears to be small. The difference in educational attainment between teen mothers and their sister is reduced to zero when restricting the sample to identical twins.

Continued on next page

Table 3 – continued from previous page

Authors	Context	Outcome(s)	Main Result
Webbink et al. (2013)	Australian Twin Registry. Compiled by 2 surveys between 1989-1990 and 1996-2000.	Using fixed effects to control for unobservables, they seek to see if crime reduces investment in human capital or if education reduces criminal activity.	Within twin pairs it is found that early arrests reduce educational attainment by 0.7 to 0.9 years and reduce the probability of completing upper secondary school by 20 to 23 percentage points.

5 Summary

Twin birth has the flavour of a miracle. At first glance it seems plausible that the occurrence of twin birth delivers an unexpected additional child and, thus, a shock to fertility. However, it is not always so. There are three sorts of reasons why. The first is hormonal, or biological-factors like age are associated with an increase in the follicle stimulating hormone, which elevates the chances of twin birth. This matters because older women may invest differently, whether in their children or in their careers, but age is typically observable and can be held constant in the analysis. The second is a feature of a technological procedure – more often than not, IVF involves a double embryo transfer designed to increase the chances of pregnancy success but, as a fallout, it also significantly increases the chances of twin birth. This matters because there is selection into IVF on traits like education of the parents which are predictive of economic outcomes for children and parents. However, where IVF use is observable, it can be controlled for. The third factor pertains to the chances that a twin conception, which may be random, survives to the status of live birth. These chances are dependent on the health and the health environment of the mother. Twins are more likely to miscarry than singletons and more so in less healthy mothers. This is much harder to fully purge with controls because the intrinsic health of the mother, her health-related behaviours, and features of the natural and the medical environment of the mother. However, the instrument tends to have sufficient power and, as long as some measures of maternal health are available, inference can proceed with partial identification methods.

Even if twin birth is conditionally quasi-randomly assigned, the twin instrument strategy faces other challenges. Inference is challenged by power because the share of twins in the population, at 2 to 3 percent is small, making this an area in which administrative data containing the entire population are particularly useful. With a view to allowing that fertility following twin birth is endogenous, the twin instrument approach looks exclusively at outcomes for older siblings. It cannot illuminate impacts of twin birth on younger siblings. Twin birth, by its nature, does not illuminate impacts of extensive margin fertility, at least insofar as this involves the birth of one child. A reduced form specification can however leverage twin birth at first parity, comparing parental outcomes between parents who have a twin vs a singleton at birth. Alternative approaches to studying impacts of the birth of a first child on women's earnings are also discussed. These include an event study design, idiosyncratic success in IVF treatments, and the implementation of mandates for single embryo transfer in

IVF.

The literature has highlighted the importance of generalizations to the simple linear model that has been used in the first generation of twin instrument studies. Impacts of the marginal birth – incident through a twin occurrence – are likely to depend upon the birth order at which the twin occurs. When the outcome is for an older sibling, the impact may depend upon the birth interval between the older sibling and the twins, and the age at which the outcome is measured. When the outcome is for a parent, the impact may depend upon the age at which the parent has the twin birth, as this will indicate where in their career process they are. Average causal response functions are useful in depicting the variation in fertility generated by the twin instrument and, in particular how the weighting function over parities induced by the instrument changes as controls are added.

The between-twin strategy that seeks to purge genetic influences on outcomes using monozygotic twins also faces identification challenges. Even if twins are genetically identical at birth, they may have different early exposures that are correlated with differences in their endowments (e.g. birth weight) and that impact the outcomes of interest. In a similar vein, parents or schools may reinforce or compensate any initial differences in endowments and these behaviours may directly influence the outcome. While the twin instrument strategy suffers from noisy estimates because twins are a relatively small share of all births, the between-twin strategy suffers from noisy estimates because any differences between twins tend to be small. As highlighted for the twin instrument, so also for between-twin strategies, we may expect that the relevant relationships are non-linear. There is then a question of external validity, led by where in the distribution the difference between twin endowments lies.

On a positive note, as discussed in the paper, there are recent innovations to the twin identification approaches that help the researcher acknowledge and possibly overcome some of these challenges.

References

D. Aaronson, R. Dehejia, A. Jordan, C. Pop-Eleches, C. Samii, and K. Schulze. The Effect of Fertility on Mothers' Labor Supply over the Last Two Centuries. *The Economic Journal*, 131(633):1–32, 08 2020. ISSN 0013-0133. doi: 10.1093/ej/ueaa100. URL https://doi.org/10.1093/ej/ueaa100.

D. Almond and B. Mazumder. Fetal origins and parental responses. *Annual Review of Economics*, 5(1):37–56, 2013. doi: 10.1146/annurev-economics-082912-110145.

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, L. Edlund, and M. Palme. Chernobyl's subclinical legacy: prenatal exposure to radioactive fallout and school outcomes in Sweden. *Quarterly Journal of Economics*, 124 (4):1729–1772, 2009.

V. Amin and J. Behrman. Do more-schooled women have fewer children and delay childbearing? evidence from a sample of u.s. twins. *Journal of Population Economics*, 27, 12 2011.

V. Amin, P. Lundborg, and D.-O. Rooth. Mothers do matter: New evidence on the effect of parents' schooling on children's schooling using swedish twin data. 09 2011.

V. Amin, J. Behrman, and T. Spector. Does more schooling improve health outcomes and health related behaviors? evidence from u.k. twins. *Economics of Education Review*, 35: 134–148, 08 2013.

V. Amin, J. Behrman, and H.-P. Kohler. Schooling has smaller or insignificant effects on adult health in the us than suggested by cross-sectional associations: New estimates using relatively large samples of identical twins. *Social science & medicine (1982)*, 127, 07 2014a.

V. Amin, J. Behrman, H.-P. Kohler, Y. Xiong, and J. Zhang. Causal inferences: Identical twins help and clarity about necessary assumptions is critical. *Social Science & Medicine*, 127, 12 2014b.

V. Amin, J. R. Behrman, and H.-P. Kohler. Schooling has smaller or insignificant effects on adult health in the us than suggested by cross-sectional associations: New estimates using relatively large samples of identical twins. *Social Science & Medicine*, 127:181–189, 2015. ISSN 0277-9536. doi: https://doi.org/10.1016/j.socscimed.2014.07.065. URL https://www.sciencedirect.com/science/article/pii/S027795361400505X. Special Issue: Educational Attainment and Adult Health: Contextualizing Causality.

M. Andresen and E. Nix. What causes the child penalty? evidence from adopting and same-sex couples. *Journal of Labor Economics*, 0(ja):null, 2022. doi: 10.1086/718565. URL https://doi.org/10.1086/718565.

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. Schooling and Labor Market Consequences of the 1970 State Abortion Reforms. NBER Working Papers 5406, National Bureau of Economic Research, Inc, Jan. 1996.

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.

J. D. Angrist and G. W. Imbens. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442, 1995. doi: 10.1080/01621459.1995.10476535.

O. Ashenfelter and A. B. Krueger. Estimates of the economic returns to schooling from a new sample of twins. *American Economic Review*, 84(5):1157–73, December 1994.

O. Ashenfelter and C. Rouse. Income, schooling, and ability: Evidence from a new sample of identical twins. *The Quarterly Journal of Economics*, 113(1):253–284, 1998.

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

A. Baranowska, K. Barclay, and M. Kolk. The effect of number of siblings on adult mortality: Evidence from swedish registers for cohorts born between 1938 and 1972. *Population Studies*, 71, 03 2017.

S. Beemsterboer, R. Homburg, N. Gorter, R. Schats, P. Hompes, and C. Lambalk. The paradox of declining fertility but increasing twinning rates with advancing maternal age. *Human Reproduction*, 21(6):1531–1532, 02 2006. ISSN 0268-1161. doi: 10.1093/humrep/del009. URL https://doi.org/10.1093/humrep/del009.

J. Behrman. Twin studies in economics. *The Oxford Handbook of Economics and Human Biology*, 1:21, November 2015.

J. Behrman and P. Taubman. Intergenerational transmission of income and wealth. *The American Economic Review*, 66(2):436–440, 1976.

J. Behrman, H.-P. Kohler, V. Jensen, D. Pedersen, I. Petersen, P. Bingley, and K. Christensen. Does More Schooling Reduce Hospitalization and Delay Mortality? New Evidence Based on Danish Twins. *Demography*, 48(4):1347–1375, November 2011.

J. R. Behrman and M. R. Rosenzweig. Ability biases in schooling returns and twins: a test and new estimates. *Economics of Education Review*, 18(2):159–167, April 1999.

J. R. Behrman and M. R. Rosenzweig. Does increasing women's schooling raise the schooling of the next generation? *American Economic Review*, 92(1):323–334, March 2002.

J. R. Behrman and M. R. Rosenzweig. Returns to Birthweight. *The Review of Economics and Statistics*, 86(2):586–601, May 2004.

J. R. Behrman and P. Taubman. Is schooling "mostly in the genes"? nature-nurture decomposition using data on relatives. *Journal of Political Economy*, 97(6):1425–1446, 1989.

J. R. Behrman, M. R. Rosenzweig, and P. Taubman. Endowments and the allocation of schooling in the family and in the marriage market: The twins experiment. *Journal of Political Economy*, 102(6):1131–74, December 1994.

J. R. Behrman, M. R. Rosenzweig, and P. Taubman. College choice and wages: Estimates using data on female twins. *The Review of Economics and Statistics*, 78(4):672–685, 1996.

J. R. Behrman, Y. Xiong, and J. Zhang. Cross-sectional schooling-health associations misrepresented causal schooling effects on adult health and health-related behaviors: Evidence from the chinese adults twins survey. *Social Science & Medicine*, 127:190–197, 2015. ISSN 0277-9536. doi: https://doi.org/10.1016/j.socscimed.2014.11.011. URL https://www.sciencedirect.com/science/article/pii/S0277953614007345. Special Issue: Educational Attainment and Adult Health: Contextualizing Causality.

V. Beral, P. Doyle, S. L. Tan, B. A. Mason, and S. Campbell. Outcome of pregnancies resulting from assisted conception. *British Medical Bulletin*, 46(3):753–68, 1990.

I. Berniell, L. Berniell, D. de la Mata, M. Edo, and M. Marchionni. Gender gaps in labor informality: The motherhood effect. *Journal of Development Economics*, 150:102599, 2021. ISSN 0304-3878.

S. Bhalotra and D. Clarke. Twin Birth and Maternal Condition. *The Review of Economics and Statistics*, 101(5):853–864, 12 2019.

S. Bhalotra and D. Clarke. The Twin Instrument: Fertility and Human Capital Investment. *Journal of the European Economic Association*, 12 2020. ISSN 1542-4766. doi: 10.1093/jeea/jvz058. URL https://doi.org/10.1093/jeea/jvz058. jvz058.

S. R. Bhalotra, D. Clarke, H. Mühlrad, and M. Palme. Health and Labor Market Impacts of Twin Birth: Evidence from a Swedish IVF Policy Mandate. IZA Discussion Papers 14990, Institute of Labor Economics (IZA), Jan. 2022. URL https://ideas.repec.org/p/iza/izadps/dp14990.html.

P. Bharadwaj, J. P. Eberhard, and C. A. Neilson. Health at birth, parental investments, and academic outcomes. *Journal of Labor Economics*, 36(2):349–394, 2018. doi: 10.1086/695616.

A. Bjorklund, M. Jäntti, and G. Solon. Influences of nature and nurture on earnings variation: A report on a study of various sibling types in sweden. *Unequal Chances Family Background and Econonomic Success*, 05 2003.

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.

S. E. Black, P. J. Devereux, and K. G. Salvanes. From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes. *The Quarterly Journal of Economics*, 122(1): 409–439, 2007.

S. E. Black, P. J. Devereux, and K. G. Salvanes. Does Grief Transfer across Generations? Bereavements during Pregnancy and Child Outcomes. *American Economic Journal: Applied Economics*, 8(1):193–223, January 2016.

D. Blanchflower and P. Elias. Ability, schooling and earnings: Are twins different? 1999.

J. Boardman and J. Fletcher. To cause or not to cause? that is the question, but identical twins might not have all of the answers. *Social science & medicine (1982)*, 127, 10 2014.

P. Böckerman and T. Maczulskij. The education-health nexus: Fact and fiction. *Social Science & Medicine*, 150:112–116, 2016. ISSN 0277-9536. doi: https://doi.org/10.1016/j.socscimed.2015.12.036. URL https://www.sciencedirect.com/science/article/pii/S0277953615302951.

D. Bonjour, L. F. Cherkas, J. E. Haskel, D. D. Hawkes, and T. D. Spector. Returns to Education: Evidence from U.K. Twins. *American Economic Review*, 93(5):1799–1812, December 2003.

J. Bound and G. Solon. Double trouble: on the value of twins-based estimation of the return to schooling. *Economics of Education Review*, 18(2):169–182, April 1999.

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.

N. Braakmann and J. Wildman. Reconsidering the effect of family size on labour supply: the twin problems of the twin birth instrument. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 179(4):1093–1115, 2016.

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.

M. G. Bulmer. *The Biology of Twinning in Man*. Oxford Clarendon Press, Oxford, UK, 1970.

A. Bütikofer, D. N. Figlio, K. Karbownik, C. W. Kuzawa, and K. G. Salvanes. Evidence that prenatal testosterone transfer from male twins reduces the fertility and socioeconomic success of their female co-twins. *Proceedings of the National Academy of Sciences*, 116 (14):6749–6753, 2019.

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

J. Cáceres-Delpiano. Can We Still Learn Something From the Relationship Between Fertility and Mother's Employment? Evidence From Developing Countries. *Demography*, 49(1): 151–174, February 2012.

D. Cesarini, C. T. Dawes, M. Johannesson, P. Lichtenstein, and B. Wallace. Genetic variation in preferences for giving and risk taking. *The Quarterly Journal of Economics*, 124(2): 809–842, 2009a.

D. Cesarini, P. Lichtenstein, M. Johannesson, and B. Wallace. Heritability of overconfidence. *Journal of the European Economic Association*, 7(2/3):617–627, 2009b.

K. Chay and M. Greenstone. The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *The Quarterly Journal of Economics*, 118(3):1121–1167, 2003.

D. Conley, K. Strully, and N. G. Bennet. A pound of flesh or just proxy? Using twin differences to estimate the effect of birth weight on life chances. NBER Working Papers 9901, National Bureau of Economic Research, Inc, Aug. 2003.

D. Conley, K. Strully, and N. Bennett. Twin differences in birth weight: The effects of genotype and prenatal environment on neonatal and post-neonatal mortality. *Economics and human biology*, 4:151–83, 07 2006.

H. Cronqvist and S. Siegel. The origins of savings behavior. *Journal of Political Economy*, 123(1):123–169, 2015.

J. Currie and D. Almond. Chapter 15 - Human capital development before age five. volume 4 of *Handbook of Labor Economics*, pages 1315–1486. Elsevier, 2011. doi: https://doi.org/10.1016/S0169-7218(11)02413-0. URL https://www.sciencedirect.com/science/article/pii/S0169721811024130.

M. Dayioglu, M. G. Kirdar, and A. Tansel. Impact of Sibship Size, Birth Order and Sex Composition on School Enrolment in Urban Turkey. *Oxford Bulletin of Economics and Statistics*, 71(3):399–426, 06 2009.

C. Ekechi. Addressing inequality in fertility treatment. *The Lancet*, 398(10301):645–646, 2021. doi: 10.1016/s0140-6736(21)01743-8.

H. Farbmacher, R. Guber, and J. Vikström. Increasing the credibility of the twin birth instrument. *Journal of Applied Econometrics*, 33(3):457–472, 2018. doi: https://doi.org/10.1002/jae.2616. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2616.

D. Figlio, J. Guryan, K. Karbownik, and J. Roth. The effects of poor neonatal health on children's cognitive development. *American Economic Review*, 104(12):3921–55, December 2014. doi: 10.1257/aer.104.12.3921. URL https://www.aeaweb.org/articles?id=10.1257/aer.104.12.3921.

E. Fitzsimons and B. Malde. Empirically Probing the Quantity-Quality Model. IFS Working Papers W10/20, Institute for Fiscal Studies, Sep 2010.

J. Gangadharan, J. Rosenbloom, J. Jacobson, and I. Pearre, James Wishart. The effects of child-bearing on married women's labor supply and earnings: Using twin births as a natural experiment. Working Paper 5647, National Bureau of Economic Research, July 1996. URL http://www.nber.org/papers/w5647.

C. D. Geyter. Single embryo transfer in all infertile couples treated with assisted reproduction produces excellent results and avoids multiple births. *Swiss Medical Weekly*, 2021. doi: 10.4414/smw.2021.20499.

A. S. Goldberger. Heritability. *Economica*, 46(184):327–347, 1979.

Z. Griliches. Sibling models and data in economics: Beginnings of a survey. *Journal of Political Economy*, 87(5):S37–64, October 1979.

J. G. Hall. Twinning. *The Lancet*, 362(9385):735–743, August 2003.

E. A. Hanushek. The Trade-Off between Child Quantity and Quality. *Journal of Political Economy*, 100(1):84–117, February 1992.

G. Isacsson. Estimates of the return to schooling in sweden from a large sample of twins. *Labour Economics*, 6(4):471–489, 1999.

G. Isacsson. Estimating the economic return to educational levels using data on twins. *Journal of Applied Econometrics*, 19:99–119, 01 2004.

J. P. Jacobsen, J. W. Pearce III, and J. L. Rosenbloom. The Effects of Childbearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment. *Journal of Human Resources*, 34(3):449–474, 1999.

H. Kleven, C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller. Child Penalties across Countries: Evidence and Explanations. *AEA Papers and Proceedings*, 109:122–26, May 2019a. doi: 10.1257/pandp.20191078. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20191078.

H. Kleven, C. Landais, and J. E. Søgaard. Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, 11(4):181–209, October 2019b. doi: 10.1257/app.20180010. URL http://www.aeaweb.org/articles?id=10.1257/app.20180010.

H.-P. Kohler, J. R. Behrman, and J. Schnittker. Social science methods for twins data: Integrating causality, endowments, and heritability. *Biodemography and Social Biology*, 57: 141 – 88, 2011.

M. Kolk. The causal effect of an additional sibling on completed fertility: An estimation of intergenerational fertility correlations by looking at siblings of twins. *Demographic Research*, 32:1409–1420, 2015.

H. Levine, N. Jørgensen, A. Martino-Andrade, J. Mendiola, D. Weksler-Derri, I. Mindlis, R. Pinotti, and S. H. Swan. Temporal trends in sperm count: a systematic review and meta-regression analysis. *Human Reproduction Update*, 23(6):646–659, 07 2017. ISSN 1355-4786. doi: 10.1093/humupd/dmx022. URL https://doi.org/10.1093/humupd/dmx022.

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.

P. Lundborg. The health returns to schooling—what can we learn from twins? *Journal of Population Economics*, 26(2):673–701, 2013.

P. Lundborg, A. Nilsson, and D.-O. Rooth. The health-schooling relationship: evidence from swedish twins. *Journal of Population Economics*, 29(4):1191–1215, 2016.

P. Lundborg, E. Plug, and A. W. Rasmussen. Can women have children and a career? iv evidence from ivf treatments. *American Economic Review*, 107(6):1611–37, June 2017. doi: 10.1257/aer.20141467. URL https://www.aeaweb.org/articles?id=10.1257/aer.20141467.

L. J. Marteleto and L. R. de Souza. The Changing Impact of Family Size on Adolescents' Schooling: Assessing the Exogenous Variation in Fertility Using Twins in Brazil. *Demography*, 49(4):1453–1477, July 2012.

B. Mazumder and Z. Seeskin. Breakfast skipping, extreme commutes and the sex composition at birth. *Biodemography and Social Biology*, 61(2):187–208, 2015.

M. E. McGovern. How much does birth weight matter for child health in developing countries? estimates from siblings and twins. *Health Economics*, 28(1):3–22, 2019. doi: https://doi.org/10.1002/hec.3823. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.3823.

P. Miller, C. Mulvey, and N. Martin. Family characteristics and the returns to schooling: Evidence on gender differences from a sample of australian twins. *Economica*, 64(253): 119–136, 1997.

P. Miller, C. Mulvey, and N. Martin. Birth weight and schooling and earnings: estimates from a sample of twins. *Economics Letters*, 86(3):387–392, 2005.

P. Miller, C. Mulvey, and N. Martin. The return to schooling: Estimates from a sample of young australian twins. *Labour Economics*, 13(5):571–587, 2006.

P. W. Miller, C. Mulvey, and N. Martin. What Do Twins Studies Reveal about the Economic Returns to Education? A Comparison of Australian and U.S. Findings. *American Economic Review*, 85(3):586–599, June 1995.

M. Mogstad and M. Wiswall. Instrumental variables estimation with partially missing instruments. *Economics Letters*, 114(2):186–189, 2012. doi: 10.1016/j.econlet.2011.10. URL https://ideas.repec.org/a/eee/ecolet/v114y2012i2p186-189.html.

M. Mogstad and M. Wiswall. Testing the quantity–quality model of fertility: Estimation using unrestricted family size models. *Quantitative Economics*, 7(1):157–192, 2016. doi: https://doi.org/10.3982/QE322.

C. Monden, G. Pison, and J. Smits. Twin Peaks: more twinning in humans than ever before. *Human Reproduction*, 36(6):1666–1673, 03 2021. ISSN 0268-1161. doi: 10.1093/humrep/deab029. URL https://doi.org/10.1093/humrep/deab029.

D. Neumark. Biases in twin estimates of the return to schooling. *Economics of Education Review*, 18(2):143–148, April 1999.

P. Oreopoulos, M. Stabile, R. Walld, and L. L. Roos. Short-, Medium-, and Long-Term Consequences of Poor Infant Health: An Analysis Using Siblings and Twins. *Journal of Human Resources*, 43(1):88–138, 2008.

G. Pison and A. V. D'Addato. Frequency of twin births in developed countries. *Twin Research and Human Genetics*, 9(2):250–259, 2006. doi: 10.1375/twin.9.2.250.

T. J. C. Polderman, B. Benyamin, C. A. de Leeux, P. F. Sullivan, A. van Bochoven, P. M. Visscher, and D. Posthuma. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, 47(7):702–709, May 2015.

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.

M. R. Rosenzweig and K. I. Wolpin. Life-Cycle Labor Supply and Fertility: Causal Inferences from Household Models. *Journal of Political Economy*, 88(2):328–348, 1980a.

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 1980b.

M. R. Rosenzweig and K. I. Wolpin. Natural "natural experiments" in economics. *Journal of Economic Literature*, 38(4):827–874, December 2000. doi: 10.1257/jel.38.4.827. URL https://www.aeaweb.org/articles?id=10.1257/jel.38.4.827.

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. E. Rouse. Further estimates of the economic return to schooling from a new sample of twins. *Economics of Education Review*, 18(2):149–157, 1999.

H. Royer. Separated at Girth: US Twin Estimates of the Effects of Birth Weight. *American Economic Journal: Applied Economics*, 1(1):49–85, January 2009.

Ö. Sandewall, D. Cesarini, and M. Johannesson. The co-twin methodology and returns to schooling — testing a critical assumption. *Labour Economics*, 26:1–10, 2014.

A. Sanz-de Galdeano and A. Terskaya. Sibling Differences in Educational Polygenic Scores: How Do Parents React? IZA Discussion Papers 12375, Institute of Labor Economics (IZA), May 2019. URL https://ideas.repec.org/p/iza/izadps/dp12375.html.

P. A. Savelyev, B. C. Ward, R. F. Krueger, and M. McGue. Health endowments, schooling allocation in the family, and longevity: Evidence from us twins. *Journal of Health Economics*, 81:102554, 2022. ISSN 0167-6296. doi: https://doi.org/10.1016/j.jhealeco.2021.102554. URL https://www.sciencedirect.com/science/article/pii/S0167629621001399.

M. A. Silles. The impact of children on women's labour supply and earnings in the uk: evidence using twin births. *Oxford Economic Papers*, 68(1):197–216, 2016. ISSN 00307653, 14643812. URL http://www.jstor.org/stable/43772905.

P. Taubman. The determinants of earnings: Genetics, family, and other environments: A study of white male twins. *The American Economic Review*, 66(5):858–870, 1976a.

P. Taubman. Earnings, education, genetics, and environment. *The Journal of Human Resources*, 11(4):447–461, 1976b.

F. Torche and G. Echevarría. The effect of birthweight on childhood cognitive development in a middle-income country. *International Journal of Epidemiology*, 40(4):1008–1018, 2011.

J. P. Vere. Fertility and parents' labour supply: new evidence from US census data: Winner of the OEP prize for best paper on Women and Work *. *Oxford Economic Papers*, 63(2):211–231, 03 2011. ISSN 0030-7653. doi: 10.1093/oep/gpr003. URL https://doi.org/10.1093/oep/gpr003.

D. Webbink, N. G. Martin, and P. M. Visscher. Does teenage childbearing reduce investment in human capital? *Journal of Population Economics*, 24(2):701–730, 2011.

D. Webbink, P. Koning, S. Vijič, and N. G. Martin. Why are criminals less educated than non-criminals? evidence from a cohort of young australian twins. *Journal of Law, Economics, & Organization*, 29(1):115–144, 2013.

J. Yi, J. J. Heckman, J. Zhang, and G. Conti. Early health shocks, intra-household resource allocation and child outcomes. *The Economic Journal*, 125(588):F347–F371, 2015. doi: https://doi.org/10.1111/ecoj.12291. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ecoj.12291.

J. Zhang. A dilemma of fertility and female labor supply: Identification using taiwanese twins. *China Economic Review*, 43:47–63, 2017. ISSN 1043-951X. doi: https://doi.org/10.1016/j.chieco.2016.12.005. URL https://www.sciencedirect.com/science/article/pii/S1043951X16301675.