ORIGINAL RESEARCH

When Size Matters: IV Estimates of Sibship Size on Educational Attainment in the U.S.

Christina J. Diaz[1] · Jeremy E. Fiel[1]

Abstract

Children with additional siblings appear to fare worse on a variety of developmental and educational outcomes across social contexts. Yet, the causal relation between sibship size and later attainment remains dubious, as factors that influence parents' fertility decisions also shape children's socioeconomic prospects. We apply instrumental variables methods that treat multiple births (e.g., twins, triplets) and same-sex composition as natural experiments to test whether increases in sibship size have a causal effect on the educational attainment of older siblings in the U.S. We pool several nationally representative datasets, including the Child and Young Adult Cohorts of the NLSY79 and the Panel Study of Income Dynamics, to obtain adequate sample sizes for these methods. Although results indicate that the presence of an additional sibling has a trivial effect on the attainment of older siblings for most families (those with two to four siblings), a large penalty arises with the introduction of a fifth sibling. Our findings imply that the costs associated with sibship size are likely concentrated among the largest families.

Keywords Siblings · Sibship size · Educational attainment · Instrumental variables

Introduction

Though children with multiple siblings appear to fare worse on a variety of educational and developmental outcomes (Steelman et al. 2002), there remains uncertainty about the extent to which sibship size *causes* lower attainment. Parents likely make

---

Christina J. Diaz and Jeremy E. Fiel have contributed equally to this wok.

✉ Christina J. Diaz
    cjdiaz@email.arizona.edu

    Jeremy E. Fiel
    jfiel@email.arizona.edu

[1] School of Sociology, University of Arizona, 1145 E. South Campus Dr., Tucson, AZ 85721, USA

decisions about the quantity and quality of their children simultaneously (Becker and Lewis 1973), such that the factors shaping parents' fertility decisions also influence children's socioeconomic prospects. It is thus challenging to isolate the effects of sibship size from other family background effects (Black et al. 2005; Bras et al. 2010; Guo and VanWey 1999; Hanushek 1992; Li et al. 2017).

We borrow an instrumental variables (IV) strategy pioneered by economists studying fertility effects on economic and family processes (Rosenzweig and Wolpin 1980a, b) to identify the effect of sibship size on educational attainment for those born in the twentieth century. We treat multiple births (e.g., twins, triplets) and same-sex sibling composition as natural experiments that increase family size but are ostensibly unrelated to confounding family background variables. Although researchers have used these strategies to examine sibship size effects on parental investment and early academic outcomes in the U.S. (Cáceres-Delpiano 2006; Conley and Glauber 2006), and on educational attainment in other countries (Angrist et al. 2010; Black et al. 2005; Li et al. 2008, 2014), data requirements have limited their capacity to study educational attainment in the U.S. These methods require large samples with accurate birth information on all children in a family as well as information on long-term attainment.

We meet this challenge by pooling nationally representative data from the National Longitudinal Survey of Young Women (NLS), the National Longitudinal Survey of Youth 1979 (NLSY79), the Child and Young Adult Cohorts of the NLSY79 (NLSCYA), and the Panel Study of Income Dynamics (PSID). We use all of these data sources for our sex composition analyses, and we use the NLSCYA and PSID for our multiple birth analyses. It is not possible to examine the effects of a first sibling with these methods, but we are able to examine the effects of subsequent transitions, from a second through a fifth sibling. This range of sibships highlights numerous transitions that remain relevant to many U.S. families.[1] We use two-stage least squares estimation, which allows us to control for potential threats to our identification strategy. We also address concerns with both IV strategies, including whether our multiple birth instrument is likely to be tainted by the adoption of medically assisted reproduction techniques. Our new findings from the U.S. add to the growing skepticism about sibship size effects on educational attainment in developed countries, but also reveal potential exceptions worthy of further research.

---

[1] Our primary analytic sample focuses on children mostly born in the 1970s and 80s, though some were born in earlier years. Estimates from the Current Population Survey suggest that this range of sibship size generalizes to a nontrivial number of families (Downs 2003). The share of women aged 40–44 who had three or more children was around 43% in 1985 (with nearly 10% having five or more children), and was 32% (and 5%) by 1990.

Theories of Sibship Size Effects

Over a century of scholarship has debated whether (and how) family size impacts children's life chances (Blau and Duncan 1967; Galton 1874; Murray 1984). This work has honed in on two possibilities. First, there may be detrimental effects of additional children on families' ability to promote human capital formation among children. Second, the relation between sibship size and child outcomes could be a result of confounding factors that affect both fertility and later well-being. Given the size and scope of this literature—as well as our own goals—we primarily review the extent to which sibship size influences educational attainment.

Causal Effects: Quality Vs. Quantity and Resource Dilution

A popular causal interpretation of sibship size rests on the logic that the resources needed to promote child well-being are finite and are diluted as the number of children increases. Becker et al. incorporated this logic into their economic model of household decision-making via interactions between the quality and quantity of children (Becker and Lewis 1973; Becker and Tomes 1976). Put simply, increasing the quality (e.g., skills) of children is costly, and these costs increase with the number of children. Costs may include parents' time and effort, which are relevant to children's general skill development, as well as financial resources that shape the affordability of higher education. Parents thus face a quality-quantity tradeoff: additional children reduce their ability to invest in each child's human capital and ultimately leads to lowered educational attainment.

Although this model predicts negative sibship size effects, closer inspection yields some ambiguity. First, not all resources or experiences benefit children, and the dilution of negative effects could help children in larger sibships (Steelman et al. 2002). Second, resources are not necessarily diluted with additional children. Parents may gain experience and knowledge from rearing older children that informs subsequent child-rearing practices, or they may strategically shift priorities to prevent the dilution of resources and experiences that are critical to children's educational success (Angrist et al. 2010; Frenette 2011; Phillips 1999). Because some resources, particularly financial ones, may be more prone to dilution than others, sibship size effects may vary across levels of education.

Confounding Factors and Spurious Effects

If parents are aware of the costs of child-rearing and make fertility decisions with their children's attainment in mind, they must simultaneously make decisions about the quantity and quality of children. Hence, many of the same parental attributes that influence the number of children in a family also influence parental investments in children (Guo and VanWey 1999; Page and Grandon 1979; Rodgers et al. 2000). Scholars occasionally point to the genetic transmission of intelligence and other skills that may be conducive to educational success (Plomin et al. 1994). More

often, there is a focus on proxies for values, skills, and resources that are likely related to parental fertility and child outcomes—including socioeconomic background, race/ethnicity, maternal age, family structure, and region (Guo and VanWey 1999; Hanushek 1992; Heer 1985; Steelman 1985).

Empirical Evidence and Issues of Causality

Scholars employ a variety of strategies to account for confounders when assessing the causal effects of sibship size on human capital accumulation. One approach statistically controls for potential confounders related to family background in regression or structural equation models. Almost invariably, controlling for such variables reduces the negative association between sibship size and educational outcomes, but statistically significant and substantively important relations often persist (Blake 1981, 1989; Steelman et al. 2002). Unfortunately, these findings are based on analyses of datasets that were not designed to study family size and lack important control variables (Heer 1985). Even with ideal data, this problem may be intractable with standard regression-based (or other covariate adjustment) methods due to the difficulty of measuring all determinants of sibship size that also affect children's outcomes.

The Rising Popularity of Instrumental Variables

Within the past 30 years, a growing number of studies have adopted instrumental variables (IV) estimation to assess sibship size effects. This approach attempts to isolate "clean" variation in sibship size that is uncorrelated with confounding family background characteristics. The goal is to find an instrument that affects the number of siblings but is unrelated to the outcome through any pathway other than sibship size.

One appealing instrument is the occurrence of a multiple birth event (e.g., a twin or triplet birth). Among women who have a second (or third, etc.) birth event, a multiple birth will increase the total number of children among those who otherwise would not have had an additional child. Rosenzweig and Wolpin (1980b) pioneered this method, and others have refined it to assess sibship size effects on educational attainment across the world. Findings suggest that effects differ across regions and periods, possibly due to differences in economic development and changing cultural norms.

In developing countries, the evidence from multiple birth IV studies is nuanced. Li et al. (2008), for instance, find that negative effects are concentrated within rural areas in China. Ponczek and Souza (2012) argue that the presence of an additional child decreases contemporary educational attainment in Brazil by over half a year—but only for females. Yet, in earlier phases of Brazil's development, there were *positive* effects of larger sibship sizes on schooling. These effects waned and trended negative throughout the twentieth century as regions expanded educational systems and family planning (Marteleto and de Souza 2012).

Conversely, most studies using multiple birth IVs in developed countries find that sibship effects on educational attainment are largely nonexistent. These patterns are documented in Denmark (Bagger et al. 2013), Norway (Black et al. 2005), Sweden (Åslund and Grönqvist 2010), and Israel (Angrist et al. 2010). Two studies that employ multiple birth instruments in the U.S. reach slightly different conclusions, but with the caveat that they are not directly comparable. One found that the presence of an additional sibling slightly reduces parental investment in older siblings, at least as indicated by children's private school attendance (Cáceres-Delpiano 2006). Another, limited to families of a cohort of 1957 Wisconsin high school graduates, found no effects on educational attainment (de Haan 2010).

A second popular instrument is the sex composition of siblings (Angrist and Evans 1998). Since many parents prefer a mix of boy and girl children, those with only boys or only girls are more likely to have an additional child than those with mixed-sex offspring. And because the sex of each child is essentially random, so is the sex composition of children in a family of a given size. A few key findings from this research should be noted. First, to the extent that negative sibship effects exist, they seem to weaken over time as families across Latin America and Asia (Li et al. 2014, 2017) adjust to economic change and declining fertility. Regardless of temporal trends, the penalty for having additional siblings is magnified for girls relative to boys across developing countries (Lee 2008; Li et al. 2017). Yet, it is striking that others report negligible sibship effects on educational attainment in Israel (Angrist et al. 2010), Mexico (Fitzsimons and Malde 2014), and in the U.S. (de Haan 2010).

Scholars posit that sibship effects are muted in more developed regions—especially throughout Northern Europe—because families benefit from centralized education systems and public investment in children, which might offset the dilution of human capital investments in large families (e.g., Li et al. 2008). Although the U.S. qualifies as a developed nation and was a leader in the expansion of public education, it is also characterized by a decentralized and unequal education system, high levels of economic inequality, and relatively limited governmental support for families (Goldin and Katz 2009). As such, it is important to further assess how sibship size impacts the accumulation of human capital for families living in the U.S.

Multiple Births and Sex Composition as Natural Experiments

Our study represents one of the first efforts to apply multiple birth and sex composition IV methods to assess sibship size effects on educational attainment in the U.S. We use nationally representative data from surveys collected throughout the twentieth century to assess possible penalties across a variety of sibship sizes. Here we describe these methods more thoroughly.

Multiple Birth IV Design

Consider the following scenario: a woman with one child from her first birth experiences a second birth, which is a potential multiple birth event. She may have twins

upon the birth of her second child, giving her three children and giving her oldest child two siblings. If she does not have twins, her oldest child will have one sibling. Assuming she has no more children in either case, the oldest child's full sibship size will be two in the twin case and one in the non-twin case.[2] Although we cannot identify individual effects, because we never observe both conditions for any individual, we can estimate population-level effects by comparing the children of women with multiple second births to the children of women with singleton second births.

This captures variation in sibship size among children whose families had more children than planned due to a multiple birth. The resulting estimand is known as a local average treatment effect (LATE) (Angrist et al. 1996), which comes with some important caveats. First, it only generalizes to compliers, who are the children whose sibship size increased *because* their parents had a multiple birth. Angrist et al. (2010) demonstrate that the multiple birth IV estimand is the average treatment on the untreated effect (TOU), which means it generalizes to all children whose mothers did not have an additional child. In other words, this captures the effect of an additional *unexpected* sibling. Second, this method captures the effect of small, specific changes in sibship size. Since most multiple births are twin births, it tends to capture the effect of a *single additional sibling* at a given level of sibship size. Although this prevents comparisons of children with few (1) and many (10) siblings, local estimates are more consistent with decisions about having an additional child.

Marginal fertility decisions at the population level can be assessed using the total fertility rate (TFR). Figure 1 shows that the TFR in the U.S. reached around 3.7 in the 1950s, declined steadily to fall below 2.0 in 1970s and 1980s, and has since stabilized just under two children per woman. A TFR of approximately two children suggests that many families face decisions about whether to add a second or third child (corresponding to one or two siblings).[3] However, there continues to be nontrivial variation in fertility; as of 2014, around 15% of 40 to 44-year-old women remained childless and nearly 33% had three or more children (Pew Research Center 2015). While our design cannot capture the effect of a first sibling (second child), we can capture the effect of a second sibling as well as several further additions.

Exclusion Restriction

The validity of the multiple birth IV method hinges on the exclusion restriction, which assumes that the two groups being compared—children whose mothers had multiple second births and children whose mothers had singleton second births—are similar with respect to all traits that affect their educational outcomes (except those influenced by completed sibship size).

---

[2] Among women who would have had more than three children regardless of a multiple second birth event, multiple births do not contribute to variation in sibship size. This is not problematic because such women should be equally represented in the multiple birth and singleton birth groups.

[3] The estimates of interest do not change after adjusting for temporal shifts in the average age at child-bearing (e.g. tempo effects). Authors' calculations are based on U.S. Vital Statistics.
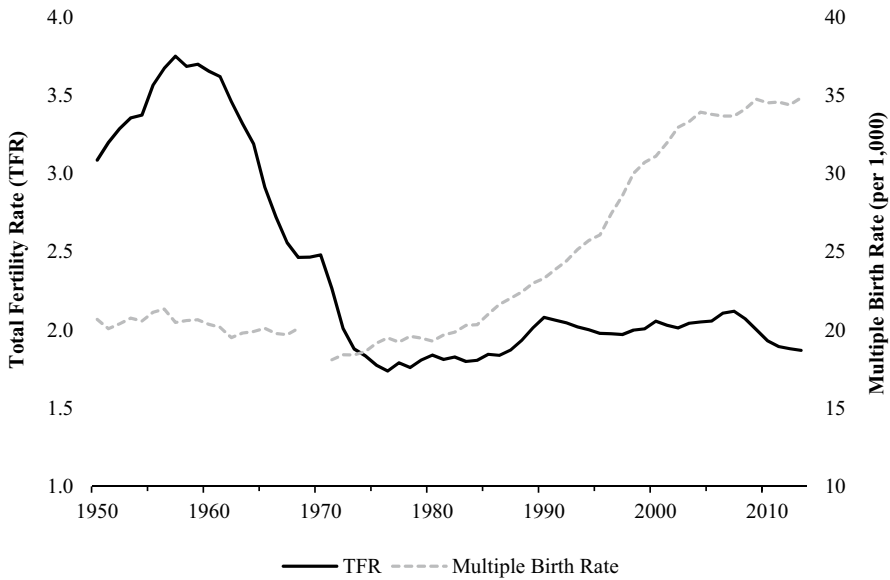
Fig. 1 Fertility and multiple births, 1950–2013. This plot shows the total fertility rate (solid black) and multiple birth rate (per 1000 births; dashed gray) from 1950 to 2013, based on authors' calculations from U.S. Vital Statistics. Multiple birth data are not available for 1969 and 1970

The exclusion restriction may not hold for the twins themselves (or triplets), who tend to be lower in birth weight and cognitive ability than singletons, and who may be treated differently than singletons (Black et al. 2005; Rosenzweig and Zhang 2009). Hence, we follow prior work by limiting our attention to the older children born prior to the potential multiple birth events. When examining the effects of third and higher siblings, we include all older children born prior to the potential multiple birth event.

A potential problem is that multiple birth events are not completely random—they are typically higher among women who are older, black, or have a family history of multiple births (Bortolus et al. 1999; Bulmer 1970; Lichtenstein et al. 1996). Black families are more likely to be socioeconomically disadvantaged, and their children tend to complete less education than white children for reasons other than differences in family size (Kao and Thompson 2003). And women who delay fertility tend to be more socioeconomically advantaged than younger mothers, but have a higher risk for complications that impact child health (Fretts et al. 1995)—either of which may influence children's educational outcomes. It will thus be important to control for mothers' race and age at birth.

Another issue that warrants concern is medically assisted reproduction (MAR), which includes in vitro fertilization, ovulation induction, and other fertility treatments that increase the likelihood of multiple births (Pison and D'Addato 2006;

Pison et al. 2015; Schieve et al. 1999).[4] If couples who use and do not use such treatments differ from each other in ways that influence children's educational outcomes, this could introduce bias into our IV estimates. Although we cannot directly observe individuals who used MAR, we can review population trends in multiple births and MAR to assuage such concerns.

In 1950, multiple births accounted for approximately 2% of all live births—a number that remained relatively stable until the late 1980s (see Fig. 1). The multiple birth rate increased after 1985, accounting for over 3% of all live births in the early 2000s. This trend coincided with an increase in fertility treatments and was driven by an increase in twin births, as well as a dramatic spike in triplet and higher plurality births; these rates subsided as medical professionals began limiting the transfer of multiple embryos (Practice Committee of the American Society for Reproductive Medicine 2013). Nevertheless, estimates available beginning in the late 1990s suggest that fertility treatments were then responsible for nearly 20% of all twin births and over 80% of triplet and higher-order births (Kulkarni et al. 2013).

We believe that systematic MAR-related selection into multiple births is unlikely to pose a major problem in our study. First, only three (1.5%) multiple births in our study are triplets and none are of a higher-order. Because MAR increases the likelihood of high plurality births, we confirm that all results are substantively identical when excluding triplets from our analytic sample (available upon request). Second, because we limit our analyses to individuals old enough to complete most of their schooling (age 25 and over) by the latest survey (2013–2014 in our data), the majority of respondents were born in the 1980s or earlier—before the mainstream expansion of fertility treatments.[5] Third, MAR treatments are concentrated among families struggling with infertility—more than 71% of women who receive fertility treatments have not previously given birth (Centers for Disease Control 2007). These women are unlikely to remain in our samples as they must have had at least one child prior to the focal birth event; at higher birth numbers, families who use MAR are even less likely to remain in the sample because they must have multiple prior children.

As an additional check, we estimated logistic regressions with our data to assess changes in the odds of multiple birth events over time that could be driven by MAR treatments. Net of the control variables described in the following section, we found no significant increases in multiple births occurring after 1980 or after 1985.[6] Nonetheless, if women who received MAR treatment remain our sample, we expect them to be relatively socioeconomically advantaged and to have stronger preferences for children. In this case, our multiple birth IV estimates may be upwardly biased as such offspring are likely to be positively selected with respect to educational

---

[4] Although delayed childbearing, which is also associated with MAR, is an important contributor to the increased rate of twinning, it cannot solely account for increases in multiple births.

[5] Nearly all multiple births in our analytic sample (96.1–99.2%, depending on the *n*th birth event) occurred before 1997—the year in which the Centers for Disease Control (CDC) began systematically collecting data on assisted reproductive technologies.

[6] Results are available upon request.

outcomes. Our two-stage least squares analyses thus includes controls, such as parental education, to mitigate this bias. Moreover, we found no evidence of differential IV estimates for birth events after 1985, which might have captured bias due to selection into MAR.

Finally, it is worth noting that because twins are more likely to have early health and developmental problems, parents may respond by redirecting resources (Rosenzweig and Zhang 2009). If parents compensate less-endowed children, twins may draw resources away from older siblings. If parents reinforce their better-endowed children, older non-twins may draw resources away from twins. These scenarios suggest that the effects of siblings added from multiple births may differ (perhaps more negatively) from the effect of siblings added for other reasons.

The Sex Composition IV Design

Similar logic extends to using same-sex composition—defined as having two boys or two girls—as an instrument, though a few differences should be highlighted. Here, we can include the oldest two children (rather than just the oldest child) in our analysis of adding a second sibling. The LATE interpretation still applies, but compliers represent those who had an additional sibling because their parents preferred mixed-sex children enough to add another child. Unlike the multiple birth case, the added child induced by same-sex composition is expected or planned by parents. If this provides parents better a chance to prepare for the added child, we might expect more favorable sibship effects when using same sex composition rather than multiple births (Black et al. 2010). In addition, parents' preferences for a certain sex composition might also influence their allocation of resources across children (Conley 2000). It seems plausible that such allocations would favor the added children, some of whom will be the (desired) opposite sex. If so, IV estimates for older children might muddle child sex effects with sibship size effects, perhaps resulting in downward bias.

Data and Measures

Both methods have substantial data requirements. Multiple births have strong effects on the number of added children but are rare, so large samples are required to obtain sufficient variation in added siblings for informative estimates. Sex composition is a weaker IV. Because same-sex children seem to induce relatively few families to add another child, even larger samples may be required to detect effects (Angrist and Evans 1998). Both approaches require information on the educational attainment and birth order of all siblings; the same-sex analyses require each sibling's gender, and the multiple birth analyses require sibling birth dates. Most applications of these methods use data from other countries that link census records to administrative data (Angrist et al. 2010; Black et al. 2005). The previously described studies of U.S. families use Census microdata without examining educational attainment

(Cáceres-Delpiano 2006; Conley and Glauber 2006), or rely on the less generalizable Wisconsin Longitudinal Study (de Haan 2010).

Data Sources

We pool data from the National Longitudinal Survey of Young Women (NLS), the National Longitudinal Survey of Youth 1979 (NLSY79), the Child and Young Adult Cohorts of the NLSY79 (NLSCYA), and the Panel Study of Income Dynamics (PSID). Both the NLSCYA and PSID provide maternal reports of all children's birth month and year, and both follow children long enough to measure their completed educational attainment. The NLS and NLSY79 only sample subsets of children from families; information on full sibships comes from follow-ups in which those surveyed report rosters with each sibling's gender, age, and educational attainment. This is sufficient for the sex composition analyses but problematic for the multiple birth analyses. When we derived multiple birth indicators from siblings' reported ages in these rosters, we found higher-than expected multiple birth rates, discrepant reports across siblings, and poor performance in the IV analyses (weaker than expected effects on added siblings). Moreover, we are concerned that errors in the multiple birth indicators are correlated with sibship size. Hence, we use all datasets for sex composition analyses but restrict the multiple birth analyses to the NLSCYA and PSID.

Outcomes

Our primary outcome variable is completed years of schooling. We restrict the sample to those who were at least 25 years old at the latest follow-up; analyses of those at least 30 years old yield similar results. To explore how sibship effects may bear on specific educational transitions, we also examine whether respondents completed at least 12, 13, and 16 years of school; these roughly correspond to high school completion, college attendance, and college completion, respectively. While our continuous measure captures the growth of skills and knowledge that presumably accompanies additional exposure to the education system, degree completion reflects prestige as well as the accumulation of credentials that are rewarded in the labor market (e.g., David 1999). Moreover, this allows us to consider whether certain sibship configuration effects may be more sensitive to financial resources (college attendance) or basic skill development (high school completion).

Control Variables

Multiple births are not completely random, particularly with respect to race and maternal age. To adjust for racial differences, we use white as the reference group and include indicators of whether the mother is identified as black or as other

nonwhite.[7] We control for mother's age and age-squared in years; age is measured at the potential multiple birth event, or at the most recent birth in the sex composition analyses. We also interact race with mother's age at birth. Dummy indicators of the data source and a continuous measure of the older sibling's age during the latest survey help protect against confounding and increase precision. Continuous measures of the highest grade of school completed by either parent and each child's year of birth are included as controls. Categorical indicators of each child's gender, birth order, and cohort (decade of birth) are also added. In the multiple birth analyses, we include the child's age at the potential multiple birth event, the year of that birth event, and an indicator of whether it occurred in or after 1985 to account for the possible increase in fertility treatments after this time period. Additional analyses (available upon request) include family income as a control to further adjust for potential selection into MAR; results reinforce the findings reported here but yield smaller samples due to missing income data.

Methods

We follow prior work and separately estimate sibship size effects at different levels (n), beginning with the effect of a second sibling ($n=2$; the third child), which we describe first. Our sex composition analyses use same-sex pairs among the first two children as the instrument, and the multiple birth analyses use a multiple birth event at the second child's birth. Our preferred specification of sibship size uses a dummy indicator of whether the instrument added "extra" siblings—e.g., whether there were two or more ($n+$) siblings for the effects of the second (nth) sibling. An alternative specification uses a continuous measure of the total number of siblings. The dummy indicator better aligns with the literal interpretation of the IV design: it captures the effect of having additional siblings due to same-sex composition or a multiple birth. The continuous specification is intuitive in that it captures the per-child effect of these added siblings, but it imposes a linear functional form on these effects that may be unrealistic. Nonetheless, our analyses suggest that the IV analyses using the dummy indicator tend to approximate the effects of one additional sibling, and most findings are consistent across specifications.

We conduct OLS analyses designed to be as comparable as possible to the IV analyses, using the same older siblings as observations. Equation 1 shows the specification, with S representing sibship size and $\mathbf{X}$ representing the control variables described previously.

$$Y_i = \beta_0 + \beta_1 S_i + \mathbf{X}_i \boldsymbol{\beta} + u_i \tag{1}$$

We use two-stage least squares (2SLS) estimation to implement the IV strategies. For the multiple birth analyses, the first stage (Eq. 2.1) predicts sibship size S using a multiple birth indicator specific to the birth of the second child (Z_i) and the

---

[7] PSID treats race as a household variable, so each family is coded as the race of the head of household.

same set of covariates **X** included in the OLS specification. For the sex composition analyses, the first stage replaces the multiple birth indicator with an indicator of same-sex composition among the first two children (Z_i). The second-stage (Eq. 2.2) regresses educational attainment on the predicted values of sibship size ($\widehat{S}_i$) from the first stage and covariates **X**. We adjust standard errors for clustering whenever the samples include multiple siblings from the same family.

$$S_i = \alpha_0 + \alpha_1 Z_i + \mathbf{X}_i \boldsymbol{\alpha} + v_i \tag{2.1}$$

$$Y_i = \beta_0 + \beta_1 \widehat{S}_i + \mathbf{X}_i \boldsymbol{\beta} + u_i \tag{2.2}$$

We replicate these analyses for sibship sizes up to five using the multiple birth approach. These instruments are based on multiple births at the birth of the third ($n=3$), fourth ($n=4$), and fifth ($n=5$) child; we include all older children born prior to the focal birth event as observations. We conduct analogous OLS analyses at these levels but cannot use the sex composition approach. The effects of same-sex composition on additional children are too weak at these higher levels, presumably because same-sex composition is less likely as sibship size increases, and because the costs of adding more children may outweigh preferences for mixed-sex composition in most families.

Results

Descriptive Statistics

Table 1 summarizes the characteristics of each analytic sample. For instance, "2+ sibs" represents the effect of moving from one to two or more siblings; the sex composition analyses include the first two children, and the multiple birth analyses only include the oldest child. The "3+ sibs" column corresponds to the oldest two siblings included in the multiple birth analyses at the third child's birth, and so on. Focusing on the 2+ sample for the sex composition analyses, average educational attainment is 13.3 years. About half of the children are from same-sex sibships, which is expected given the randomness of child sex. About 76% have two or more siblings ("$N+$ siblings"), and average sibship size is 3.2. Though the latter is higher than average throughout this period, this makes sense given that our sample excludes all single-child families. Most of these children were born between the 1940s and 1980s and appear in the NLS (Women), NLSY79, NLSCYA, and PSID data.

In the 2+ sample for the multiple birth analyses, average attainment is 13.7 years, 57% of the children have two or more siblings, and average sibship size is about 2. Compared to the sex composition sample, educational attainment is higher and sibship size is smaller because these children come from more recent birth cohorts—mainly the 1970s and 1980s—because we draw exclusively on the PSID and NLS-CYA. Educational attainment declines across birth numbers, which is expected given the negative association between sibship size and schooling. All children in the 2+ sample are first-borns; second-, third-, and fourth-born children enter the

samples at subsequent levels. Other attributes vary predictably across birth numbers: mothers' age at birth increases, blacks are increasingly represented, and parental education declines.

U.S. vital statistics indicate that roughly 2.0% of all births during the 1980s were multiple births. Because vital statistics do not disaggregate multiple births by parity, and because multiple births increase with maternal age, we expect below-average multiple birth rates at lower parities and above-average rates at higher parities. Accordingly, our multiple birth rates are 1.2% at birth two, 1.1% at birth three, 1.3% at birth four, and 2.3% at birth five. Of all multiple births in our sample, 98.7% are twins (not shown), consistent with vital statistics (97–99%) during this period.

First Stage Results

Table 2 summarizes our preferred analyses using the binary "at least one additional sibling" indicator. At the 2+ level, the first column shows the first-stage effect of each instrument on having more than one additional sibling and the corresponding *F*-statistic. The second column shows the total "reduced form" effects of the instruments on older siblings' educational attainment, the third shows the corresponding 2SLS estimate of the sibship effect on educational attainment, the fourth shows the corresponding OLS estimate with no controls, the fifth shows the OLS estimate with controls, and the sixth reports the p-value of an endogeneity test based on whether the IV and OLS estimates (with controls) are significantly different.

Here we focus on first-stage results. Being one of two same-sex siblings increases the probability of having two or more siblings by only 4% points (Table 2). The *F*-statistic is 34.2, which is above the typical rule of thumb ($F = 10$) for diagnosing weak instruments (Staiger and Stock 1997). Turning to the multiple birth analyses, a multiple second birth increases the probability that the oldest child will have two or more siblings by 44.6% points, with an *F*-statistic of 293. The first-stage effects are stronger at subsequent birth events, which is expected because multiple births should only increase the total number of children among families who would not have had additional children otherwise; there are likely more families, for instance, willing to have a third child than families willing to have a sixth child. The first-stage effects found here are comparable to prior work in other developed contexts (e.g., Angrist et al. 2010). That multiple births are much stronger instruments than sex composition is not surprising and suggests that multiple birth IV estimates will be more informative.

We present an (Appendix Table 4) that summarizes parallel analyses using the total number of siblings. With respect to same-sex composition, the *F*-statistic of 2.5 is low enough to worry about a weak instrument and finite sample bias—that is, estimates may be sensitive to chance correlations between the instrument and the error term due to sampling error. The first-stage coefficients for our multiple birth instruments are consistent with past work (Cáceres-Delpiano 2006; Cáceres-Delpiano and Simonsen 2012), and the *F*-statistics remain above ten but are considerably lower than when using the binary "additional sibling" indicator. We focus our discussion

Table 1 Descriptive statistics

	Sex comp 2+ sibs N=22,230		Mult. birth 2+ sibs N=4849		Mult. birth 3+ sibs N=5072		Mult. birth 4+ sibs N=3257		Mult. birth 5+ sibs N=1754	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Years of school	13.308	2.659	13.721	2.212	13.425	2.220	13.047	2.240	12.843	2.261
Sibship size	3.211	2.335	2.006	1.262	2.851	1.266	3.904	1.375	5.134	1.564
$N+$ siblings	0.758	0.428	0.565	0.496	0.470	0.499	0.459	0.498	0.514	0.500
Multiple birth	–	–	0.012	0.109	0.011	0.106	0.013	0.113	0.023	0.151
Same sex	0.504	0.500	–	–	–	–	–	–	–	–
2 boys	0.220	0.414	–	–	–	–	–	–	–	–
2 girls	0.284	0.451	–	–	–	–	–	–	–	–
Mother's age at birth N	24.307	4.922	25.133	4.808	27.374	5.008	28.958	5.050	30.153	5.017
Year of Nth birth	–	–	1978.3	11.568	1979.0	12.640	1977.6	13.596	1974.9	14.310
Birth event N 1985+	–	–	0.355	0.479	0.431	0.495	0.411	0.492	0.364	0.481
Black	0.296	0.456	0.355	0.478	0.394	0.489	0.445	0.497	0.493	0.500
Other	0.122	0.327	0.127	0.333	0.133	0.340	0.134	0.341	0.142	0.349
Female	0.539	0.498	0.504	0.500	0.512	0.500	0.516	0.500	0.523	0.500
Age at Nth birth	–	–	3.919	3.110	5.690	3.967	6.928	4.388	7.586	4.497
Year of birth	1961.8	14.544	1974.4	10.824	1973.4	11.524	1970.9	12.342	1967.6	13.057
Cohort: 1920s	0.002	0.042	0.000	0.020	0.001	0.024	0.000	0.000	0.000	0.000
Cohort: 1930s	0.025	0.155	0.001	0.038	0.000	0.020	0.000	0.000	0.001	0.024
Cohort: 1940s	0.190	0.393	0.030	0.171	0.032	0.177	0.041	0.199	0.062	0.241
Cohort: 1950s	0.319	0.466	0.104	0.305	0.142	0.349	0.215	0.411	0.307	0.462
Cohort: 1960s	0.158	0.364	0.124	0.330	0.138	0.345	0.165	0.371	0.182	0.386
Cohort: 1970s	0.105	0.307	0.317	0.465	0.274	0.446	0.228	0.419	0.152	0.359
Cohort: 1980s	0.200	0.400	0.424	0.494	0.413	0.492	0.352	0.478	0.295	0.456

Table 1 (continued)

	Sex comp		Mult. birth		Mult. birth		Mult. birth		Mult. birth	
	2+ sibs		2+ sibs		3+ sibs		4+ sibs		5+ sibs	
	N=22,230		N=4849		N=5072		N=3257		N=1754	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Cohort: 1990s	0.001	0.035	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Age at last survey	34.834	7.321	34.564	9.377	35.089	9.880	36.489	10.537	38.194	11.239
Data: NLS-Men	0.005	0.070	–	–	–	–	–	–	–	–
Data: NLS-Women	0.202	0.401	–	–	–	–	–	–	–	–
Data: NLSCYA	0.175	0.380	0.414	0.493	0.387	0.487	0.325	0.469	0.263	0.441
Data: NLSY79	0.386	0.487	–	–	–	–	–	–	–	–
Data: PSID	0.233	0.423	0.586	0.493	0.613	0.487	0.675	0.469	0.737	0.441
Birth order: 1st	0.532	0.499	1.000	0.000	0.537	0.499	0.370	0.483	0.279	0.449
Birth order: 2nd	0.468	0.499	–	–	0.463	0.499	0.346	0.476	0.268	0.443
Birth order: 3rd	–	–	–	–	–	–	0.284	0.451	0.249	0.432
Birth order: 4th	–	–	–	–	–	–	–	–	0.205	0.404
Birth order: 5th	–	–	–	–	–	–	–	–	–	–
Parental education	12.263	3.096	13.243	2.431	12.951	2.534	12.384	2.748	11.943	2.826

Note N+ siblings and Birth #*N* refer to the analyses corresponding to the effect of an *N*th sibling. Sample restricted to those with education data at age 25 or older

Table 2 Analyses for years of schooling (N+ Siblings)

	First stage	Reduced form	IV	OLS	OLS+ contr	Endog. p(IV=OLS)	Obs
IV: same sex (PSID, NLSCYA, NLSY79, NLS)							
2+ Siblings	0.040*	0.039	0.973	− 1.000*	− 0.272*	0.205	22,230
	(0.007)	(0.039)	(1.004)	(0.046)	(0.044)		
	$F=34.23$						
IV: multiple births (PSID, NLSCYA)							
2+ Siblings	0.446*	− 0.077	− 0.173	− 0.545*	− 0.190*	0.977	4849
	(0.026)	(0.249)	(0.557)	(0.063)	(0.061)		
	$F=292.78$						
3+ Siblings	0.539*	0.043	0.079	− 0.631*	− 0.174*	0.653	5072
	(0.033)	(0.306)	(0.567)	(0.073)	(0.068)		
	$F=262.67$						
4+ Siblings	0.572*	− 0.114	− 0.200	− 0.310*	0.122	0.454	357
	(0.039)	(0.245)	(0.430)	(0.104)	(0.098)		
	$F=210.98$						
5+ Siblings	0.595*	− 0.867*	− 1.458*	− 0.672*	− 0.292*	0.051[†]	1754
	(0.061)	(0.361)	(0.616)	(0.149)	(0.145)		
	$F=95.67$						

Sibship variable (*N*+ siblings) specified as having *N* or more siblings. Household-clustered standard errors in parentheses. Endogeneity test refers to the *p* value for the test of no difference between the IV estimate and the OLS estimate with controls

*$p < .05$, [†]$p < .10$

on analyses using the binary indicator given that it is a stronger instrument and does not impose a linear functional form.

We can assess the performance of both instruments by estimating their effects on achieving different levels of family size with linear probability models (Angrist et al. 2010). The graphs in Fig. 2 plot the effect of two same-sex children or a multiple second birth on the probability of having at least 2, 3, 4, etc., siblings. If either instrument adds one "extra" child, we expect it to increase the probability of having at least two siblings, but not to increase the probability of having many more additional siblings.

The plot for sex composition shows that two same-sex children increases the probability of at least two siblings (2+; one additional sibling) by about 4% points, increases the probability of at least three siblings by 2–3 points, and increases the probability of at least four siblings by about 1 point. There is no effect on five or more siblings. This suggests that a small but significant fraction of families who fail to reach the desired sex composition on the third birth continue to have additional children. Hence, the appropriate LATE generalization may not be limited to the effect of a second additional sibling; this makes the estimates more generalizable but also prevents a clear and specific interpretation.

The multiple birth plot illustrates that a second multiple birth increases the probability of at least two siblings by 45% points, increases the probability of at
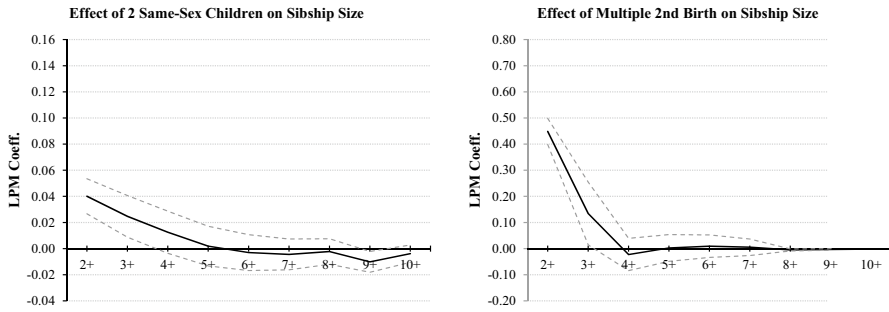
Fig. 2 Instrument effects on sibship size. These graphs show the effect (*y*-axis) of same sex composition (first 2 children) and multiple births (2nd birth) on the probability of older siblings having a given number of siblings or more (*x*-axis), estimated via linear probability model (LPM). The gray lines are 95% confidence intervals
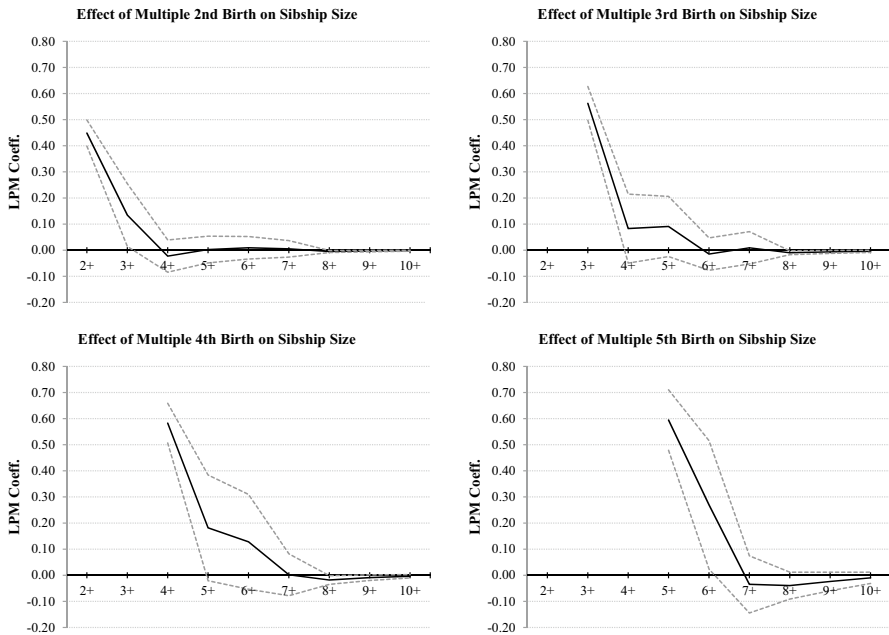

Fig. 3 Multiple birth effects on sibship size

least three siblings by 13 points, and has negligible effects on additional siblings. This is consistent with most multiple births adding a single unexpected child through twinning, but also with a few families adding a second unexpected child through triplets (1.5% of multiple births in our sample). Hence, the LATE interpretation applies to the approximate effect of a second sibling. Figure 3 shows comparable plots for the multiple birth effects at all four levels (2, 3, 4, and 5). Multiple births typically increase sibship size by one and more rarely by two

siblings. There are slight deviations at the third and fourth births, where families with multiple births have slightly more children than expected. Nonetheless, these instruments appear to mainly be capturing the effect of a single added sibling at each level.

OLS and 2SLS Results

Turning to the educational attainment analyses, the "Reduced Form" estimate in Table 2 shows that same-sex composition is associated with a non-significant 0.04-year advantage in educational attainment. The IV analysis scales this relative to the first-stage effect, yielding a coefficient of 0.97, a very imprecise estimate (SE of 1.0) of a one-year *advantage* of having more than one sibling. This deviates substantially (but not statistically significantly) from the unadjusted (− 1.00) and adjusted OLS estimates (− 0.27)—both of which are statistically significant (Fig. 4). These findings are not shocking in light of previous research that finds modest positive effects when using sex composition (Angrist et al. 2010; Black et al. 2010; de Haan 2010).

The multiple birth IV estimates tell a different story. Multiple second births decrease older children's educational attainment by 0.08 years (reduced form). This translates to a 0.17-year disadvantage of having more than one sibling (IV). Estimates are modest and not statistically significant; they are also similar to the adjusted OLS estimates. Others using this method in developed contexts rarely find substantial negative effects of a second sibling on educational attainment (Angrist, Lavy and Schlosser 2010; Åslund and Grönqvist 2010; Bagger et al. 2013; Black et al. 2005).

Multiple birth IV estimates at higher sibship sizes are also negative: a third sibling comes with a 0.08-year advantage in schooling, and a fourth comes with a 0.20-year disadvantage. Again, these are not statistically significant, and they are not significantly different than the adjusted OLS estimates. The addition of a fifth sibling appears to have a stronger negative effect. The IV estimate indicates that a fifth sibling reduces schooling by 1.46 years, which is statistically significant and substantively large: it is over twice the female advantage (0.55) in the same regression and over eight times the parental education coefficient (0.19, not shown). The IV estimate for five or more siblings is also much larger than the corresponding adjusted OLS estimate (− 0.29), a marginally significant difference ($p = 0.051$).

Levels of Schooling

To assess the levels of schooling at which sibship size effects may operate, we replicate analyses with linear probability models that predict whether the older siblings completed high school (at least 12 years of school), attended college (at least 13), or completed college (at least 16). Figure 4 summarizes results using our dummy indicator as the sibship size variable.[8]

---

[8] To increase sample sizes, we include all observations with educational data reported at age 19 or older for high school completion, at age 20 or older for college attendance, and at age 23 or older for bachelor's degree attainment.
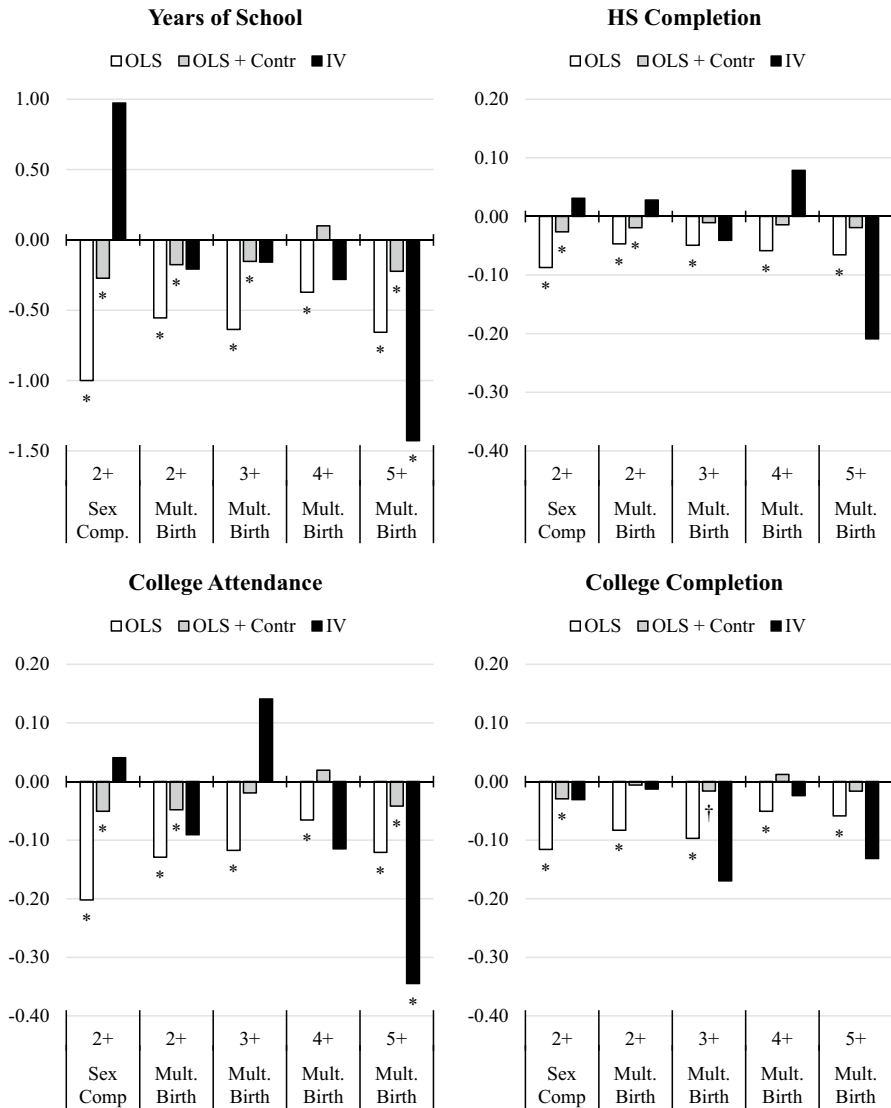
Fig. 4 Sibship effects on educational outcomes: OLS and IV Estimates. Sibship size is specified as a dummy indicator. The *y*-axis represents the sibship size effect for the corresponding outcome. The white bars are OLS estimates; the gray bars are OLS estimates with controls, and the black bars are IV estimates from 2SLS models with controls

The sex composition analyses of two or more siblings yield small positive coefficients for high school completion and college attendance and small negative coefficients for college completion, none of which are statistically significant. The multiple birth analyses reveal negative IV estimates for a second and fourth sibling concentrated at the transition to college attendance, and a negative estimate for a

third sibling at college completion, but none of these are significant. These fluctuations compound the uncertainty reflected in the imprecise sibship size effect estimates reported earlier. The substantial reduction in schooling upon the addition of a fifth sibling, however, appears at all levels of attainment, reducing the probability of completing each educational milestone by at least 11% points. This is most pronounced at the transition to college, reducing the probability of attendance by 38% points; though imprecise, this is estimate is statistically distinguishable from zero.

Additional Considerations

Although we cannot directly test the exclusion restriction—that same-sex composition or multiple births are only correlated with schooling through sibship size—we can indirectly test it by assessing whether these instruments are associated with observable factors that may be independently related to educational attainment (Black et al. 2010). We focus on parental education, which is available in all datasets. Although our 2SLS analyses control for parental education, we might still be concerned if it is associated with our instruments. We test this by regressing parental education on our instruments and other controls. None of the associations are statistically significant, although they still warrant attention (Table 3).

Parental education is 0.08 years higher among families with two same-sex children; this is about twice the association between same-sex composition and children's education (0.04, the reduced form estimate in Table 2). If same-sex composition is positively associated with other unobserved variables that influence child attainment net of parental education and other controls, our IV estimates may be upwardly biased. With respect to multiple births, parental education is about 0.18 years *lower* among families who experience multiple second births; this is over twice as large as the reduced-form association with children's education (− 0.08 in Table 2) and associations are larger at higher birth numbers. Additional analyses of the smaller samples with family income data found more modest, non-significant, negative associations between multiple births and family income (not shown). This suggests that positive selection related to MAR is unlikely, although it does raise concerns about negative selection. If

Table 3 Specification checks: instrument associations with parental education

	Coeff	SE	*p* value	# Obs
IV: same sex (PSID, NLSCYA, NLSY79, NLS)				
First 2 children (2+ Siblings)	0.076	(0.056)	0.180	22,230
IV: multiple births (PSID, NLSCYA)				
Birth 2 (2+ Siblings)	−0.185	(0.275)	0.501	4849
Birth 3 (3+ Siblings)	−0.199	(0.449)	0.658	5072
Birth 4 (4+ Siblings)	−0.782	(0.668)	0.242	3257
Birth 5 (5+ Siblings)	−0.425	(0.586)	0.468	1754

Note Results based on regressions predicting parental education (years of school) with same sex composition, or with multiple births at birth *n* (conditional on having an *n*th birth). Controls not shown (see text). Standard errors are clustered by household, accompanied by 2-tailed *p* values

multiple births are negatively associated with other unobserved variables that predict child attainment, our IV estimates may be negatively biased.

The only clear theoretical rationale for these associations is finite sampling bias (sampling error), so we examine heterogeneous findings across data sources. Appendix Fig. 5 presents results from analyses using the binary additional sibling indicator. Sex composition IV estimates fluctuate across surveys, with very large positive coefficients in the NLSY79, modestly positive coefficients in the NLSCYA and PSID, and negative coefficients in the NLS. While this fluctuation is not surprising with a weaker instrument, it does highlight sensitivity to sampling error. The multiple birth IV estimates differ less across the NLSCYA and PSID. They hover around zero at all sibship size levels except the fifth, where the IV estimates are negative in both surveys but more negative in the NLSCYA.

Overall, these exercises suggest caution when interpreting our IV estimates. The parental education analyses are troubling in some respects, and there is some variation in estimates across datasets. Our most interesting finding—the negative effect of a fifth sibling—seems the most robust, however. Parental education is more weakly associated with sibship size at this transition (− 0.42 for 5+ siblings) than at the previous one (− 0.78 for 4+ siblings), and it is substantially weaker than the corresponding association with children's education (− 0.87, Table 2, reduced form, 5+ siblings). Moreover, the IV estimate of the effect of a fifth sibling is negative in both the NLSCYA and PSID, it is even more negative after controlling for family income (not shown), and this is the level at which the conditional OLS estimates are also most negative. While it would also be useful to test whether our instruments are associated with other key observables, such as maternal health, we are unable to do with these data; this is a limitation shared with other IV studies of sibship size and attainment.

Discussion

The long-held suspicion that sibship size deters children's educational attainment has faced continued scrutiny. Because parents jointly decide how many children to have and how to invest in their human capital, it is difficult to ascertain whether negative associations between sibship size and education are truly causal. We follow recent work that uses multiple births and same-sex composition as natural experiments that induce additional children in some families (Angrist et al. 2010; Black et al. 2005). Regrettably, this does not allow us to examine the influence of a first sibling (a second child), a transition many contemporary families likely weigh. But it does allow us to assess the effects of subsequent sibship size transitions, from the addition of a second sibling all the way up to a fifth. This covers a range of transitions that has been and remains relevant to many families. We also examine different stages in the attainment process (high school completion, college attendance and completion) to better understand if sibship size effects are sensitive to specific educational transitions.

Regardless of the outcome, our IV estimates do not reveal any meaningful effects of adding a second, third, or fourth sibling. If we favor the OLS estimates with rudimentary controls, which are more precise, these sibship sizes appear to modestly deter older siblings' educational attainment (less than one-fourth of a year of

school), if at all. Although many of our IV estimates are slightly more negative, they are modest and imprecisely estimated. For the second sibling's effect, estimates are positive with the same-sex composition IV and negative with the multiple birth IV. Black et al. (2010) find the same pattern in their analyses of IQ in Norway, speculating that parents better prepare for the planned addition of a child induced by sex composition preferences than for the unexpected addition of a child induced by a multiple birth. But we caution that our sex composition IV estimates are too imprecise to infer much from them, and it will take very large samples to obtain informative estimates using this method. Nevertheless, for the fertility decisions made by the majority of U.S. families, evidence suggests that the costs of additional children do not come at great expense to children already in the home.

We do, however, find robust evidence that adding a fifth sibling reduces completed schooling substantially. The OLS estimates are largest and significant at this level net of controls, and the IV estimates are significantly larger (over a one-year reduction). Although the latter are imprecise, they are statistically significant, and they appear in both datasets. This negative effect appears most consequential for college attendance, a critical transition that has substantial implications for earnings potential and job security. That we find differences across levels of education implies researchers should expect their estimates of sibship size effects to depend on their specified measure(s) of attainment. Effects would likely be muted, for instance, if we solely focused on high school or college completion and ignored the transition to college. Future work should continue to explore the extent to which the quantity-quality trade-off depends on specific levels of schooling.

The modest findings at low-to-moderate sibship sizes parallel recent studies of other countries (Angrist et al. 2010; Black et al. 2005; Frenette 2011). We echo these scholars in suggesting that the costs imposed by additional children may be absorbed by parents reducing their own consumption or increasing labor force participation—especially when additional children are planned. But how should we explain the abrupt drop-off in educational attainment upon the addition of a fifth sibling? One possibility is that there is a threshold at which resource dilution becomes unavoidable—that is, families can no longer offset the costs of additional children and the quality-quantity tradeoff emerges. Our analyses suggest this may involve resources that determine college attendance and may take hold in especially large families. Given the substantial direct financial costs of college attendance and the emerging opportunity costs of foregoing work in young adulthood, college entry represents a transition where financial resources likely become more salient. College costs are large and fixed, and families cannot make college cheaper as the number of children increases in the same way that they become more efficient in other human-capital building activities (i.e., increased parental experience). This is consistent with prior work claiming that capital-intensive investments are subject to threshold effects (Downey 1995), but we cannot tease out the mechanisms to explain why the threshold occurs at the fifth sibling. Future work could shed additional light on mechanisms by using similar methods to more rigorously examine sibship effects on potential mediating factors such as family contributions to financing college.

These IV methods have many strengths, and we feel that our analysis makes a convincing case that if sibship size reduces educational attainment, it only does so

in the upper range of U.S. family sizes. These methods also have notable limitations, however. One is that we cannot reliably estimate effects at higher sibship sizes that may further illuminate possible threshold effects. Another is that we cannot rule out differential effects for younger siblings excluded from the analyses. Although older siblings tend to fare better than younger ones on many outcomes (Black et al. 2005; Zajonc and Markus 1975), we doubt that sibship size effects are concentrated among younger siblings: families' socioeconomic resources tend to increase over time and should be more favorable to later-born children; parents may gain experience in child-rearing practices that benefit younger siblings; and older siblings may sacrifice their own pursuits to support younger siblings. Another limitation is the uncertainty in our estimates; larger samples would yield more precision, but this is only attainable with large administrative data sets not commonly accessible in the U.S. These analyses also likely lack the power to detect interaction effects. We examined differential effects by cohort, gender, birth order, and SES (parental education), but found none of the corresponding interactions to be substantively meaningful or statistically significant. While the findings here are compelling, this study leaves many questions to be answered.
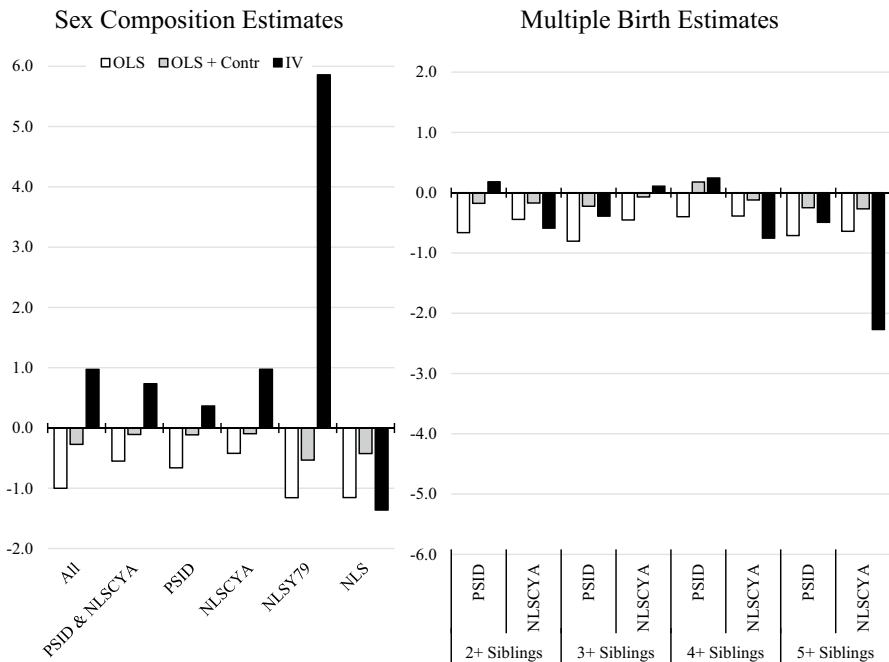
Appendix

See Fig. 5 and Table 4.

Fig. 5 Sibship effects by data source. The plot at left shows the estimates from the sex composition analyses (2+ siblings) across data sets. The plot at right shows the estimates from the multiple birth analyses at each sibship size level. Note different scales

Table 4 Analyses for years of schooling (# siblings)

	First stage	Reduced form	IV	OLS	OLS+ contr	Endog. p(IV = OLS)	Obs
2+ Siblings	0.063*	0.039	0.615	−0.313*	−0.113*	0.240	22,230
	(0.040)	(0.039)	(0.760)	(0.014)	(0.014)		
	F = 2.54						
2+ Siblings	0.573*	−0.077	−0.135	−0.290*	−0.109*	0.952	4849
	(0.114)	(0.249)	(0.432)	(0.026)	(0.027)		
	F = 25.49						
3+ Siblings	0.698*	0.043	0.061	−0.279*	−0.079*	0.747	5072
	(0.161)	(0.306)	(0.440)	(0.030)	(0.030)		
	F = 18.79						
4+ Siblings	0.897*	−0.114	−0.127	−0.213*	−0.040	0.750	3257
	(0.197)	(0.245)	(0.272)	(0.040)	(0.041)		
	F = 20.67						
5+ Siblings	0.759*	−0.867*	−1.142*	−0.262*	−0.167*	0.037*	1754
	(0.189)	(0.361)	(0.485)	(0.048)	(0.054)		
	F = 16.20						

Sibship variable specified as total number of siblings. Household-clustered standard errors in parentheses. Endogeneity test refers to the *p* value for the test of no difference between the IV estimate and the OLS estimate with controls

*$p < .05$

References

Angrist, J. D., & Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review, 88*(3), 450–477.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*(434), 444–455.

Angrist, J., Lavy, V., & Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics, 28*(4), 773–823.

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

Bagger, J., J. A. Birchenall, H. Mansour and S. Urzúa. (2013). Education, birth order, and family size. *National Bureau of Economic Research Working Paper Series* No. 19111.

Becker, G. S., & Gregg Lewis, H. (1973). Interaction between quantity and quality of children. *Journal of Political Economy, 81*(2), S279–S288.

Becker, G. S., & Tomes, N. (1976). Child endowments and quantity and quality of children. *Journal of Political Economy, 84*(4), S143–S162.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005). The more the merrier? The effect of family size and birth order on children's education. *Quarterly Journal of Economics, 120*(2), 669–700.

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

Blake, J. (1981). Family size and the quality of children. *Demography, 18*(4), 421–442.

Blake, J. (1989). Number of siblings and educational attainment. *Science, 245*(4913), 32–36.

Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure*. New York: Wiley.

Bortolus, R., Parazzini, F., Chatenoud, L., Benzi, G., Bianchi, M. M., & Marini, A. (1999). The epidemiology of multiple births. *Human Reproduction Update, 5*(2), 179–187.

Bras, H., Kok, J., & Mandemakers, K. (2010). Sibship size and status attainment across contexts: Evidence from the Netherlands, 1840–1925. *Demographic Research, 23*(4), 73–104.

Bulmer, M. G. (1970). *The Biology of twinning in man*. Oxford: Clarendon.

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

Cáceres-Delpiano, J., & Simonsen, M. (2012). The toll of fertility on mothers' wellbeing. *Journal of Health Economics, 31*(5), 752–766.

Centers for Disease Control and Prevention. (2007). American Society for Reproductive Medicine, Society for Assisted Reproductive Technology: 2005. Assisted Reproductive Technology Success Rates, National Summary and Fertility Clinic Reports: Atlanta, GA

Conley, D. (2000). sibship sex composition: Effects on educational attainment. *Social Science Research, 29*(3), 441–457.

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

David, C. (1999). The causal effect of education on earnings. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (pp. 1802–1859). Amsterdam: Elsevier.

de Haan, M. (2010). Birth order, family size and educational attainment. *Economics of Education Review, 29*(4), 576–588.

Downey, D. B. (1995). When bigger is not better: Family size, parental resources, and childrens educational performance. *American Sociological Review, 60*(5), 746–761.

Downs, B. (2003). Fertility of American Women: June 2002. Current Population Reports, *U.S. Census Bureau*

Fitzsimons, E., & Malde, B. (2014). Empirically probing the quantity-quality model. *Journal of Population Economics, 27*(1), 33–68.

Frenette, M. (2011). Why do larger families reduce parental investments in child quality, but not child quality per se? *Review of Economics of the Household, 9*(4), 523–537.

Fretts, R. C., Schmittdiel, J., Mclean, F. H., Usher, R. H., & Goldman, M. B. (1995). Increased maternal age and the risk of fetal death. *New England Journal of Medicine, 333*(15), 953–957.

Galton, F. (1874). *English men of science: Their nature and nurture*. London: Macmillan & Co.

Goldin, C., & Katz, L. F. (2009). *The Race between Education and Technology*. Cambridge: Harvard University Press.

Guo, G., & VanWey, L. K. (1999). Sibship size and intellectual development: Is the relationship causal? *American Sociological Review, 64*(2), 169–187.

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

Heer, D. M. (1985). Effects of sibling number on child outcome. *Annual Review of Sociology, 11*, 27–47.

Kao, G., & Thompson, J. S. (2003). Racial and ethnic stratification in educational achievement and attainment. *Annual Review of Sociology, 29*, 417–442.

Kulkarni, A. D., Jamieson, D. J., Jones, H. W., Kissin, D. M., Gallo, M. F., Macaluso, M., et al. (2013). Fertility treatments and multiple births in the United States. *New England Journal of Medicine, 369*(23), 2218–2225.

Lee, J. (2008). Sibling size and investment in children's education: An asian instrument. *Journal of Population Economics, 21*, 855–875.

Li, H., Zhang, J., & Zhu, Yi. (2008). The quantity-quality trade-off of children in a developing country: identification using Chinese Twins. *Demography, 45*(1), 223–243.

Li, J., Dow, W. H., & Rosero-Bixby, L. (2014). The declining effect of sibling size on children's education in Costa Rica. *Demographic Research, 31*(48), 1431–1454.

Li, J., Dowm, W. H., & Rosero-Bixby, L. (2017). Education gains attributable to fertility decline: Patterns by gender, period, and country in Latin America and Asia. *Demography, 54*, 1353–1373.

Lichtenstein, P., Olausson, P. O., & Bengt Kallen, A. J. (1996). Twin births to mothers who are twins: A registry based study. *British Medical Journal, 312*(7035), 879–881.

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

Murray, C. A. (1984). *Losing ground: American Social Policy, 1950–1980*. New York: Basic Books.

Page, E. B., & Grandon, G. M. (1979). Family configuration and mental ability: Two theories contrasted with U.S. Data. *American Educational Research Journal, 16*(3), 257–272.

Pew Research Center. (2015). Childlessness falls, family size grows among highly educated women. Washington, DC.: May.

Phillips, M. (1999). Sibship size and academic achievement: What we now know and what we still need to know. *American Sociological Review, 64*(2), 188–192.

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

Pison, G., Monden, C., & Smits, J. (2015). Twinning rates in developed countries: Trends and explanations. *Population and Development Review, 41*(4), 629–649.

Plomin, R., Owen, M. J., & McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science, 264*(5166), 1733–1739.

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

Practice Committee of the American Society for Reproductive Medicine and the Practice Committee of the Society for Assisted Reproductive Technology. (2013). Criteria for number of embryos to transfer: A committee opinion. *Fertility and Sterility, 99*(1), 44–46.

Rodgers, J. L., Harrington Cleveland, H., van den Oord, E., & Rowe, D. C. (2000). Resolving the debate over birth order, family size, and intelligence. *American Psychologist, 55*(6), 599–612.

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

Rosenzweig, M. R., & Wolpin, K. I. (1980b). Testing the quantity-quality fertility model: Use of twins as a natural experiment. *Econometrica, 48*(1), 227–240.

Rosenzweig, M. R., & Zhang, J. S. (2009). 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.

Schieve, L. A., Peterson, H. B., Meikle, S. F., Jeng, G., Danel, I., Burnett, N. M., et al. (1999). Live-birth rates and multiple-birth risk using in vitro fertilization. *Jama-Journal of the American Medical Association, 282*(19), 1832–1838.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557–586.

Steelman, L. C. (1985). A tale of two variables: A review of the intellectual consequences of sibship size and birth order. *Review of Educational Research, 55*(3), 353–386.

Steelman, L. C., Powell, B., Werum, R., & Carter, S. (2002). Reconsidering the effects of sibling configuration: Recent advances and challenges. *Annual Review of Sociology, 28*, 243–269.

Zajonc, R. B., & Markus, G s B. (1975). Birth order and intellectual development. *Psychological Review, 82*, 74.