Downloaded from by guest on April 18, 2024. Copyright 2017

The Impact of Family Composition on Educational Achievement

Stacey H. Chen
Yen-Chien Chen
Jin-Tan Liu

ABSTRACT

Parents preferring sons tend to go on having more children until a boy is born and to concentrate investment in boys for a given number of children (sibsize). Thus, having a brother may affect a child's education in two ways: an indirect effect by keeping sibsize lower and a direct rivalry effect where sibsize remains constant. We estimate the direct and indirect effects of a next brother on the first child's education conditional on potential sibsize. We address endogenous sibsize using twins. We find new evidence of sibling rivalry and gender bias that cannot be detected by conventional methods.

Stacy H. Chen is an associate professor of economics at National Graduate Institute for Policy Studies (GRIPS). Yen-Chien Chen is an associate professor of economics at National Chi-Nan University, Taiwan. Jin-Tan Liu is a professor of economics at National Taiwan University and NBER. *This is a heavily revised version of Chen, Chen and Liu (2014). Earlier versions of this paper were circulated under the following titles: "We Prefer Sons, But Does It Matter? Evidence from Matched Administrative Data from Taiwan," "Estimating the Causal Effects of Sibling Gender Composition on Child Mortality and Education Using Twin Gender Shocks," and "The Impact of Family Composition on Educational Achievement." The authors thank the Ministry of Interior Affairs and the Joint Board of the College Recruitment Commission for providing administrative data. The work benefited from feedback from Josh Angrist, John Ham, Yu-Chin Hsu, Masao Ogaki, Nancy Qian, Tyler VanderWeele, Francis Vella, and participants in universities and the 2009 NBER Education and Children Programs. The authors acknowledge financial support from JSPS (17H02537), the Academia Sinica Career Award (103-H02), Taiwan's NSC (101-2628-H-001-001-MY3), and National Health Research Institute. The data used in this article can be obtained beginning September 2019 through September 2022 from Stacey Chen at GRIPS (s-chen@grips.ac.jp).*

[Submitted September 2015; accepted June 2017]; doi: 10.3368/jhr.54.1.0915.7401R1
JEL Classification: J1, I25, and O15
ISSN 0022-166X E-ISSN 1548-8004 © 2019 by the Board of Regents of the University of Wisconsin System
Supplementary materials are freely available online at: http://uwpress.wisc.edu/journals/journals/jhr-supplementary.html

doi:10.3368/jhr.54.1.0915.7401R1

This open access article is distributed under the terms of the CC-BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0) and is freely available online at: http://jhr.uwpress.org

I. Introduction

Gender bias in families has been persisting across generations in many regions. Girls in India, for example, get weaned earlier, receive less childcare and health care, and suffer from higher mortality.[1] However, some studies find no evidence that Indian females receive less care than males under normal circumstances (Duflo 2005). Deaton (1997, 2003) suggests that equal parental spending and vaccination were carried out for both genders in the same regions. Using data from Taiwan—a society with a long tradition of preferring sons over daughters[2]—we also find surprising evidence of females having the advantage in completing high school or attaining university education.[3]

One explanation for these mixed results is that women are more enduring than males given the same care (Waldron 1983); consequently, gender differences in child outcomes often understate the degree of gender bias. For the purpose of detecting gender bias, one empirical strategy for cancelling out the endowment deficit of males is to estimate the sibling gender effect on child outcomes, while keeping the realized number of siblings (sibsize) constant. The literature on sibling rivalry (or cross-sibling competition) has adopted that strategy.[4] However, the method of keeping the realized sibsize constant does not work for households that would go on to have more children until they get a son—the so-called son-preferring stopping rule of childbearing (Yamaguchi 1989; Jensen 2005; Filmer, Friedman, and Schady 2009). Since the estimation method excludes households following the son-preferring stopping rule, previously estimated effects of gender composition of siblings might have understated the degree of gender bias.

Son preference and the son-preferring stopping rule suggest that having a brother may affect the human capital formation of a boy or a girl in two ways: an indirect effect (*IE*) by decreasing potential sibsize and a direct effect (*DE*) where potential sibsize remains constant. Under son preference, *DE* captures the rivalry effect of a son (relative to a daughter) on child outcomes, keeping other things (including potential sibsize) constant. Under the son-preferring stopping rule, *IE* captures the gain/penalty from the reduced sibsize by the presence of a brother, keeping other things (including potential sibling gender) constant. If child quality is independent of quantity, then *IE* is zero. However, if child quality decreases with quantity, then *IE* is positive. Because *DE* and *IE* may go in opposite directions, the overall impact of a brother on child outcomes can appear to be too small, particularly in countries with strong pro-son bias.[5] Therefore, understanding the relative importance of *DE* and *IE* is necessary in detecting gender discrimination.

Downloaded from by guest on April 18, 2024. Copyright 2017

1. See Chen, Huq, and D'Souza (1981); Basu (1989); Sen (1990); Ganatra and Hirve (1994); Borooah (2004); Jayachandran and Kuziemko (2011); Singh (2012); Rosenblum (2013); and Barcellos, Carvalho, and Lleras-Muney (2014).

2. Illustrations of Taiwanese preference for sons include higher death rates among girls during the first half of the 20th century (Barclay 1954) and fewer opportunities for young women to work or be educated during the postwar period (Greenhalgh 1985). For more recent evidence, see Parish and Willis (1993) and Lin, Liu, and Qian (2014).

3. Rudd (1993) finds no evidence of discrimination against females in the expenditure data of Taiwan during the early 1990s.

4. See Parish and Willis (1993); Butcher and Case (1994); Kaestner (1997); Garg and Morduch (1998); Morduch (2000); Steelman, Powell, Werum, and Carter (2002); and Lafortune and Lee (2014).

5. Vogl (2013) estimates the overall impact of a next sister on older sisters' marriage and parental coresidence, using data from South Asia where the son-preferring stopping rule is prevalent. Having a next sister not only increases family size but also increases the firstborn girl's marriage risk. Both *DE* and *IE* of a next sister on firstborn outcomes in this context are likely negative.

Downloaded from by guest on April 18, 2024. Copyright 2017

Given that the gender composition of the existing children affects fertility choice, the conventional methods of "keeping the realized sibsize constant" creates bad-control problems, which not only understate *DE* but also leave *IE* undefined (because no variation in sibsize is left for defining *IE* after the realized sibsize has been fixed). The bad-control problem is more than an issue of endogenous sibsize. Even if sibsize were uncorrelated with unobserved determinants of child outcomes, the conventional method would still bias downward because sibsize changes with sibling gender. Perhaps *IE* has been recognized and formulated in the literature, but the relative magnitude of *IE* to *DE* is still an open question.

Rather than keeping the realized sibsize constant, new methods have been proposed recently to address the bad-control problem. The most notable work comes from Barcellos, Carvalho, and Lleras-Muney (2014), who shut down the *IE* channel by restricting their sample to infants under 15 months of age. This strategy works because mothers are unable to respond to a child's gender by having more babies in such a short window. They find strong evidence of gender imbalance in receiving childcare, which is not masked by the son-preferring fertility-stopping rule.

Our objective is not to provide an explanation for female advantage in education; rather, we seek to detect gender bias in family settings, although the aggregate statistics show patterns of female advantage in education. We provide causal estimates of the *DE* and *IE* of a next brother on the education of firstborn females/males.[6] We focus on the sibling gender effects on the firstborn, not on children born later, because later-born children only exist if realized firstborn gender is such that parents continue fertility. The previous studies about family size effects have addressed the sample selection issue by estimating the effect of the later-born sibsize on the existing children.[7] This selection issue is particularly relevant for Eastern Asian countries and several South-Eastern Asian fast-growing economies, where fertility has fallen to below the replacement level. As more than 50 percent of Taiwanese families in our data have only one or two children, we address this selection issue by focusing on firstborn outcomes.

This study confronts four challenges. The first is to formulate *DE* and *IE* by overcoming the bad-control problem. Rather than using the realized sibsize to define *DE* and *IE*, we use potential sibsize because in counterfactual worlds it is possible to fix or change potential sibsize with a change in sibling gender composition. We decompose the overall impact of sibling gender on human capital accumulation into two separate causal channels: the active reallocation of parental resources along the direction of gender given potential sibsize (*DE*) and the passive effect of sibling gender through changing potential sibsize upon a change in sibling gender composition (*IE*).[8] We show both theoretically and empirically that the coefficient of sibling gender in a human capital formation model cannot be interpreted causally as the *DE* of sibling gender, even if consistently estimated. The coefficient of sibling gender could be interpreted causally only in the absence of the son-preferring fertility-stopping behavior, or in the absence

6. Our method is a simple modification of a broad literature on quality–quantity tradeoffs (for example, Black, Devereux, and Salvanes 2005, 2010; Angrist, Lavy, and Schlosser 2010).

7. For example, Black, Devereux, and Salvanes (2005, 2010); Li, Zhang, and Zhu (2008); Rosenzweig and Zhang (2009); and Angrist, Lavy, and Schlosser (2010).

8. Oaxaca's decomposition cannot work when the grouping variable (sibsize in this paper) is a mediating variable which affects outcomes (child education) and is affected by the treatment variable (sibling gender composition).

of an interaction term between sibling gender and family size in child outcomes. As expected, empirical results indicate a large difference between the coefficient of sibling gender and the average *DE*, particularly for firstborn daughters, whose family size is more responsive to the next sibling gender compared with firstborn sons.

The second challenge is to address the endogeneity of family size. While an extensive empirical literature has taken the endogeneity of family size seriously on various outcomes including child education,[9] this is less of a recognized issue in studies on the effect of gender composition of siblings. Given that the sibling gender effect works in part through its effect on family size, it is important to treat endogenous family size properly. Simply controlling for the realized sibsize in regressions of child outcomes on sibling gender can yield misleading results. Following the previous literature on family size effects on child outcomes, we address the endogeneity issue by exploiting the plausibly exogenous variation in sibsize due to twinning at the second birth, conditional on family background characteristics. Critiques of the twins instrument have noted that the tradeoff between child quality and quantity is understated (or overstated) if parents who have secondborn twins invest more (or less) in the first child than those who have a secondborn singleton, due to an endowment reinforcing (or compensating) motive. Rosenzweig and Zhang (2009) recommend a remedy for this problem by including the mean birthweight of the secondborn children. Because fetal conditions may reflect the adult risk of disease or lifestyle, relating to unobserved factors of child outcomes, we prefer not to include in regressions. Nevertheless, we discuss the results from Rosenzweig and Zhang's approach as robustness checks.

The third challenge is to address potential concerns about endogenous child gender, due to potential sex-selective abortion or recall errors. We minimize this possibility by restricting the Birth Registry data to firstborn children who were born before 1985 when abortion was not legal, and sex-testing technology was not widely available. Using the same data as ours, Lin, Liu, and Qian (2014) have noted that the sex ratio at birth for the first two parities remained within the normal range between 1980 and 1992. The sex ratio started diverging from this after 1986 but only for the third- and higher-parity births. Consistent with their findings, our test statistics show that endogeneity of child gender for the first two births is not a concern in our data.

The fourth challenge is to overcome data limitations. Typically, child outcomes could not be observed directly after infancy, and family size and sibling gender composition could only be observed partially or indirectly.[10] Using Birth Registry data for all of Taiwan since 1978, we ensure that data on the family size and sibling gender composition are complete and accurate, by tracing at least 15 years of fertility history for each mother who first gave birth before 1985. By matching Birth Registry records with University Entrance Test records, we can observe each firstborn child's educational outcomes during adolescence, in addition to complete family size and sibling gender composition.

The empirical results of our modified approach indicate that both *DE* and *IE* are near zero for firstborn males. In contrast, firstborn females have a negative direct effect and a positive indirect effect, which almost cancel each other out, the result being a near-zero total effect. This finding offers new evidence of gender bias in family settings, evidence

9. See Footnote 7.

10. One exception is the data from Norway in Black, Devereux, and Salvanes (2005, 2010), but decomposed effects of sibling gender composition are not their focus.

Downloaded from by guest on April 18, 2024. Copyright 2017

that cannot be detected by conventional methods. Conventional measures, such as gender gaps, have suggested unambiguous female advantages of completing high school and entering university. Ordinary least squares (OLS) estimates also suggest much smaller *DE* and *IE* than those constructed from our instrumental variables (IV) estimates.

This study represents the first attempt to formulate and estimate the indirect effect of sibling gender on child outcomes via a change in potential family size. The result pointing to a large and positive indirect effect has important implications for policy. If parents' ability to control their total fertility is restricted (as in the case of the two-child policy in China), the overall impact of sibling gender could be negative and much greater. Although we study a particular economy where son preference is strong, it is exactly in countries where son preference is strongest that we may expect the coexistence of *DE* and *IE,* driven by gender discrimination among siblings and son-preferring stopping rules, respectively.

As a by-product of our analysis, we find the effects of family size on child education, as well as the direction of omitted variable bias, highly depend on the gender composition of children. OLS overstates family size effects if the next sibling is male and understates them if it is female. This contrast is particularly clear among firstborn girls. If the next sibling is also female, parents' utility gain from a larger family is greater. If the next sibling is a brother rather than a sister, the son-preferring stopping rule kicks in and parents' utility gain from a larger family is smaller. The two-stage least squares (2SLS) estimates suggest that a third child in the family would lower firstborn daughters' high school completion rate or university admission rate by about one-third if the next sibling is female too. The effect of family size on firstborn males' education is also reduced by a next brother, although the estimates are imprecise and much smaller in magnitude than the effect on firstborn daughters'.

The remainder of this paper is organized as follows. Section II introduces our data sets, reports evidence of strong preference for multiple sons, and discusses the endogeneity of child gender. Section III clarifies the definitions of direct and indirect effects and describes our empirical strategies to identify them. Section IV summarizes the empirical findings, discusses the exogeneity of child gender and the twins instrument, and examines the robustness of our main results. We explore various interpretations for the patterns of effect heterogeneity. Section V concludes.

Downloaded from by guest on April 18, 2024. Copyright 2017

II. Data and Descriptive Analysis

Identifying the impact of a change in sibling gender composition on educational achievement requires a large amount of detailed data. The data should contain information about sibling gender composition of completed families and child education levels up to the late teens. To fulfill this requirement, we link two Taiwanese national administrative datasets—Birth Registry and University Entrance Test records.

A. Data

1. Sample construction

Our master data file is the Birth Registry of Taiwan since 1978 (the initial year of the digitization of the data). It contains information on each newborn child's birthweight,

birth order, birthplace (by district), parents' education levels, and everyone's exact birth date. The data also contain everyone's identifier, which allows us to link all children to mothers. We restrict our data to 965,330 mothers whose first child was born when the mother was 18 or older, prior to 4 January 1985 when the Eugenics Protection Law (which lists the cases where pregnant women could legally abort a fetus and thus opens the door to some doctors assisting the abortion of unwanted female fetuses) began to be enforced. We limit our data to fathers no younger than 18 when the first child was born and exclude observations if the birthday of the second parity is missing or if the number of births at the second parity exceeds three.

2. Completed family size and sex composition

To measure the sibling gender composition of completed families, we trace all births of the 965,330 mothers for 15–22 years after their first birth, until 1999. No mother in our data had a child in either 1998 or 1999, so the measures of completed family size and sibling gender composition can be considered accurate.[11] Taiwan has no forceful birth-control policy (such as China's one-child policy) promulgated, so son-preferring parents have no incentive to underreport a female birth, as Hull (1990) has raised as a concern for the case of China. Rather, they can keep having children until a boy is born. Thus, our data are not distorted by underreported female births.

Our treatment group is firstborn children whose next sibling is a brother ($D = 1$). In the control group, the next sibling is a sister ($D = 0$). If the second birth results in mixed-gender twins or triplets, then we randomize sibling gender using the fraction of males from the birth—rather than assign the sex ratio or the number of male siblings to D—to maintain the exclusion restriction condition for the twins instrument and to separate the effect of a next brother from the effect of family size. We set D to 1 with probability 0.5 for mixed-gender twins, probability 0.33 for triplets with one male, and probability 0.66 for triplets with two males; otherwise, we set it to 0.[12] Overall, 214,846 firstborn females and 222,915 firstborn males are in the treatment group (who have a secondborn brother), while 201,469 firstborn females and 211,814 firstborn males are in the control group (who have a secondborn sister). We discuss our results from balancing checks in Section II.D.

3. Educational outcomes

We acquired education data from the University Entrance Test records of 1996–2003 when the firstborn just turned 18. The data include two sets of tests: General Tests

11. We are able to trace and match almost all births. The fraction of births that cannot be matched is less than 1 percent (21,776/2,388,560) in the data; those are the children with no birthday information, although they satisfy all of the sample inclusion rules; that is, children whose firstborn siblings were singleton and born between 1978 and 1984, whose parents' ages are 18 or over, and whose number of siblings at the same parity is three or fewer. We also exclude five mothers who have no ID information and 60 firstborn children whose birth registry shows zero birthweight.

12. The reason we cannot use the sex ratio among the secondborn children is that the son-preferring fertility-stopping behavior is based on the number of boys, not the sex ratio. If we had used the sex ratio (0.5) for the mixed-gender twins, then we would understate treatment intensity for the male sibling in the twins. The only way to capture the correct treatment intensity, while maintaining the joint distribution of sibling gender composition, is to randomize sibling gender for mixed-gender twins.

Downloaded from by guest on April 18, 2024. Copyright 2017

(conducted in February during the high school senior year) and Union Entrance Tests (carried out in July after high school graduation). These tests offer two distinct channels for university education. Students can apply for university admission using their General Test scores and skip the tests in July. If their application results are unsatisfactory, students can forgo early admissions and take the Union Entrance Tests in July after graduation.

We construct an indicator for high school completion using "having taken general tests in February" as a proxy because most graduating seniors take the tests. Typically, the brightest high school graduates attend public universities. During our sample years (1996–2003), tuition fees in public universities were about 14 percent of average yearly family income, whereas the cost of attending private university was about 25 percent.

Although we only observe test scores among admitted students, we observe each test taker's university admission receipt. Combining this information with Birth Registry records, we construct the university admission indicator for each firstborn child being admitted to university at age 18. Our calculation of the rates of high school completion and university admission excludes vocational high schools and vocational colleges and uses the firstborn cohort size as the denominator. As Table 1 summarizes, 24 percent of firstborns from families with two or more children complete (academic) high school, and 17 percent attain university education, about the same as the education level of the overall firstborn population.

4. Parental education by category

The Birth Registry has detailed categorical information about parental education. Because the years of education on academic versus vocational tracks are not comparable, we capture the variation in parental education using five indicators: college degree or higher, professional training degree, high school diploma, vocational high school diploma, and junior high school diploma. The excluded category—primary school or lower—is the reference group.

5. The urban dummy

For groupwise comparisons, we construct an urban indicator for the five special municipalities (Taipei, New Taipei, Kaohsiung, Taichung, and Tainan), which are home to 60 percent of the total population. In most specifications, we include district fixed effects for firstborn birthplace, and the results show no difference with or without the urban dummy.

6. Distribution of families by sibsize

Table 1 summarizes the distribution of the 965,330 families by number of children (or sibsize). To causally link child education to sibling gender composition conditioned on the birth order, we focus on the education of 851,044 firstborn singletons from families with two or more children, which account for 88 percent of all families and have a sex ratio of firstborn boys to firstborn girls at birth of around 1.044.

The sex ratio of firstborn boys to firstborn girls at birth drops rapidly with sibsize because many families adopt the son-preferring fertility-stopping rule. The sex ratio

Downloaded from by guest on April 18, 2024. Copyright 2017

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 1
Firstborns' Education Achievement and Family Characteristics by Completed Number of Children

	All Families	Number of Children									
		1	2+	2	3	4	5	6	7	8	9+
Frequency	965,330	114,286	851,044	411,597	332,727	86,220	16,238	3,314	703	181	64
Percentage	100	11.8	88.2	42.6	34.5	8.9	1.7	0.3	0.07	0.02	0.01
Sex ratio (firstborn boys/girls)	1.071	1.300	1.044	1.384	0.938	0.507	0.381	0.410	0.400	0.484	0.561
High school completion	0.24	0.26	0.24	0.30	0.21	0.14	0.10	0.08	0.07	0.06	0.05
University admission	0.17	0.17	0.17	0.20	0.14	0.09	0.07	0.05	0.04	0.03	0.03
Urban (firstborn's birthplace)	0.36	0.50	0.34	0.42	0.29	0.23	0.20	0.19	0.21	0.19	0.33
Firstborn's year of birth	1981	1981	1981	1981	1981	1981	1980	1980	1980	1980	1980
Mother's year of birth	1957	1955	1957	1957	1958	1958	1958	1958	1958	1959	1959
Father's year of birth	1954	1951	1954	1953	1954	1955	1954	1954	1953	1953	1954
Mother's highest qualification											
College degree+	0.03	0.07	0.03	0.05	0.01	0.01	0.00	0.00	0.00	0.00	0.00
Professional degree	0.05	0.08	0.04	0.06	0.03	0.01	0.01	0.01	0.00	0.00	0.00
High school diploma	0.07	0.09	0.06	0.08	0.05	0.03	0.02	0.02	0.02	0.01	0.00
Vocational high school diploma	0.19	0.21	0.19	0.24	0.16	0.11	0.08	0.06	0.05	0.05	0.05
Junior high school diploma	0.25	0.21	0.26	0.24	0.28	0.26	0.25	0.23	0.21	0.15	0.23
Father's highest qualification											
College degree+	0.07	0.13	0.07	0.10	0.04	0.01	0.01	0.01	0.01	0.01	0.00
Professional degree	0.08	0.10	0.08	0.10	0.06	0.03	0.02	0.02	0.02	0.01	0.00
High school diploma	0.10	0.12	0.09	0.12	0.08	0.06	0.05	0.05	0.04	0.03	0.05
Vocational high school diploma	0.18	0.17	0.18	0.20	0.18	0.14	0.12	0.11	0.07	0.07	0.03
Junior high school diploma	0.23	0.18	0.23	0.20	0.26	0.26	0.25	0.23	0.22	0.21	0.17

Notes: This table reports descriptive statistics of characteristics of all families whose firstborn is a singleton and was born between 1978 and 1984. We exclude families with father or mother younger than 18, with missing information about child birth year, or with the number of children at some parity exceeding three.

goes from 1.38 for families with two children to 0.94 or lower for those with three or more children. Additionally, firstborns' education outcomes and their parents' socioeconomic status also decrease with sibsize. Comparing families with two children and those with three or more, we find high school completion rates dropping from 30 percent to 21 percent or lower, and university admission rates decreasing from 20 percent to 14 percent or less. Among families with two or more children, mother's and father's highest qualifications tend to decrease as the number of children rises. These statistics suggest a high correlation among family size, parents' education, and their preference for sons, as discussed further below.

B. Demand for Multiple Sons—Son-Preferring Stopping Rules

Downloaded from by guest on April 18, 2024. Copyright 2017

The Taiwanese have a long tradition of pro-male bias, for cultural and economic reasons. Confucianism—the grounding philosophy in Taiwan, Japan, Korea, and imperial China—dictates social statutes and provides rationales for the subordination of women to men, within a strict family hierarchy. According to Confucianism, family line and wealth should be transmitted from father to son, irrespective of ability, except in cases where there is no direct male line. In return, sons and their spouses assume responsibility for taking care of the parents if they are too infirm to work. In contrast, daughters move out of the family household at the time of marriage. These social norms have acted as old-age social security for the elderly for centuries in the form of extended families composed of sons (and their spouses, if married), unmarried daughters, parents, and grandparents. Although old-age social security (not based on employment) in Taiwan began in 2008, extended families (even if they do not live together) are still the primary source of support for the elderly. Thus, the demand for old-age social security is more likely to be met by having more sons.

Statistics suggest that a firstborn son reduces family size. Other observed family backgrounds being equal, firstborn males have 0.26 fewer siblings than firstborn females, as the top panel of Table 2 shows. This is over 10 percent of the average sibsize of all families (2.47). The effect of having a firstborn son on sibsize is greater among families in rural areas (Columns 4–5) or with less well-educated parents (Columns 6–9). Because a firstborn son significantly reduces the chances of having a second child, the families with a firstborn son or firstborn daughter with the same number of siblings are not comparable because the former group probably has preferences for larger families. Thus, our analysis separates 434,729 firstborn sons from 416,315 firstborn daughters out of all families with two or more children.

We report the demand for multiple sons in the middle and bottom panels of Table 2, where we estimate the effect of a change in sibling gender composition on sibsize, conditional on observed family backgrounds. Taiwanese parents not only prefer sons to daughters but also prefer multiple sons to mixed-sex composition—unlike American families, who prefer mixed-sex composition, as documented by Angrist and Evans (1998). This tendency gets stronger if the mother or the father is less educated or if the child was born in a rural area. The first two columns of Model I show that having a son, regardless of the birth order, decreases sibsize by 0.43. Because the result indicates that birth order is not important in explaining the demand for sons in our data, we further use Model II where we focus on the impact of sibling gender composition on sibsize, leaving out the factor of birth order. The estimates suggest that compared with families with two

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 2
Demand for (Multiple) Sons—Effect of Sibling Gender Composition on Sibsize

Dependent Variable = *Sibsize*	(1)	Add District Fixed Effect (2)	Add Parents' Education (3)	Birthplace		Mother's Education		Father's Education	
				Rural (4)	Urban (5)	HS– (6)	HS+ (7)	HS– (8)	HS+ (9)
All Families									
Boy1st	–0.263	–0.263	–0.263	–0.288	–0.218	–0.295	–0.201	–0.296	–0.219
	(0.002)*	(0.002)*	(0.002)*	(0.002)*	(0.003)*	(0.002)*	(0.003)*	(0.002)*	(0.002)*
Average sibsize	2.47	2.47	2.47	2.59	2.26	2.62	2.20	2.63	2.27
Adjusted *R*-squared	0.158	0.172	0.172	0.145	0.152	0.138	0.125	0.139	0.151
Number of families	965,330	965,330	965,330	614,305	351,025	635,910	329,420	550,224	415,106
Families with One or More Children									
Model I									
Boy1st	–0.428	–0.429	–0.428	–0.463	–0.362	–0.471	–0.341	–0.469	–0.371
	(0.002)*	(0.002)*	(0.002)*	(0.003)*	(0.004)*	(0.003)*	(0.003)*	(0.003)*	(0.003)*
Boy2nd	–0.429	–0.429	–0.428	–0.463	–0.363	–0.469	–0.344	–0.470	–0.370
	(0.002)*	(0.002)*	(0.002)*	(0.003)*	(0.004)*	(0.003)*	(0.003)*	(0.003)*	(0.003)*
Boy1st × *Boy2nd*	0.328	0.327	0.328	0.347	0.293	0.350	0.282	0.347	0.300
	(0.003)*	(0.003)*	(0.003)*	(0.004)*	(0.005)*	(0.004)*	(0.004)*	(0.004)*	(0.004)*
Adjusted *R*-squared	0.146	0.169	0.193	0.181	0.165	0.158	0.139	0.159	0.165

(*continued*)

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 2 (*continued*)

Dependent Variable = *Sibsize*	(1)	Add District Fixed Effect (2)	Add Parents' Education (3)	Birthplace		Mother's Education		Father's Education	
				Rural (4)	Urban (5)	HS– (6)	HS+ (7)	HS– (8)	HS+ (9)
Model II									
Mixed gender	0.101	0.101	0.101	0.116	0.070	0.120	0.060	0.122	0.070
	(0.002)*	(0.002)*	(0.002)*	(0.002)*	(0.003)*	(0.002)*	(0.002)*	(0.002)*	(0.002)*
Two girls	0.530	0.530	0.529	0.579	0.433	0.590	0.403	0.592	0.441
	(0.002)*	(0.002)*	(0.002)*	(0.003)*	(0.004)*	(0.003)*	(0.003)*	(0.003)*	(0.003)*
Average sibsize	2.67	2.67	2.67	2.76	2.51	2.79	2.42	2.81	2.49
Adjusted *R*-squared	0.146	0.169	0.193	0.181	0.165	0.158	0.139	0.159	0.165
Number of families	851,044	851,044	851,044	557,683	293,361	572,704	278,340	494,864	356,180

Note: This table reports the OLS estimated effect of sibling gender composition on family size. Other covariates included in Columns 1–3 are the full set of dummies for urban, the subject's age, parents' years of birth, and maternal age at the first birth. Columns 4–5 split the sample by the urban dummy and Columns 6–9 by the indicator for mother's or father's high school diploma. The top panel reports the estimates for all families whose firstborn singleton was born between 1978 and 1984, and the second part of the table reports the estimates for the same set of firstborns but restricted to those having at least one sibling. Robust standard errors are in parentheses. * indicates significance at the 5% level. We assume in Model II that the coefficients of *Boy1st* and *Boy2nd* are approximately equal, as confirmed by the result in Model I.

males, those with two females have about 0.53 more children, and those with mixed-sex composition still have about 0.10 more. These estimates are extraordinarily large, since they account for 20 percent and 4 percent, respectively, of the average sibsize (2.67).

Several previous studies have suggested that gender bias is stronger in rural areas (for example, Barcellos, Carvalho, and Lleras-Muney 2014). Columns 4 and 5 of Table 2 show that the son-preferring stopping behaviors are also more evident in rural areas of Taiwan. Family size reduction because of a firstborn son in rural areas is more than 30 percent (=0.288/0.218 – 1) higher than the urban counterpart. Additionally, in urban areas, parents who have a boy and a girl in the first two births have 0.07 more children than those who have two boys (standard error = 0.003). This magnitude in rural areas is more than 65 percent (=0.116/0.070 – 1) higher than the urban counterpart. While Taiwanese families in both rural and urban areas have high demand for sons, it is even higher among rural families than among urban families.

The remainder of Table 2 examines heterogeneity in demand for sons by parental education. Columns 6–9 show that family size reduction due to a firstborn son is more than 35 percent (0.295/0.201 – 1; 0.296/0.219 – 1) stronger if the mother or father has no (vocational or academic) high school diploma (HS–). Similar to the rural–urban comparison, the difference in the demand for multiple sons by parental education levels is also more evident at higher parities, as the bottom panel of the table shows. It should be noted that since parental education is higher in urban areas than in rural areas, the comparison across subpopulations does not isolate the comparative static of interest.

C. Concerning Sex-Selective Abortion

With exceedingly strong demand for multiple sons, Taiwanese families might have adopted sex-testing technologies to select child gender. Endogeneity of child gender would invalidate our empirical strategy that requires sibling gender composition at the first two births to be conditionally exogenous. This subsection examines this concern.

Although prenatal sex testing by ultrasound was introduced in Taiwan during the early 1980s, it was only after 1986 that the technology for sex testing became widely available. Using the same data source as ours, Lin, Liu, and Qian (2014, Figure 1) show that Taiwanese sex ratios at birth start being unbalanced after 1986, and the unbalanced trends are limited to singletons of third and higher parities.[13] Motivated by their results, we limit our firstborn data to those born before 1985. For these children, the sex ratio of boys to girls at birth is 1.044 (= 434,729/416,315), and the sex ratio of their next siblings at birth is between 1.053 and 1.070. Both ratios are within the range (1.04–1.08) that demographers consider normal on the basis of historical evidence (Chahnazarian 1988; Johansson and Nygren 1991).

Although the firstborn children in our data were all born before 1985, some have next siblings born after 1985 and exposed to sex-testing technology. Table 3 tests whether or not the firstborn populations whose next siblings were born before and after 1985 are different. A mean difference test in the top row suggests the sex ratio does not differ significantly across the firstborn populations whose next siblings were born earlier or

13. Although the sex ratio of secondborn children in their study is not conditional on firstborn gender, they use the same data source as ours and show statistics that support our assumption (that is, conditional exogeneity of secondborn gender).

Downloaded from by guest on April 18, 2024. Copyright 2017

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 3
Mean Education and Characteristics of Firstborn Children in Families with Two or More Children

	Firstborn Females				Firstborn Males			
		Second Birth				Second Birth		
	(1)	Before 1985 (2)	After 1985 (3)	Difference (4)	(5)	Before 1985 (6)	After 1985 (7)	Difference (8)
Sex ratio of next siblings (boys/girls)	1.066	1.070	1.062	−0.009	1.053	1.053	1.049	−0.004
High school completion	0.246	0.207	0.323	0.116*	0.239	0.210	0.296	0.086*
University admission	0.177	0.155	0.221	0.066*	0.153	0.138	0.183	0.045*
More than two children (*Morethan2*)	0.585	0.653	0.452	−0.202*	0.450	0.516	0.323	−0.193*
Complete family size (*sibsize*)	2.803	2.923	2.566	−0.357*	2.547	2.636	2.373	−0.263*
Age of mother at the 1st birth	23.71	23.54	24.03	0.495*	23.67	23.53	23.95	0.423*
Age of mother at the 2nd birth	26.22	25.40	27.83	2.422*	26.22	25.42	27.78	2.360*
Mother's year of birth	1957.2	1956.5	1958.8	2.301*	1957.3	1956.5	1958.9	2.388*
Father's year of birth	1954.0	1953.1	1955.6	2.496*	1954.0	1953.2	1955.7	2.552*
Urban	0.346	0.327	0.383	0.056*	0.343	0.325	0.378	0.053*
Mother's highest qualification								
College degree+	0.030	0.024	0.041	0.017*	0.030	0.025	0.039	0.014*
Professional degree	0.042	0.036	0.056	0.020*	0.042	0.036	0.053	0.017*
High school diploma	0.063	0.055	0.080	0.026*	0.063	0.055	0.079	0.024*
Vocational high school diploma	0.192	0.170	0.236	0.066*	0.192	0.171	0.233	0.062*
Junior high school diploma	0.258	0.242	0.291	0.050*	0.260	0.242	0.294	0.052*

Downloaded from by guest on April 18, 2024. Copyright 2017

Father's highest qualification								
College degree+	0.066	0.056	0.085	0.030*	0.065	0.057	0.082	0.026*
Professional degree	0.076	0.067	0.093	0.026*	0.076	0.069	0.089	0.021*
High school diploma	0.095	0.089	0.108	0.020*	0.094	0.088	0.107	0.019*
Vocational high school diploma	0.182	0.171	0.205	0.035*	0.183	0.171	0.205	0.034*
Junior high school diploma	0.231	0.214	0.265	0.051*	0.232	0.212	0.269	0.057*
Sample Size	416,315	276,151	140,164		434,729	287,144	147,585	

Notes: This table reports descriptive statistics of family characteristics of firstborn singletons who were born between 1978 and 1984 and had at least one sibling. Columns 2 and 6 include those whose next sibling was born before 1985, and Columns 3 and 7 include those whose next sibling was born afterwards. We exclude families with father or mother younger than 18 when their first child was born, or with missing information about child birth year, or with the number of children at some parity exceeding three. * indicates significance at the 5% level.

Table 4
Regressions of Birth Spacing (Measured in Days) between the First Two Births

	First Children Born			
			Post-1980	
Dependent Variable = Spacing in Days	1978–1984 (1)	Pre-1980 (2)	(3)	Next Sibling Born by 1985 (4)
Girl1st × *Boy2nd*	3.97	4.23	3.80	−3.28
	(2.81)	(3.42)	(4.88)	(3.01)
Girl1st	−15.51	−16.61	−13.17	−9.52
	(2.04)*	(2.48)*	(3.55)*	(2.16)*
Boy2nd	−8.13	−7.92	−8.29	1.32
	(1.98)*	(2.42)*	(3.42)*	(2.11)
Adjusted *R*-squared	0.06	0.06	0.05	0.07
Sample size	850,198	598,777	251,421	241,033

Notes: This table reports the OLS estimated coefficients for a regression of birth spacing between the first two births on their sex composition. We use the same sample as the regressions in Tables 6 to 11, except for 811 families whose second born children have missing or erroneous birthday information. We restrict the sample in Column 3 to those whose first child was born from 1980 onwards the sample in Column 4 to those whose first child was born from 1980 onwards and with the second birth prior to 1985. Additional covariates include the subject's age and district of birth, indicators for urban, parents' education and years of birth, mother's age at first birth. Robust standard errors are reported in parentheses. * indicates significance at the 5% level.

Downloaded from by guest on April 18, 2024. Copyright 2017

later. In fact, the families of firstborn children whose next siblings were born later are less male-dominated. In contrast, the rest of the table shows significant differences in characteristics between the firstborn populations with older versus younger secondborn siblings. In particular, those with younger secondborn siblings have older and more educated parents and smaller family size. This is because some parents in our firstborn data experienced the expansion of compulsory education from six to nine years during the late 1960s. Compulsory schooling expansion might have postponed the timing of marriage and childbearing.

Furthermore, we use the birth interval between the first two births to detect sex-selective abortion at the second birth. If a male birth is more likely to be selected by son-preferring parents who have a firstborn daughter, the birth spacing could be distorted by the time it takes to abort a female fetus. Regressions of the spacing between the first two births on sibling gender composition and family background factors suggest no evidence of sex-selective abortion at the second birth. As Table 4 shows, the coefficient of the interaction between the indicator for firstborn females (*Girl1st*) and the indicator for secondborn males (*Boy2nd*) is small (3.97 days) and statistically insignificant (standard error = 2.81 days), suggesting no evidence of sex-selective abortion at second birth. This estimate is similar to the result when we restrict our sample to firstborns for the pre- or post-1980 birth cohorts. Additionally, the coefficient of *Girl1st* is significantly negative,

Table 5
Balance Check—Regression of Firstborn Demographics on the Indicator for a Next Brother

	Firstborn Females		Firstborn Males	
Dependent Variable	Unconditional (1)	Control for the Other Covariates (2)	Unconditional (3)	Control for the Other Covariates (4)
Urban (firstborn birthplace)	0.0016	0.0000	–0.0003	0.0000*
Mother's highest qualification				
College degree+	0.0014*	0.0008	0.0001	0.0000
Professional degree	0.0006	0.0003	–0.0003	–0.0002
High school diploma	–0.0001	–0.0001	0.0003	0.0003
Vocational high school diploma	0.0030*	0.0017	0.0028*	0.0014
Junior high school diploma	–0.0019	0.0004	0.0006	0.0010
Father's highest qualification				
College degree+	0.0013	0.0000	0.0010	0.0012*
Professional degree	0.0010	0.0002	0.0004	0.0007
High school diploma	0.0015	0.0006	0.0009	0.0011
Vocational high school diploma	0.0012	0.0003	0.0017	0.0013
Junior high school diploma	–0.0033*	–0.0011	–0.0008	0.0002

Notes: The table reports the OLS estimated coefficients of *Boy2nd* in a regression of urban or each parental education level. Covariates not listed here are the full set of indicators for the district of firstborn's birthplace, parents' years of birth, and maternal age at the first birth. * indicates the 5% significant level.

Downloaded from by guest on April 18, 2024. Copyright 2017

indicating that with a firstborn daughter, the next child is born on average only 16 days *sooner* than if the firstborn had been a son.

All of these estimates suggest that although parents have strong demand for sons, sex-selective abortion is not a concern in our data if we control for the full set of dummies for parental education and year of birth, maternal age at the first birth, and district-level fixed effects.

Additionally, if we restrict the data to mothers who had their first two births before 1985, the estimated decomposed effects of sibling gender on the education of firstborn females are greater because parents with smaller birth spacing after a firstborn girl might have stronger preferences for sons. We report the estimates using the restricted sample in Online Appendix Tables A1–A4. Since birth spacing can be endogenous, our main analysis uses the unrestricted sample.

Downloaded from by guest on April 18, 2024. Copyright 2017

D. Balancing Tests

Table 5 implements balance tests between samples with a secondborn son versus daughter in samples split by firstborn gender. Columns 1 and 3 show the unconditional mean differences in urban versus rural or parental education by the gender of the secondborn sibling. The estimates show a small but positive association between maternal education and secondborn sons. Using the largest association in the full sample as an example, we find mothers with secondborn sons are 0.3 percentage points more likely to have a vocational high school diploma as their highest qualification than those with secondborn daughters, but this is only less than 2 percent of the sample mean (19 percent; see Table 1).

Columns 2 and 4 additionally include the other covariates, including the full set of indicators for parents' years of birth, maternal age at the first birth, and the district of the first child's birthplace. The previously estimated associations now either decrease to nearly zero or become statistically insignificant. Overall, we find little evidence that parental education or the urban dummy differs by secondborn gender if the other covariates are included. Throughout our analysis below, we include the full set of these control variables.

III. Empirical Strategy

Following the previous literature, we assume sibling gender composition for the first two births is conditionally exogenous during the sample period. Using the same data from Taiwan, recent work by Lin, Liu, and Qian (2014) has provided strong evidence supporting this assumption. As shown in their Figure 1, the fraction of males at birth was around 0.52 across the first three parities. Only after 1985 is there "a small increase in the fraction of boys for second-parity births and a dramatic increase for third- and higher-parity births." We reconfirm in Section II that evidence of sex selection is absent during our sample period (1978–1984).

Since a firstborn son considerably reduces family size, we split the firstborn population by gender throughout our analysis. We study the causal effects of having a next brother on the firstborn's education, not on the education of children born later, since having any later-born child is a parental choice that can be affected by the gender composition of existing children.

A. Defining the Decomposed Effects at Individual Levels

Consider that the first child's education Y is determined by the sex of the next sibling D and family size M, conditional on the first child's gender and observed family backgrounds.[14] Let D be the indicator for a next brother, and M_1 and M_0 denote the potential

14. In addition to the firstborn's gender, we include the full set of indicators for urban, the subject's age and district of birth, parents' education and years of birth, and mother's age at the first birth. We exclude child's initial health (for example, birthweight or gestation length) from our regressions because those are bad controls, which can be affected by child gender. Additionally, we estimate standard errors conservatively using robust standard errors. Results remain unchanged when the standard errors are calculated by clustering at the birth district level.

family size as if the next sibling were male or female, respectively, regardless of the realized sibling gender. Motivated by VanderWeele's (2013, 2014) causal inference methods, we decompose the total effect (*TE*) of sibling gender into *DE* and *IE* as follows:

$$
\begin{aligned}
(1)\quad & DE(M_1) = E[Y|M_1, D=1] - E[Y|M_1, D=0], \\
& IE(M_1, M_0) = E[Y|M_1, D=0] - E[Y|M_0, D=0], \\
& TE(M_1, M_0) = E[Y|M_1, D=1] - E[Y|M_0, D=0],
\end{aligned}
$$

We measure *DE* by fixing *potential sibsize* M_1 as if the gender of the next sibling were always male and *IE* by fixing the gender of the next sibling while comparing the potential sibsize with a different sibling gender. The sum of *DE* and *IE* is the total influence of a change in sibling gender on the firstborn.

The "controlled direct effect" *CDE* is defined by the previous studies under the assumption that sibsize can be fixed:

$$
(2)\quad CDE(M) = E[Y|M, D=1] - E[Y|M, D=0],
$$

When sibsize is fixed and cannot change with sibling gender, families cannot adjust fertility based on the gender composition of the existing children; that is, no families in the estimation can apply the son-preferring stopping rule. We call this measure a naive measure because it assumes parents' fertility choice *M* is independent of the firstborn gender. Given the son-preferring stopping rule, a female second birth increases potential sibsize; that is, M_0 tends to exceed M_1. Thus, the naive comparison does not necessarily give a right measure for *DE*.

B. The Outcome Regression and the Parameters of Interest

Both naive and proposed measures, $CDE(M)$ and $DE(M_1)$, immediately imply that the direct effect of sibling gender on firstborn outcomes is a function of family size.

To allow this possibility, we include an interaction term between family size and sibling gender in the firstborn's outcome equation. Since we run the outcome equation separately for firstborn males and females, we omit the notation of firstborn gender (and covariates) in this section for ease of exposition:

$$
\begin{aligned}
(3)\quad Y &= \beta_0 + (\beta_1 + \beta_3 M) \times D + \beta_2 M + \epsilon_{DM} \\
&= \beta_0 + \beta_1 D + \beta_2 M + \beta_3 D \times M + \epsilon_{DM},
\end{aligned}
$$

where the Greek letters are coefficients, and the error term ϵ_{DM} can vary with fertility choice *M* and sibling gender *D* to allow each firstborn to have idiosyncratic gains from a next brother or a smaller family.[15]

15. For example, in a random-coefficient model, $Y = \beta_0 + (\beta_1 + \rho_1)D + (\beta_2 + \rho_2)M + (\beta_3 + \rho_3)D \times M + \epsilon$, where ϵ is independent of *D* and *M*, ρ_1 captures parents' utility gain from having a younger son and (ρ_2, ρ_3) capture parents' utility gains from family size. Collecting all error terms yields $\epsilon_{DM} = \epsilon + \rho_1 D + \rho_2 M + \rho_3 D \times M$. This model is less restrictive than a constant-coefficient model, and it can help predict the direction of omitted variable bias (see Section III.C).

Downloaded from by guest on April 18, 2024. Copyright 2017

Under the assumption of randomized child gender, the following two properties help us to proceed: (i) ϵ_{1M} and ϵ_{0M} have the same distribution given the same family size M; (ii) the conditional mean of potential sibsize equals the conditional mean of realized sibsize, given randomized sibling gender. Using these two properties, we derive the sample analogs of the average *DE* and *IE*, as functions of the regression coefficients and the conditional means of realized sibsize, by averaging over all possible values of potential sibsize (Online Appendix 1). Again, firstborn gender is omitted for ease of exposition:

$$(4) \quad ADE = \beta_1 + \beta_3 E[M|D=1],$$

$$(5) \quad AIE = \beta_2\{E[M|D=1] - E[M|D=0]\}.$$

Their sum is the sample analog of the total effect, *ATE*. Both *ADE* and *AIE* are independent of the gender of the next sibling because both effects measure the impact of a change in the gender of the next sibling. Also, both are independent of family size because we derive both measures by averaging over all possible values of potential sibsize. For comparisons, we also estimate the sample analog of the naive comparison:

$$(6) \quad CDE = \beta_1 + \beta_3 E[M].$$

Although the conventional measures for direct effects have omitted the interaction term (by assuming $\beta_3 = 0$), we include the interaction to compare traditional methods with ours.

Definitions 4 and 6 immediately imply the following results: (i) Suppose $\beta_3 \neq 0$. Then $ADE = CDE$ if and only if $E[M|D=1] = E[M|D=0] = E[M]$. (ii) If $\beta_3 = 0$, then $ADE = CDE = \beta_1$.

Case (i) considers a general case where β_3 is not necessarily zero. The necessary and sufficient condition for gender neutrality to hold (for example, exclusion of the son-preferring stopping rule; $E[M|D] = E[M]$) is equivalent to equating the naive comparison *CDE* with the true average direct effect. As confirmed by our empirical results, the naive comparison indeed produces biased results under the son-preferring stopping rule.

Case (ii) considers a special case where we force β_3 to be zero. This functional form assumption makes the naive comparison seemingly a correct measure for the average direct effect, but purely due to the imposed restriction. Either assuming gender neutrality as in Case (i) or imposing the functional form assumption as in Case (ii) will make the naive comparison appear to be a correct measure for the average direct effect. For the purpose of allowing for more general cases, our estimation does not impose $\beta_3 = 0$ and does not assume $E[M|D=1] = E[M|D=0] = E[M]$.

Furthermore, the assumption of $\beta_3 = 0$ restricts the effect of family size to be independent of sibling gender. However, intensive empirical work has noted that family size effects may vary across firstborn genders (for example, Black, Devereux, and Salvanes 2005, Table 8; Angrist, Lavy, and Schlosser 2010, Table 8). Motivated by the previous literature, our model allows the family size effects to vary with the gender of the next sibling.

Downloaded from by guest on April 18, 2024. Copyright 2017

C. Addressing Endogeneity of Family Size

It is crucial to distinguish the family size effect from the indirect effect of the next brother. Unlike the indirect effect of a next brother through a reduction in family size, the family size effect is the impact of having one additional sibling (regardless of sibling gender) on firstborn outcomes. Although this study aims to estimate the decomposed effects of sibling gender on firstborn outcomes, we also estimate the family size effect and address the endogeneity of family size because unbiased and consistent estimates of *ADE* and *AIE* require unbiased and consistent estimates of (β_1, β_2, β_3).

In the presence of the interaction term in the outcome equation, the family size effect is specific to firstborn gender and the gender of the next sibling. Again, we run regression analysis separately for firstborn males and females so that we can omit the notation of firstborn gender for ease of exposition. The outcome Equation 3 can be rephrased as

$$Y = \beta_0 + \beta_1 D + (\beta_2 + \beta_3 D) \times M + \epsilon_{DM}. \quad (7)$$

Conditional on firstborn gender, the family size effect is β_2 given a next sister and $\beta_2 + \beta_3$ given a next brother. If the estimated coefficient of the interaction term is significantly positive ($\beta_3 > 0$), then any negative impact of one additional sibling on firstborn outcomes can be lessened by the next brother, as confirmed by our empirical result.

Problems of endogeneity arise because fertility choice M can be correlated with unobserved parental preferences, family background, and child characteristics in ϵ_{DM}. To address this, we apply 2SLS using the incidence of twins at the second birth. The second stage is the outcome Equation 3 or 7. In the first stage, we instrument M with an indicator of whether the second birth is twins ($Z = Twin2nd$), and we instrument the interaction $D \times M$ with an interaction $D \times Z$, as indicated by these two equations:

$$M = \alpha_0 + \alpha_1 D + (\alpha_2 + \alpha_3 D) \times Z + u, \quad (8)$$

$$D \times M = \gamma_0 + \gamma_1 D + (\gamma_2 + \gamma_3 D) \times Z + v. \quad (9)$$

Error terms u and v can be correlated with ϵ_{DM}. Again for ease of exposition, we suppress the gender of the firstborn and the same set of covariates as in the outcome equation. Since the decomposed effects are both linear in regression coefficients, as Equations 4 and 5 indicate, we calculate their standard errors using conventional methods (by testing for linear restrictions).[16]

We predict the direction of omitted variable bias by considering a random-coefficient model,

$$Y = \beta_0 + (\beta_1 + \rho_1) D + (\beta_2 + \rho_2) M + (\beta_3 + \rho_3) D \times M + \epsilon, \quad (10)$$

16. Most studies on sibling rivalry overlook the problem of endogenous family size. One important exception is Vogl (2013), who addresses the problem by estimating two separate reduced-form regressions of firstborn sisters' outcomes on two twin sisters relative to a singleton sister and on two twin brothers relative to a singleton brother (the combination of D and M by multiplicity). The difference in these reduced-form estimates, divided by the first stage coefficient, will provide a Wald statistic for β_3.

Downloaded from by guest on April 18, 2024. Copyright 2017

where ϵ is independent of D and M, ρ_1 captures parents' utility gain from having a younger son, and (ρ_2, ρ_3) capture parents' utility gains from family size. Collecting all error terms yields

$$\epsilon_{DM} = \epsilon + \rho_1 D + \rho_2 M + \rho_3 D \times M,$$

which is correlated with M. The OLS estimated result would overstate (understate) the magnitude of the quality–quantity tradeoff if the unobserved factor ϵ_{DM} is negatively (positively) correlated with fertility M given sibling gender composition. Under the son-preferring fertility-stopping rule, if the next sibling is female ($D = 0$), then parents' utility gain from family size ρ_2 is positive, and thus $corr(M, \epsilon_{0M}) = corr(M, \epsilon + \rho_2 M) = \rho_2 > 0$. OLS *understates* the magnitude of the quality–quantity tradeoff for firstborn outcomes given a next sister.

In contrast, if the next sibling is male ($D = 1$), then parents' utility gain from family size ($\rho_2 + \rho_3$) is negative, and thus $corr(M, \epsilon_{1M}) = \rho_2 + \rho_3 < 0$ (which implies $\rho_3 < 0$). OLS *overstates* the magnitude of the quality–quantity tradeoff for firstborn outcomes given a next brother.

Therefore, the direction of the omitted variable bias for the quality–quantity tradeoff depends on the gender composition of existing children. If the interaction term is also omitted, the bias further expands:

$$\epsilon_{DM} = \epsilon + \rho_1 D + \rho_2 M + (\beta_3 + \rho_3) D \times M.$$

Now we have $corr(M, \epsilon_{0M}) = \rho_2 > 0$ and $corr(M, \epsilon_{1M}) = \rho_2 + \rho_3 + \beta_3$. The direction of the expanded bias also depends on the gender composition of existing children. Given a next sister, 2SLS without the interaction *understates* the magnitude of the tradeoff. Given a next brother, 2SLS without the interaction can bias upward or downward, depending on the relative magnitudes of $\beta_3 > 0$ and $\rho_2 + \rho_3 < 0$.

In summary, we study how a secondborn brother affects the outcomes of the firstborn, either directly or indirectly through changing sibsize. We instrument sibsize with whether the second birth is a twin birth. Our identification strategy requires the twins instrument and the sex composition of the first two births be both conditionally exogenous. We present empirical evidence justifying these requirements in Sections IV.E and II.C, respectively. To allow fertility choice to be son-preferring, our regression model interacts sibsize with sibling gender, and it instruments this interaction by interacting the twins instrument with sibling gender in the first stage. Omitting the interaction term would lead to biased results, as the results in Section IV suggest.

IV. Empirical Results

Using data from families with at least two children, we estimate the *ADE* and *AIE* of having a next brother on the first child's education, measured by high school completion and university admission. We first present the OLS benchmark to show the importance of the included covariates and the interaction term. We implement 2SLS using twins at the second birth as an instrument for fertility choice M, measured by *Sibsize* or *Morethan2*. With *F*-statistics ranging from 488 to 12,866, the first-stage estimates suggest no concern over weak instruments, even when we include and instrument for the interaction term.

Downloaded from by guest on April 18, 2024. Copyright 2017

The second-stage results show that quality–quantity tradeoffs appear only among firstborn females whose next sibling is also female. The decomposed results suggest that the conventional measures understate the degree of gender bias, primarily because the interaction between sibsize and sibling gender is omitted and the endogeneity problem of fertility choice is overlooked. The magnitude of understatement is markedly large particularly for firstborn females' high school completion. Firstborn girls have a significant *ADE*, about 7 percent of high school completion and university admission rates. In contrast, firstborn boys' *ADE* is under 2 percent of the average value of the education outcomes.

We report empirical evidence supporting our identification assumption that both sibling gender composition and the twins instrument are conditionally exogenous. Finally, we address other specification concerns and discuss the results about effect heterogeneity.

Downloaded from by guest on April 18, 2024. Copyright 2017

A. OLS Results

Before presenting the 2SLS results, we report OLS coefficients in the outcome Regression 3 using data that will be utilized for 2SLS estimation. Table 6 shows two sets of results: one includes family composition variables (a next brother *D* and family size *M* = *Morethan2* or *Sibsize*) with no controls and the other additionally includes controls. Results in Online Appendix Table A5 using *Sibsize* as the fertility choice measure show similar patterns. In the shortest regression with sibling gender as the only regressor, the coefficient of a next brother in Columns 1 and 7 represents the unconditional mean difference in firstborn education by gender of the next sibling. In longer regressions with controls, the coefficient of a next brother in Columns 4 and 10 represents the conditional mean difference. Both conditional and unconditional mean differences are smaller than 0.4 percentage point. In particular, the conditional mean difference is small and positive for firstborn females while small and negative for firstborn males. This does not imply the absence of gender bias against girls because a negative *ADE* and a positive *AIE* of similar magnitudes might have canceled each other out.

After adding *Morethan2* to the regressions with controls, the coefficient of a next brother *D* is adjusted downward to be (more) negative. This downward adjustment is expected because the correlation between *D* and *Morethan2* is strongly negative due to the son-preferring stopping rule. A next brother induces son-preferring parents to have a smaller family and thus enables them to invest more in the first child. To allow the sibsize effect to vary with sibling gender (or the sibling gender effects to vary with sibsize), we further add the interaction between sibsize and sibling gender in the outcome regression. In the models including the interaction, the coefficients of *D* and *Morethan2* significantly decrease in magnitude with the inclusion of the control variables (for example, parents' education). This suggests that omitting parents' education and other family background variables may lead to an overstatement of the two coefficients. Contrary to what we have expected (as explained in Section III), the OLS estimated coefficient of the interaction term is negative, seemingly suggesting that firstborn children's outcomes are more hurt by a larger family if the next sibling is male. However, the endogeneity issue of fertility choice is intensified by its interaction with gender composition. The OLS estimates with or without interactions are both biased and difficult to compare/interpret.

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 6
Family Compositional Effects on the First Child's Education, OLS

	Firstborn Female						Firstborn Male					
	No Controls			Controls			No Controls			Controls		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Y = High school completion		*0.246*			*0.246*			*0.239*			*0.239*	
A next brother (*D*)	0.0037 (0.0013)*	−0.0224 (0.0014)*	−0.0189 (0.0023)*	0.0018 (0.0012)	−0.0019 (0.0013)	0.0047 (0.0022)*	−0.0018 (0.0013)	−0.0092 (0.0013)*	−0.0082 (0.0019)*	−0.0028 (0.0012)*	−0.0045 (0.0012)*	−0.0016 (0.0017)
Morethan2 (M)		−0.1183 (0.0014)*	−0.1150 (0.0022)*		−0.0170 (0.0014)*	−0.0109 (0.0021)*		−0.1161 (0.0013)	−0.1150 (0.0018)*		−0.0278 (0.0013)*	−0.0245 (0.0018)*
A next brother *(D)*×*Morethan2 (M)*			−0.0059 (0.0029)*			−0.0109 (0.0027)*			−0.0022 (0.0025)			−0.0065 (0.0024)*
Y = University Admission		*0.177*			*0.177*			*0.153*			*0.153*	
A next brother (*D*)	0.0038 (0.0012)*	−0.0159 (0.0012)*	−0.0140 (0.0021)*	0.0023 (0.0011)*	−0.0007 (0.0011)	0.0042 (0.0020)*	0.0013 (0.0011)	−0.0036 (0.0011)*	−0.0025 (0.0016)	0.0006 (0.0010)	−0.0006 (0.0010)	0.0018 (0.0015)
Morethan2 (*M*)		−0.0894 (0.0013)*	−0.0876 (0.0020)*		−0.0133 (0.0013)*	−0.0087 (0.0019)*		−0.0767 (0.0011)*	−0.0754 (0.0015)*		−0.0182 (0.0011)*	−0.0155 (0.0015)*
A next brother (*D*) × *Morethan2* (*M*)			−0.0032 (0.0026)			−0.0083 (0.0024)*			−0.0025 (0.0021)			−0.0053 (0.0021)*

Notes: This table reports the OLS results of how the firstborn's education is associated with the sex of the secondborn (*D*), having more than two children (*M*), and their interaction. We also report in Appendix Table A5 the corresponding results with *M* being measured by sibsize. We include 416,315 firstborn females and 434,729 firstborn males born between 1978 and 1984 who have at least one sibling. We control for the full set of dummies for urban, the subject's age and district of birth, parents' education and years of birth, and mother's age at the first birth. The sample mean of the educational outcome (*Y*) is in italics, and robust standard errors are in parentheses.

B. First-Stage Estimates

To address endogeneity of family size and its interaction with sibling gender, we instrument *Morethan2* using the occurrence of twins at the second birth (*Twin2nd*), and we instrument $D \times Morethan2$ using the interaction $D \times Twin2nd$. Table 7 shows a strong first stage, even after adding the interaction term. The *F*-statistic ranges from 1,372 to 27,123 with the twin instrument *Twin2nd* only and decreases to 488–12,866 after adding $D \times Twin2nd$ as an instrument for the interaction term.

The estimates in Columns 2 and 4 of the top panel suggest that a twin birth significantly increases the probability of having more than two births. Firstborn females have a smaller first-stage estimate than firstborn males because families with a firstborn daughter tend to opt for a larger family, regardless of the occurrence of twins. In particular, when the first two births are both males, the first-stage estimate is largest, at around 60 percentage points (=0.539 + 0.067), as Column 4 shows. In contrast, when the first two births are both females, the first stage is smallest, at around 32 percentage points, as Column 2 indicates. These estimates are precisely estimated, with standard errors no larger than 0.7 percentage points. In Panels 3 and 4, we also report the first-stage estimates for *Sibsize* and its interaction with a next brother *D*. The results suggest a twin birth increases the number of siblings by about 0.6–0.7 children (standard error < 0.03) for various compositions of sibling gender.

The first-stage *F*-statistic for *Morethan2* and its interaction with a next brother *D*, calculated in a manner that takes account of multiple endogenous variables (as in Angrist and Pischke 2009, p. 217–18), is between 2,046 and 12,866. The same *F*-statistic for *Sibsize* and its interaction with *D* is between 488 and 6,300, markedly lower but still outside the range where bias in 2SLS estimates could be a concern. Because the first-stage *F*-statistic for *Morethan2* is considerably larger than that for *Sibsize*, we focus our discussion on the results using *Morethan2*.

C. Second-Stage Estimates and Quality–Quantity Tradeoffs

We report the second-stage results for two educational outcomes—high school completion and university admission—at the top of Tables 8 and 9, respectively. We noted in Section III that the coefficient of a next brother could not be interpreted as his direct effect on firstborn outcomes, although it helps estimate the direct rivalry effect (see results in Section IV.D). In what follows, we explain why interactions between sibsize and sibling gender are required for unbiased estimation, discuss how family size effects change with gender composition, and relate our findings to the literature on family size effects.

Under the son-preferring fertility-stopping rule, the omission of interactions between sibsize and sibling gender is equivalent to assuming sibsize and sibling gender affect the first child's outcomes through two independent channels. This assumption is unrealistic under the stopping rule because parents' desired family size may change with the gender composition of the existing children. As a result, the magnitude of the quality–quantity tradeoff may also change with sibling gender composition.

As Columns 5 and 10 in both tables show, the 2SLS estimated coefficient of the interaction given a next brother is positive and cancels out a large part of the family size main effect. Firstborn daughters with a next brother receive almost no net family size effect, as suggested by the sum of the family size main effect and the coefficient of

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 7
First-Stage Estimates

	Firstborn Females		Firstborn Males	
	(1)	(2)	(3)	(4)
Panel 1: Dependent Variable = *Morethan2*	0.59	0.59	0.45	0.45
A next brother (*D*)	–0.219	–0.221	–0.063	–0.064
	(0.001)*	(0.001)*	(0.001)*	(0.001)*
Twin2nd	0.436	0.322	0.572	0.539
	(0.004)*	(0.005)*	(0.003)*	(0.005)*
Twin2nd × A next brother (*D*)		0.223		0.067
		(0.007)*		(0.007)*
F-statistics for twins effect	11,973	8,680	26,123	14,081
Multivariate first-stage *F*-statistic		2,046		5,892
Panel 2: Dependent Variable = *Morethan2* × A next brother (*D*)				
A next brother (*D*)	0.480	0.476	0.419	0.416
	(0.001)*	(0.001)*	(0.001)*	(0.001)*
Twin2nd	0.278	0.011	0.305	0.010
	(0.005)*	(0.003)*	(0.006)*	(0.002)*
Twin2nd × A next brother (*D*)		0.526		0.586
		(0.004)*		(0.003)*
F-statistics for twins effect	2,875	19,251	2,797	27,981
Multivariate first-stage *F*-statistic		5,110		12,866
Panel 3: Dependent Variable = *Sibsize*	2.80	2.80	2.55	2.55
A next brother (*D*)	–0.428	–0.428	–0.100	–0.101
	(0.002)*	(0.002)*	(0.002)*	(0.002)*
Twin2nd	0.622	0.616	0.720	0.711
	(0.012)*	(0.019)*	(0.010)*	(0.014)*
Twin2nd × A next brother (*D*)		0.012		0.018
		(0.024)		(0.020)
F-statistics for twins effect	2,817	1,568	5,258	2,649
Multivariate first-stage *F*-statistic		488		1,191
Panel 4: Dependent Variable = *Sibsize* × A next brother (*D*)				
A next brother (*D*)	2.596	2.592	2.498	2.493
	(0.002)*	(0.002)*	(0.001)*	(0.001)*

(*continued*)

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 7 (*continued*)

	Firstborn Females		Firstborn Males	
	(1)	(2)	(3)	(4)
Twin2nd	0.317 (0.008)*	0.015 (0.004)*	367.000 (0.010)*	0.013 (0.003)*
Twin2nd × A next brother (*D*)		0.593 (0.013)*		0.701 (0.014)*
F-statistics for twins effect	1,372	1,203	1,489	1,474
Multivariate first-stage *F*-statistic		3,472		6,300

Notes: Panels 1 and 3 of this table report the first-stage estimates for fertility choice, using twinning at the second birth (*Twin2nd*) as instrument. Panels 2 and 4 report the first-stage estimates for the interaction between a next brother *D* and fertility choice, using $D \times Twin2nd$ as instrument. An interaction term for $D \times Twin2nd$ is included in Columns 2 and 4 when the interaction between *D* and fertility choice is included in the outcome equation. We include 416,315 firstborn females and 434,729 firstborn males born between 1978 and 1984 who have at least one sibling. Additional covariates include the full set of indicators for urban, the subject's age and district of birth, parents' education and years of birth, and mother's age at the first birth. The multivariate first-stage *F*-statistics are constructed as described in Angrist and Pischke (2009, p. 217–218). Means of the fertility choice variable are in italics, and robust standard errors are in parentheses.

Downloaded from by guest on April 18, 2024. Copyright 2017

the interaction term. In contrast, given a next sister, the family size effect is startlingly strong for firstborn females. If parents go on to have a third child, then firstborn daughters with a next sister face a 10 percentage point reduction in high school completion and a 7 percentage point reduction in university admission. This accounts for nearly 40 percent of the average high school completion rate and the average university admission rate.[17] The pattern of these estimates is robust, irrespective of using *Morethan2* or *Sibsize* as the measure for family size—see Tables 10 and 11. Similar results appear among firstborn sons, although all estimates regarding the effects of family size and interactions are small and insignificant.

To investigate the magnitude of omitted variable bias in estimating the family size effects, we compare the OLS and 2SLS results in Columns 4 and 5 or Columns 9 and 10, both including interactions.[18] We find that the OLS *understates* the quality–quantity tradeoff given a next sister, but *overstates* the tradeoff given a next brother. This pattern of biases is particularly evident among firstborn females. Specifically, for firstborn daughters with a next sister, the OLS estimated impact of a third child in the family is only a 1 percentage point reduction in both high school completion and university admission. This accounts for less than 13 percent of the tradeoff suggested by the 2SLS estimates.[19] In contrast, for firstborn daughters with a next brother, OLS estimates suggest a third child in the family reduces firstborn daughters' educational outcomes by 2 percentage points, significantly greater than the zero effect that the 2SLS estimate

17. 0.0955/0.246 or 0.0669/0.177 is approximately 40 percent.
18. It is difficult to explain the change in OLS estimates from Column 1 or 6 (without interactions) to Column 4 or 9 (with interactions) because both contain omitted variables correlated with family size.
19. 0.0109/0.0955 and 0.0088/0.0669 are no more than 13 percent.

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 8
Outcome Equations and Decomposed Effects of Sibling Gender on High School Completion (with Morethan2 *as Mediating Variable)*

	Firstborn Females					Firstborn Males				
Dependent Variable = High School Completion	OLS (1)	2SLS (2)	2SLS 1st-Stage Interact (3)	OLS Interact (4)	2SLS Interact (5)	OLS (6)	2SLS (7)	2SLS 1st-Stage Interact (8)	OLS Interact (9)	2SLS Interact (10)
A Next Brother (*D*)	–0.0019	–0.0056	–0.0034	0.0047	–0.0651	–0.0045	–0.0040	–0.0040	–0.0016	–0.0128
	(0.0013)	(0.0039)	(0.0038)	(0.0020)*	(0.0261)*	(0.0012)*	(0.0015)*	(0.0015)*	(0.0016)	(0.0126)
Morethan2 (*M*)	–0.0170	–0.0342	–0.0240	–0.0109	–0.0955	–0.0278	–0.0198	–0.0198	–0.0245	–0.0307
	(0.0014)*	(0.0171)*	(0.0165)	(0.0020)*	(0.0342)*	(0.0013)*	(0.0132)	(0.0132)	(0.0018)*	(0.0203)
A Next Brother (*D*) × *Morethan2* (*M*)				–0.0109	0.0960				–0.0065	0.0193
				(0.0026)*	(0.0401)*				(0.0024)*	(0.0275)
Decomposition										
Average total effect	0.0019	0.0019	0.0019	0.0019	0.0019	–0.0028	–0.0028	–0.0028	–0.0028	–0.0028
	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)*	(0.0012)*	(0.0012)*	(0.0012)*	(0.0012)*
Average indirect effect	0.0038	0.0075	0.0053	0.0024	0.0211	0.0018	0.0013	0.0013	0.0016	0.0020
	(0.0003)*	(0.0038)	(0.0036)	(0.0004)*	(0.0075)*	(0.0001)*	(0.0008)	(0.0008)	(0.0001)*	(0.0013)
Average direct effect	–0.0019	–0.0056	–0.0034	–0.0005	–0.0192	–0.0045	–0.0041	–0.0040	–0.0043	–0.0048
	(0.0013)	(0.0039)	(0.0038)	(0.0013)	(0.0076)*	(0.0012)*	(0.0015)*	(0.0015)*	(0.0012)*	(0.0018)*

Downloaded from by guest on April 18, 2024. Copyright 2017

Controlled direct effect	–0.0019 (0.0013)	–0.0056 (0.0039)	–0.0034 (0.0038)	–0.0017 (0.0013)	–0.0089 (0.0045)*	–0.0045 (0.0012)*	–0.0041 (0.0015)*	–0.0040 (0.0015)*	–0.0045 (0.0012)*	–0.0041 (0.0015)*
Difference CDE – ADE				–0.0012 (0.0003)*	0.0102 (0.0043)*				–0.0002 (0.0001)*	0.0006 (0.0009)
Sample mean			0.246					0.239		

Notes: This table reports OLS and IV estimates for the educational outcome equation (with university admission as dependent variable) and decomposed effects of sibling gender on the outcome, with fertility choice being measured by *Morethan2*. The corresponding results for fertility choice being measured by *Sibsize* are reported in Table 11. The first-stage *F*-statistics are reported in Table 7. The samples include 416,315 firstborn females and 434,729 firstborn males, who were born between 1978 and 1984 and have at least one sibling. We exclude the interaction between the twins instrument and sibling gender from the first stage for Columns 2 and 7, while we include it for Columns 3 and 8 in order to be comparable to Columns 5 and 10. The decomposed effects are measured according to Equations 4–6 as summarized: $AIE = \beta_2\{E[M|D=1] - E[M|D=0]\}$, $ADE = \beta_1 + \beta_3 E[M|D=1]$, $CDE = \beta_1 + \beta_3 E[M]$. *AIE* and *ADE* are evaluated at the conditional mean of $M = Morethan2$: $(E[M|D=1],\ E[M|D=0]) = (0.4787,\ 0.6992)$ for firstborn females and (0.4191, 0.4831) for firstborn males, while *CDE* is evaluated at the unconditional mean $E[M] = 0.5854$ for firstborn females and 0.4503 for firstborn males. We control for the full set of dummies for urban, the subject's age and district of birth, parents' education and years of birth, and mother's age at the first birth. Robust standard errors are reported in parentheses.

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 9
Outcome Equations and Decomposed Effects of Sibling Gender on University Admission (with Morethan2 *as Mediating Variable)*

	Firstborn Females					Firstborn Males				
Dependent Variable = University Admission	OLS (1)	2SLS (2)	2SLS 1st-Stage Interact (3)	OLS Interact (4)	2SLS Interact (5)	OLS (6)	2SLS (7)	2SLS 1st-Stage Interact (8)	OLS Interact (9)	2SLS Interact (10)
A Next Brother (*D*)	–0.0007	–0.0035	–0.0020	0.0042	–0.0429	–0.0006	–0.0002	–0.0001	0.0018	–0.0122
	(0.0012)	(0.0036)	(0.0035)	(0.0018)*	(0.0235)	(0.0011)	(0.0013)	(0.0013)	(0.0014)	(0.0109)
Morethan2 (*M*)	–0.0133	–0.0263	–0.0196	–0.0088	–0.0669	–0.0182	–0.0119	–0.0112	–0.0155	–0.0261
	(0.0013)*	(0.0154)	(0.0149)	(0.0018)*	(0.0309)*	(0.0011)*	(0.0114)	(0.0114)	(0.0016)*	(0.0176)
A Next Brother (*D*) × *Morethan2* (*M*)				–0.0081	0.0636				–0.0053	0.0266
				(0.0024)*	(0.0362)				(0.0021)*	(0.0238)
Decomposition										
Average total effect	0.0023	0.0023	0.0023	0.0022	0.0023	0.0006	0.0006	0.0006	0.0006	0.0006
	(0.0011)*	(0.0011)*	(0.0011)*	(0.0011)*	(0.0011)*	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
Average indirect effect	0.0029	0.0058	0.0043	0.0019	0.0148	0.0012	0.0008	0.0007	0.0010	0.0017
	(0.0003)*	(0.0034)	(0.0033)	(0.0004)*	(0.0068)*	(0.0001)*	(0.0007)	(0.0007)	(0.0001)*	(0.0011)
Average direct effect	–0.0007	–0.0035	–0.0020	0.0003	–0.0125	–0.0006	–0.0002	–0.0001	–0.0004	–0.0011
	(0.0012)	(0.0036)	(0.0035)	(0.0012)	(0.0069)	(0.0011)	(0.0013)	(0.0013)	(0.0011)	(0.0015)

Downloaded from by guest on April 18, 2024. Copyright 2017

Controlled direct effect	–0.0007 (0.0012)	–0.0035 (0.0036)	–0.0020 (0.0035)	–0.0005 (0.0012)	–0.0057 (0.0040)	–0.0006 (0.0011)	–0.0002 (0.0013)	–0.0001 (0.0013)	–0.0006 (0.0011)	–0.0003 (0.0013)
Difference *CDE* – *ADE*				–0.0009 (0.0003)*	0.0068 (0.0039)				–0.0002 (0.0001)	0.0008 (0.0007)
Sample mean			0.177					0.153		

Notes: This table reports OLS and IV estimates for the educational outcome equation (with university admission as dependent variable) and decomposed effects of sibling gender on the outcome, with fertility choice being measured by *Morethan2*. The corresponding results for fertility choice being measured by *Sibsize* are reported in Table 11. The first-stage *F*-statistics are reported in Table 7. The samples include 416,315 firstborn females and 434,729 firstborn males, who were born between 1978 and 1984 and have at least one sibling. We exclude the interaction between the twins instrument and sibling gender from the first stage for Columns 2 and 7, while we include it for Columns 3 and 8 in order to be comparable to Columns 5 and 10. The decomposed effects are measured according to Equations 4–6 as summarized: $AIE = \beta_2\{E[M|D=1] - E[M|D=0]\}$, $ADE = \beta_1 + \beta_3 E[M|D=1]$, $CDE = \beta_1 + \beta_3 E[M]$. *AIE* and *ADE* are evaluated at the conditional mean of $M = Morethan2$: $(E[M|D=1], E[M|D=0]) = (0.4787, 0.6992)$ for firstborn females and (0.4191, 0.4831) for firstborn males, while *CDE* is evaluated at the unconditional mean $E[M] = 0.5854$ for firstborn females and 0.4503 for firstborn males. We control for the full set of dummies for urban, the subject's age and district of birth, parents' education and years of birth, and mother's age at the first birth. Robust standard errors are reported in parentheses.

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 10
Outcome Equations and Decomposed Effects of Sibling Gender on High School Completion (with Sibsize *as Mediating Variable)*

	Firstborn Females					Firstborn Males				
Dependent Variable = High School Completion	OLS (1)	2SLS (2)	2SLS 1st-Stage Interact (3)	OLS Interact (4)	2SLS Interact (5)	OLS (6)	2SLS (7)	2SLS 1st-Stage Interact (8)	OLS Interact (9)	2SLS Interact (10)
A Next Brother (*D*)	–0.0044 (0.0013)*	–0.0084 (0.0053)	–0.0083 (0.0053)	0.0133 (0.0045)*	–0.1498 (0.0717)*	–0.0051 (0.0012)*	–0.0044 (0.0016)*	–0.0044 (0.0016)*	0.0099 (0.0046)*	–0.0396 (0.0556)
Sibsize (*M*)	–0.0146 (0.0008)*	–0.0239 (0.0119)*	–0.0237 (0.0119)*	–0.0120 (0.0010)*	–0.0495 (0.0177)*	–0.0233 (0.0009)*	–0.0162 (0.0105)	–0.0161 (0.0105)	–0.0206 (0.0012)*	–0.0232 (0.0154)
A Next Brother (*D*) × *Sibsize* (*M*)				–0.0064 (0.0016)*	0.0502 (0.0254)*				–0.0059 (0.0017)*	0.0138 (0.0218)
Decomposition										
Average total effect	0.0019 (0.0012)	0.0019 (0.0012)	0.0019 (0.0012)	0.0019 (0.0012)	0.0019 (0.0012)	–0.0028 (0.0012)*	–0.0028 (0.0012)*	–0.0028 (0.0012)*	–0.0028 (0.0012)*	–0.0028 (0.0012)*
Average indirect effect	0.0063 (0.0004)*	0.0103 (0.0051)*	0.0102 (0.0051)*	0.0052 (0.0005)*	0.0213 (0.0076)*	0.0024 (0.0001)*	0.0017 (0.0011)	0.0016 (0.0011)	0.0021 (0.0001)*	0.0024 (0.0016)
Average direct effect	–0.0044 (0.0013)*	–0.0084 (0.0053)	–0.0083 (0.0053)	–0.0033 (0.0013)*	–0.0194 (0.0077)*	–0.0051 (0.0012)*	–0.0044 (0.0016)*	–0.0044 (0.0016)*	–0.0048 (0.0012)*	–0.0052 (0.0020)*

Downloaded from by guest on April 18, 2024. Copyright 2017

Controlled direct effect	–0.0044 (0.0013)*	–0.0084 (0.0053)	–0.0083 (0.0053)	–0.0046 (0.0013)*	–0.0089 (0.0053)	–0.0051 (0.0012)*	–0.0044 (0.0016)*	–0.0044 (0.0016)*	–0.0051 (0.0012)*	–0.0044 (0.0016)
Difference *CDE* – *ADE*				–0.0013	0.0105				–0.0003	0.0007
Sample mean			0.246					0.239		

Notes: This table reports OLS and IV estimates for the educational outcome equation (with high school admission as dependent variable) and decomposed effects of sibling gender on the outcome, with fertility choice being measured by *Sibsize*. The corresponding results for fertility choice being measured by *Morethan2* are reported in Table 8. The first-stage *F*-statistics are reported in Table 7. The samples include 416,315 firstborn females and 434,729 firstborn males, who were born between 1978 and 1984 and have at least one sibling. We exclude the interaction between the twins instrument and sibling gender from the first stage for Columns 2 and 7, while we include it for Columns 3 and 8 in order to be comparable to Columns 5 and 10. The decomposed effects are measured according to Equations 4–6 as summarized: $AIE = \beta_2\{E[M|D=1] - E[M|D=0]\}$, $ADE = \beta_1 + \beta_3 E[M|D=1]$, $CDE = \beta_1 + \beta_3 E[M]$. *AIE* and *ADE* are evaluated at the conditional mean of $M = Sibsize$: $(E[M|D=1], E[M|D=0]) = (2.5947, 3.0255)$ for firstborn females and (2.4975, 2.5993) for firstborn males, while *CDE* is evaluated at the unconditional mean $E[M] = 2.8032$ for firstborn females and 2.5471 for firstborn males. We control for the full set of dummies for urban, the subject's age and district of birth, parents' education and years of birth, and mother's age at the first birth. Robust standard errors are reported in parentheses.

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 11
Outcome Equations and Decomposed Effects of Sibling Gender on University Admission (with Sibsize *as Mediating Variable)*

	Firstborn Females					Firstborn Males				
Dependent Variable = High School Completion	OLS (1)	2SLS (2)	2SLS 1st-Stage Interact (3)	OLS Interact (4)	2SLS Interact (5)	OLS (6)	2SLS (7)	2SLS 1st-Stage Interact (8)	OLS Interact (9)	2SLS Interact (10)
A Next Brother (*D*)	–0.0028 (0.0012)*	–0.0056 (0.0048)	–0.0056 (0.0048)	0.0104 (0.0041)*	–0.0954 (0.0648)	–0.0010 (0.0011)	–0.0004 (0.0014)	–0.0004 (0.0014)	0.0103 (0.0040)*	–0.0520 (0.0482)
Sibsize (*M*)	–0.0119 (0.0008)*	–0.0184 (0.0108)	–0.0183 (0.0108)	–0.0100 (0.0009)*	–0.0347 (0.0160)	–0.0156 (0.0008)*	–0.0095 (0.0091)	–0.0094 (0.0091)	–0.0135 (0.0011)*	–0.0198 (0.0133)
A Next Brother (*D*) × *Sibsize* (*M*)				–0.0048 (0.0014)*	0.0319 (0.0229)				–0.0044 (0.0015)*	0.0202 (0.0189)
Decomposition										
Average total effect	0.0023 (0.0011)*	0.0023 (0.0011)*	0.0023 (0.0011)	0.0023 (0.0011)*	0.0023 (0.0011)	0.0006 (0.0011)	0.0006 (0.0011)	0.0006 0.0011	0.0006 (0.0011)	0.0006 (0.0011)
Average indirect effect	0.0051 (0.0003)*	0.0079 (0.0047)	0.0079 (0.0047)	0.0043 (0.0004)*	0.0149 (0.0069)	0.0016 (0.0001)*	0.0010 (0.0009)	0.0010 (0.0009)	0.0014 (0.0001)*	0.0020 (0.0014)
Average direct effect	–0.0028 (0.0012)*	–0.0056 (0.0048)	–0.0056 (0.0048)	–0.0020 (0.0012)*	–0.0126 (0.0070)	–0.0010 (0.0011)	–0.0004 (0.0014)	–0.0004 (0.0014)	–0.0008 (0.0011)	–0.0014 (0.0017)

Downloaded from by guest on April 18, 2024. Copyright 2017

Controlled direct effect	–0.0028 (0.0012)*	–0.0056 (0.0048)	–0.0056 (0.0048)	–0.0030 (0.0012)*	–0.0060 (0.0048)	–0.0010 (0.0011)	–0.0004 (0.0014)	–0.0004 (0.0014)	–0.0010 (0.0011)	–0.0004 (0.0014)
Difference *CDE* – *ADE*				–0.0010 (0.0003)*	0.0067 (0.0048)				–0.0002 (0.0001)*	0.0010 (0.0009)
Sample mean			0.177					0.153		

Notes: This table reports OLS and IV estimates for the educational outcome equation (with university admission as dependent variable) and decomposed effects of sibling gender on the outcome, with fertility choice being measured by *Sibsize*. The corresponding results for fertility choice being measured by *Morethan2* are reported in Table 9. The first-stage *F*-statistics are reported in Table 7. The samples include 416,315 firstborn females and 434,729 firstborn males, who were born between 1978 and 1984 and have at least one sibling. We exclude the interaction between the twins instrument and sibling gender from the first stage for Columns 2 and 7, while we include it for Columns 3 and 8 in order to be comparable to Columns 5 and 10. The decomposed effects are measured according to Equations 4–6 as summarized: $AIE = \beta_2\{E[M|D=1] - E[M|D=0]\}$, $ADE = \beta_1 + \beta_3 E[M|D=1]$, $CDE = \beta_1 + \beta_3 E[M]$. *AIE* and *ADE* are evaluated at the conditional mean of $M = Sibsize$: $(E[M|D=1], E[M|D=0]) = (2.5947, 3.0255)$ for firstborn females and (2.4975, 2.5993) for firstborn males, while *CDE* is evaluated at the unconditional mean $E[M] = 2.8032$ for firstborn females and 2.5471 for firstborn males. We control for the full set of dummies for urban, the subject's age and district of birth, parents' education and years of birth, and mother's age at the first birth. Robust standard errors are reported in parentheses.

would suggest.[20] The relationship between the direction of bias and the gender composition of existing children is consistent with what we predicted in Section III.C.

Omitted interaction bias is a special case of omitted variable bias. For the purposes of comparing estimation results using the same set of compliers (that is, those whose parents would respond to a change in the twins instrument) as in the full model, we interact *Twin2nd* and a next brother *D* in the first stage while omitting the interaction between *Morethan2* and *D* in the second stage. The 2SLS results in Columns 3 and 8 suggest that omitting interactions in the second stage leads to an *understatement* of the quality–quantity tradeoffs given a next sister, but *overstatement* given a next brother.[21] This is as we predicted in Section III.C for the general directions of omitted variable bias even if interactions are included.

In summary, the second-stage results suggest that under son preference and the son-preferring fertility-stopping rule, the family size effect strongly depends on sibling gender composition, particularly among firstborn daughters. Although not the main focus of this paper, this part of the analysis relates to the literature on the effect of family size on child outcomes or parental behaviors, using variation due to twin births or preferences for a particular gender composition. Rosenzweig and Wolpin (1980) were the first to use twins to identify family size effects, and Angrist and Evans (1998) were the first to use the gender composition of existing children to identify the family size effects on parents' labor supply. More recently, Black, Devereux, and Salvanes (2005) and Angrist, Lavy, and Schlosser (2010) find no evidence in data from Norway and Israel that an exogenous increase in the number of younger siblings (induced by twins or variation in sibling gender composition) affects adult outcomes. Black, Devereux, and Salvanes (2005) focus on younger Norwegian cohorts and find mixed evidence: no evidence that an exogenous increase in family size induced by gender composition has an impact on IQ scores, but some evidence that an increase in family size induced by twins has negative consequences. Conley and Glauber (2006) and Cáceres-Delpiano (2006) find negative family size effects on attending private school in the United States. However, while the former suggests an adverse impact of family size on grade retention, the latter shows no significant impact.

Other empirical work that exploits an exogenous increase in the number of younger siblings has used data from developing countries. Notable examples are Li, Zhang, and Zhu (2008) and Rosenzweig and Zhang (2009), both using Chinese twins, and Ponczek and Souza (2012) using Brazilian twins. These three studies from developing countries suggest a marked tradeoff between child quality and quantity. Rosenzweig and Wolpin (2000) have noted that in countries such as India where son preference is strong, gender composition might directly affect outcomes due to economies of scale, such as sharing clothes. More generally, gender composition directly affects adult outcomes, including parents' labor supply and earnings and the relationship between them.[22] Thus, child

Downloaded from by guest on April 18, 2024. Copyright 2017

20. The OLS estimates here are derived by summing the coefficients of *Morethan2* and its interaction with a next brother *D* (–0.0109–0.0109 from Table 8 and –0.0088–0.0081 from Table 9) in Column 4. The corresponding 2SLS estimates are derived using data in Column 5 (–0.0955 + 0.0960 and –0.0669 + 0.0636).

21. Example comparisons are $|-0.0240| < |-0.0955|$ given a next sister and $|-0.0240| > |-0.0955 + 0.0960|$ given a next brother.

22. See Rose (2000) using data from India. Lundberg and Rose (2002), Dahl and Moretti (2008), and Ananat and Michaels (2008) use data from the United States.

outcomes are also likely affected through channels other than variation in family size. In the following subsection, we show robust evidence of a *direct impact* of a next brother on the high school completion of firstborn daughters and sons in Taiwan. This suggests that in regions such as Taiwan where the son-preferring stopping rule is prevalent, gender composition is not a valid instrument for family size since it violates the exclusion restriction by directly affecting a child's education.

D. Main Results: Decomposed Effects of Sibling Gender

The baseline decomposed results appear at the bottom of Tables 8 and 9, summarizing the sample analogs of the total effect, indirect effect and direct effect of a next brother on the education of firstborn daughters and sons (denoted by *ATE*, *AIE*, and *ADE*). As simple statistics suggest that firstborn females are more likely than firstborn males to complete high school or enroll in university, even with markedly strong demand for sons, the *ATE* of a next brother on firstborn education also seemingly suggests no gender discrimination. For both firstborn females and firstborn males, the *ATE*s are nearly zero, although some are statistically significant.

However, a zero or positive *ATE* is likely a result of the son-preferring stopping rule being prevalent. A positive indirect effect of a younger brother—via reduced family size—might have masked the direct rivalry effect on his older sibling. Since the magnitude of the *AIE* has not been revealed before, we first describe in detail how we estimate *AIE* using the strategy introduced in Section III. For example, the 2SLS full model in Column 5 of Table 8 shows that $AIE = 0.0211$, which is derived by multiplying the effect of *Morethan2* on the firstborn daughter's high school completion given a next sister ($\beta_2 = -0.0955$) by the effect of a next brother on the probability of having *Morethan2* ($E[M|D=1] - E[M|D=0] = 0.4787 - 0.6991$). Similarly, $ADE = -0.0192$ is derived by the 2SLS estimated coefficient on a next brother ($\beta_1 = -0.0651$) being added to the product of the interaction's coefficient ($\beta_3 = 0.0960$) and the probability of *Morethan2* when the secondborn is male ($E[M|D=1] = 0.4787$). Since both *AIE* and *ADE* estimators are linear in the regression coefficients, we use a simple linear restriction to derive the standard errors.

Additionally, we use an alternative family size measure, *Sibsize*, to estimate the decomposed effects shown in Tables 10 and 11. Because the decomposed results are strikingly similar, irrespective of the choice of family size measure, and because the first stage for *Morethan2* is stronger than that for *Sibsize*, our discussion focuses on the 2SLS results using *Morethan2*.

The 2SLS estimated *AIE* suggests that through decreasing potential sibsize of firstborn daughters, a next brother indirectly causes a 2 percentage point increase in the probability of completing high school and a 1.5 percentage point increase in the likelihood of enrolling in university. The results are almost identical using either *Morethan2* or *Sibsize*, as shown in Column 5 of Tables 8 and 10 for high school completion and Tables 9 and 11 for university admission. The magnitude of *AIE* is not negligible since it accounts for 9 percent (0.0211/0.246) of the high school completion rate and 8 percent (0.0148/0.177) of the university admission rate, both precisely estimated. In contrast, the 2SLS estimates of the *AIE* on firstborn sons' high school completion and university admission are only 0.2 percentage points, accounting for 1 percent or less (0.0020/0.239 or 0.0017/0.153) of the sample mean.

Downloaded from by guest on April 18, 2024. Copyright 2017

The markedly large difference in the magnitude of the *AIE* between firstborn daughters and sons indicates the presence of gender bias, resulting primarily from two factors. First, firstborn daughters bear a much larger family size effect than firstborn sons do if the next sibling is female. If the next sibling is male, the quality–quantity tradeoffs are nearly zero for firstborn sons and daughters. Second, while a next brother decreases the probability for firstborn sons of having *Morethan2* by only 6 percentage points ($E[M|D=1]-E[M|D=0]=0.4191-0.4831$), the reduction is much higher for firstborn daughters, at about 22 percentage points (0.4787 – 0.6991). This result is not surprising because we now allow families to follow the son-preferring stopping rule in the model by interacting sibsize with sibling gender composition.

The large positive *AIE* on firstborn daughters almost entirely offsets the direct rivalry effect of male siblings, measured by *ADE*. The 2SLS estimated *ADE*s in Column 5 of Tables 8 and 9 suggest that a next brother directly reduces the probability for firstborn daughters of completing high school by 1.9 percentage points (standard error = 0.0076) and of enrolling in university by 1.3 percentage points (standard error = 0.0069). Both *ADE*s account for 7–8 percent (0.0192/0.246 or 0.0125/0.177) of the sample mean. In contrast, the 2SLS estimated *ADE* for firstborn sons, reported in Column 10, suggests that a next brother has little impact on firstborn sons' high school completion or university admission. The estimates have standard errors as small as 0.2 percentage points. We note that regardless of firstborn gender, the negative *ADE* is almost entirely offset by the positive *AIE*, resulting in a nearly zero total effect.

In the previous literature, the direct rivalry effect of male siblings was measured by the coefficient of sibling gender in regression models without interactions, and sibsize was assumed to be exogenous and independent of the gender composition of older siblings. Our OLS estimated coefficients of a next brother in Columns 1 and 6 indicate a nearly zero direct rivalry effect, regardless of firstborn gender. This coefficient is equal to *CDE* assuming no interaction term. Even if the interaction term is included, as in Columns 4 and 9, the OLS estimated *CDE* remains small or statistically insignificant. Overall, the OLS estimates show no sign of sibling rivalry, irrespective of whether the interaction is included.

The 2SLS estimated *CDE* excluding the interaction also considerably understates *ADE*, particularly among female firstborns. As Columns 2 and 3 show, without interactions, the 2SLS estimated *CDE* on firstborn daughters suggests a decrease in high school completion by 0.3–0.6 percentage points and a decrease in university admission by 0.2–0.4 percentage points. These only capture less than one-third of the 2SLS estimated *ADE* in Column 5 where the interaction term is included. This pattern does not appear in Columns 7 and 8 for firstborn sons.

Looking at Column 5, we can compare the 2SLS estimated *ADE* and *CDE* among firstborn females using the same regression model. The former is more than double the latter, and the downward bias of *CDE* is statistically significant for high school completion but imprecise for university admission. The fundamental difference between these two measures is that *ADE* is evaluated at the conditional mean of family size as if every firstborn's next sibling were male. In contrast, *CDE* is evaluated at the unconditional mean, assuming sibsize does not adjust for the gender composition of existing siblings. The gap between conditional and unconditional family size is larger among firstborn daughters because many parents follow the son-preferring stopping rule. The

Downloaded from by guest on April 18, 2024. Copyright 2017

stronger the demand for sons is, the greater the downward bias of *CDE*. A firstborn son reduces parental demand for boys, so the downward bias of *CDE* is smaller among firstborn sons.

Overall, the decomposed effects of sibling gender are larger on firstborn females than on firstborn males. As the estimates of the full 2SLS model suggest, *AIE* for firstborn females is nine or ten times (0.0148/0.0017 or 0.0211/0.0020) that for firstborn males, and *ADE* for firstborn women is at least quadruple that for firstborn males. This reveals a strong gender bias against firstborn females' education, which cannot be uncovered by conventional measures, such as *ATE* or *CDE*, or from the coefficient of sibling gender. The decomposition results are robust, regardless of which fertility choice measure (*Sibsize* or *Morethan2*) is adopted, as long as we include an interaction term in the 2SLS model.[23]

E. Testing for Exogeneity of the Twins Instrument and Robustness Checks

Conditional exogeneity of the twins instrument has been questioned because the birth of twin siblings likely has a direct effect on firstborn children beyond just increasing family size. Although conditional exogeneity of the twins instrument is not testable, this section discusses or addresses three potential concerns: (i) Twins might affect firstborn outcomes because of zero spacing between the next two siblings. (ii) Twins tend to have lower birthweight and/or poor health. (iii) Mothers who have better health or education might be positively selected into having twins.

First, a firstborn child without twin siblings tends to have a lower education if he or she has two siblings born more closely together. Assuming that we can extrapolate this tendency to the case of secondborn twins, in which the interval is zero, the birth of twins not only increases family size but also adversely affects the firstborn through spacing. If such an adverse effect of secondborn twins exists, then our 2SLS result would overstate the magnitudes of both the quantity–quality tradeoff and the average indirect effect.

The potentially negative effect of close spacing between the next two siblings on firstborn outcomes is inherently unmeasurable because most firstborn children have only one sibling. We explore the magnitude of the upward biases by including the logarithmic spacing between the first two births instead. Columns 4 and 8 of Table 12 and Online Appendix Table A7 show that 1 percent longer spacing between the first two births reduces the firstborn's high school completion or university admission by 0.6–0.9 percentage points. Although this effect is insignificant for the education of firstborn females, inclusion of birth spacing decreases the family size effects on their high school completion and university admission by approximately 2 percentage points (=0.96–

23. Online Appendix Table A6 suggests that more educated fathers (or families whose first child was born in an urban area) have stronger gender bias against firstborn daughters' education. This is seemingly contrary to their weaker demand for sons (Table 2). However, these two results are consistent with Goodkind's (1996) hypothesis, concerning the possible reduction in postnatal discrimination following prenatal sex selection. Although prenatal sex selection is minimal in our data period (as discussed in Section II.C), pro-male bias in fertility choice is manifested by the prevalence of son-preferring fertility-stopping rules. Columns 1–4 in the top panel suggest that a more educated father's strong bias against firstborn daughters' education may represent the substitution of his relatively weak son-preferring fertility choice for stronger discrimination in daughters' education investment in the future. The son-preferring stopping rule that passively discriminates against the birth of a daughter may involve reduced discrimination against daughters' education later in life.

Downloaded from by guest on April 18, 2024. Copyright 2017

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 12
Robustness Checks, Using Data from Firstborn Females

	High School Completion				University Admission			
	Baseline Estimates (1)	Add Birthweight Percentile (2)	Add Gestation Length (3)	Add Birth Spacing (4)	Baseline Estimates (5)	Add Birthweight Percentile (6)	Add Gestation Length (7)	Add Birth Spacing (8)
A next brother (*D*)	–0.0651 (0.0261)*	–0.0529 (0.0260)*	–0.0598 (0.0272)*	–0.0641 (0.0257)*	–0.0429 (0.0235)	–0.0325 (0.0235)	–0.0354 (0.0246)	–0.0420 (0.0232)
Morethan2 (*M*)	–0.0955 (0.0342)*	–0.0685 (0.0344)	–0.0963 (0.0354)*	–0.0737 (0.0333)*	–0.0669 (0.0309)*	–0.0446 (0.0311)	–0.0609 (0.0320)	–0.0652 (0.0301)*
A next brother (*D*) × *Morethan2* (*M*)	0.0960 (0.0401)*	0.0829 (0.0401)*	0.0836 (0.0419)*	0.0748 (0.0398)	0.0636 (0.0362)	0.0520 (0.0363)	0.0497 (0.0379)	0.0624 (0.0359)
Mean birthweight percentile of 2nd birth		0.0210 (0.0023)*				0.0176 (0.0021)*		
Gestational length of 2nd birth			–0.0004 (0.0007)				–0.0001 (0.0006)	
Ln (birth spacing between first 2 births)				–0.0071 (0.0044)				–0.0074 (0.0040)

Downloaded from by guest on April 18, 2024. Copyright 2017

Decomposition								
Average total effect	0.0019	0.0019	0.0014	0.0020	0.0023	0.0022	0.0018	0.0023
	(0.0012)	(0.0013)	(0.0013)	(0.0012)	(0.0011)*	(0.0011)*	(0.0011)	(0.0011)*
Average indirect effect	0.0211	0.0151	0.0212	0.0207	0.0148	0.0098	0.0134	0.0144
	(0.0075)*	(0.0076)	(0.0078)*	(0.0073)*	(0.0068)*	(0.0069)	(0.0071)	(0.0066)*
Average direct effect	–0.0192	–0.0133	–0.0199	–0.0187	–0.0125	–0.0076	–0.0116	–0.0121
	(0.0076)*	(0.0076)	(0.0079)*	(0.0074)*	(0.0069)	(0.0069)	(0.0072)	(0.0067)
Sample mean	0.246	0.248	0.247	0.246	0.177	0.178	0.178	0.177
Sample size	416,315	410,203	403,775	415,897	416,315	410,203	403,775	415,897

Notes: This table reports robustness checks on the 2SLS results for the educational outcome equation and the decomposed effects of sibling gender on firstborn females. Though not reported here, we find no evidence of nonlinearity in gestation length. Since boys are heavier than girls at birth on average, we use gender-specific birthweight percentiles and calculate the mean percentile of secondborn twins. We control implicitly for the full set of dummies for urban, the subject's age and district of birth, parents' education and years of birth, and maternal age at the first birth. Robust standard errors are reported in parentheses. Estimates for firstborn males are reported in Appendix Table A7.

0.76) or one-half of the standard error, if the next sibling is also female. The inclusion of birth spacing also decreases the decomposed effects of sibling gender on schooling by approximately 0.4 percentage (=0.0154–0.0192; 0.0173–0.0211), about one-half of the standard errors. Because families with firstborn daughters tend to rush to have another child, the impact of sibsize or sibling gender decreases after controlling for birth spacing, as expected. In contrast, the effect of 1 percent longer spacing is significant for the education of firstborn males, as Appendix Table A7 shows. However, the inclusion of birth spacing almost does not change the decomposed effects of sibling gender for firstborn males. Overall, the pattern of our estimation results remains regardless.

Second, weaker health conditions of secondborn twins may induce parents to divert family resources from the twins to the firstborn singleton (if parents have efficiency concerns) or the other way around (if parents have inequality aversion). In either case, estimation results regarding the family size effect and the decomposed effects of sibling gender will be biased. Rosenzweig and Zhang (2009) recommend controlling for the mean birthweight of secondborn twins to address the issue of twins' endowment deficit. The idea is that by fixing the birthweight of the second birth (in addition to family background), the only channel through which twinning at the second birth can affect the firstborn child's education is through changing the sibsize. However, this approach is not suitable for studies aiming to detect gender bias because boys are heavier than girls at birth on average. Part of the rivalry effect of a younger brother would be mistaken as a birthweight effect if we included the mean birthweight of the secondborn. In addition, Black, Devereux, and Salvanes (2010) note that inclusion of an initial health condition (such as birthweight or gestational duration) might invalidate conditional exogeneity of *Twin2nd* because initial health is shaped by other determinants of firstborn education, such as maternal health, the introduction of social programs, and the interplay of genes and environments (Almond, Chay, and Lee 2005; Currie 2009). Appendix Table A8 shows the secondborn twins are lighter at birth in larger families. This suggests the presence of unobserved factors (such as household income or wealth) that are positively correlated with birthweight but negatively correlated with family size. Thus, our main results exclude the endogenous birthweight of secondborn siblings.

Nevertheless, we evaluate this endogeneity issue with initial health by including the conditional mean birthweight percentile of the second birth given gender, instead of the unconditional mean birthweight, to remove the correlation between gender and birthweight. As Columns 2 and 6 of Table 12 show for firstborn females, the inclusion of the birthweight percentile decreases the estimated decomposed effects and family size effects by at least 40 percent (0.0125/0.0211 – 1 using *AIE* for high school completion as an example). In contrast, Columns 3 and 7 suggest that inclusion of gestation duration has almost no impact on the estimation results. Since we have no exogenous sources of variation in initial health, our baseline result excludes initial health status.

Third, we find that the family size impact is large among firstborn daughters whose next sibling is also female, but nearly zero or insignificant among firstborn daughters and sons whose next sibling is male. No or low quantity–quality tradeoff can be driven by positive selection of mothers into twinning, as Bhalotra and Clarke (2016) have noted. If more educated/healthier mothers have a higher chance of conceiving twins and prefer a smaller family, then the observed negative impact of family size will be biased downward.

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 13
OLS Regression of Twin2nd on Parental Education

	Dependent Variable: *Twin2nd*			
	Firstborn Female		Firstborn Male	
Observed Family Background	(1)	(2)	(3)	(4)
Panel A: Full Sample				
Sample mean	0.0070	0.0070	0.0064	0.0064
Mother's highest qualification				
College degree+	0.0019*	0.0020	0.0006	0.0003
Professional degree	0.0013	0.0014	–0.0001	–0.0006
High school diploma	0.0003	0.0004	0.0009	0.0004
Vocational high school diploma	0.0009*	0.0009*	–0.0001	–0.0005
Junior high school diploma	0.0006	0.0006	0.0001	–0.0001
Father's highest qualification				
College degree+		–0.0001		0.0006
Professional degree		–0.0001		0.0016*
High school diploma		0.0002		0.0008
Vocational high school diploma		–0.0002		0.0010*
Junior high school diploma		0.0004		0.0007*
Sample size	416,315	416,315	434,729	434,729
Adjusted *R*-square	0.00015	0.00014	0.00010	0.00012
Panel B: First Two Births Before 1985				
Sample mean	0.0065	0.0065	0.0059	0.0059
Mother's highest qualification				
College degree+	0.0024	0.0020	0.0004	0.0010
Professional degree	0.0013	0.0010	0.0007	0.0005
High school diploma	0.0007	0.0006	0.0015*	0.0013
Vocational high school diploma	0.0009	0.0008	0.0001	–0.0002
Junior high school diploma	0.0001	0.0001	0.0004	0.0002
Father's highest qualification				
College degree+		0.0006		–0.0011
Professional degree		0.0005		0.0015*
High school diploma		–0.0002		0.0007
Vocational high school diploma		–0.0003		0.0009*
Junior high school diploma		0.0003		0.0004
Sample size	276,151	276,151	287,144	287,144
Adjusted *R*-square	0.00016	0.00015	0.00022	0.00025

Notes: This table reports the OLS estimated coefficient of demographic covariates in the regression *D* or *Twin2nd* by firstborn gender, conditional on the full set of indicators for urban, the firstborn's age and district of birthplace, parents' years of birth, and maternal age at the first birth. * indicates significance at the 5% level.

Downloaded from by guest on April 18, 2024. Copyright 2017

Downloaded from by guest on April 18, 2024. Copyright 2017

Table 14
First-Stage, Reduced-Form, Second-Stage, and Decomposition Results

	Firstborn Females Dependent Variable					Firstborn Males Dependent Variable				
	More-than2	High School Completion		University Admission		*More-than2*	High School Completion		University Admission	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A next brother (*D*)	–0.221*	0.002	–0.065*	0.002	–0.429	–0.064*	–0.003*	–0.013	0.001	–0.012
Twin2nd	0.322*	–0.030*		–0.021*		0.539*	–0.016		–0.014	
Twin2nd × A next brother (*D*)	0.223*	0.029		0.019		0.067*	0.009		0.014	
Morethan2			–0.096*		–0.067*			–0.031		–0.026
Morethan2 × A next brother (*D*)			0.096*		0.064			0.019		0.027
Decomposition										
Average total effect			0.002		0.002*			–0.003*		0.001
Average indirect effect			0.021*		0.015*			0.002		0.002
Average direct effect			–0.019*		–0.013			–0.005*		–0.001
Sample mean	0.585	0.246	0.246	0.177	0.177	0.450	0.239	0.239	0.153	0.153
N	416,315	416,315	416,315	416,315	416,315	434,729	434,729	434,729	434,729	434,729
Restricted sample with close spacing ≤18 months										
A next brother (*D*)	–0.199*	0.000	–0.086	0.002	–0.111	–0.086*	–0.002	0.002	0.001	–0.027
Twin2nd	0.172*	–0.009		–0.017		–0.373*	–0.010		–0.020	
Twin2nd × A next brother (*D*)	0.204*	0.031		0.030		–0.085*	–0.008		0.014	

Downloaded from by guest on April 18, 2024. Copyright 2017

Morethan2			–0.055		–0.107			–0.027		–0.054
Morethan2 × A next brother (*D*)			0.117		0.142			–0.012		0.042
Decomposition										
Average total effect			0.000		0.017			–0.002		0.001
Average indirect effect			0.011		0.021			0.002		0.005
Average direct effect			–0.011		–0.019			–0.005		–0.003
Sample mean	0.736	0.199	0.199	0.142	0.142	0.604	0.200	0.200	0.129	0.129
N	143,975	143,975	143,975	143,975	143,975	143,975	143,724	143,724	143,724	143,724

Notes: This table reports the first-stage, reduced-form, second-stage, and decomposition results by firstborn gender. All regressions include the full set of indicators for urban, the subject's age and district of birth, parents' education and years of birth, and maternal age at the first birth. In the bottom panel, we restrict the sample to families with an interval between the first two births no more than 18 months. The samples include firstborn children who were born between 1978 and 1984 and have at least one sibling. * indicates significance at the 5% level.

Downloaded from by guest on April 18, 2024. Copyright 2017

Although our analysis included both mother's and father's education levels and birth years and their district of residence when the first child was born, in Table 13 we examine whether the observed correlation between twin births and mother's education can be captured by other family background covariates, such as father's education. We present the results using the full sample in the top panel, compared with those using a restricted sample in the bottom panel where the first two births were prior to 1985 (before the enactment of the abortion law, and before the sex ratio of boys to girls for the third birth started rising while that for the first two births remained around the normal range). Columns 1 and 3 show some positive associations between maternal education and secondborn twins, particularly in families with firstborn females. In families with firstborn males in the full sample, the association appears to be absent. After adding father's education, Columns 2 and 4 suggest that the estimated coefficients of mother's education levels on twin births become smaller or imprecise. The only exception is the coefficient of mother's vocational high school diploma among families with firstborn females; it suggests a 0.09 percentage point increase in the probability of having twins at the second birth, about 13 percent (= 0.0009/0.007) of the sample mean. As for families with firstborn males, mother's education is less important than father's education in explaining the occurrence of secondborn twins. All the coefficients of maternal education levels are tiny or negative. In contrast, father's professional degree is associated with a 0.16 percentage point increase in secondborn twinning, which accounts for a quarter (= 0.0016/0.0064) of the sample mean. Overall, the evidence for mothers positively selecting into having twins is not as clear in Taiwan as in those countries examined in Bhalotra and Clarke (2016). Our 2SLS results have controlled for both parents' education levels to minimize the downward bias caused by positive selection.

Finally, to further examine the exclusion restrictions of the twins instrument, we estimate in Table 14 the reduced-form twins effect on the education of children from families whose twins have a smaller impact on fertility. These families are likely to have a lower first stage and to have larger families anyway. The top panel of the table uses the full sample while the lower panel limits our sample to families with 18 months or less between the first two births. For comparison, Table 14 summarizes the first-stage, reduced-form, second-stage, and decomposition results for each sample (using *Morethan2* for illustration because of a stronger first stage). Columns 1 and 6 show that the restricted sample has a smaller first stage. However, the reduced-form result for that sample is either close to zero (Column 2) or imprecisely resembles the result using the full sample (Columns 4, 7, and 9).

Thus, twinning has no statistically significant impact on the education of firstborns unless it strongly changes their parents' fertility choice. This is consistent with the exclusion restrictions required for identification.

V. Conclusion

Gender bias in family settings is often masked by the practice of son-preferring fertility-stopping rules, particularly in regions where gender bias is strongest. Through stopping fertility after a son, the indirect effect of a younger brother is positive if child outcomes would have been hurt more by a larger family size potentially induced

by a younger sister. As a typical result, the younger brother effect on child outcomes appears too small, or the younger sister effect seems too large. Importantly, fertility choice is determined not only by the gender composition of existing children but also by the parents' unobserved desire for a large or small family, which affects the family size and child outcomes simultaneously. Since parents' preference for family size also changes with gender composition, unbiased estimation of younger brother effects and family size effects requires a flexible setup, such as the inclusion of interactions between family size and sibling gender composition in IV models. Using this approach, we illustrate a decomposition method to estimate the relative importance of the direct and indirect effects in an integrated framework.

Our decomposed results show that the positive indirect effect almost entirely offsets the negative direct rivalry effect on firstborn girls' high school completion and university admission, leading to a nearly zero total effect. In contrast, neither the direct nor the indirect effect of a younger brother is important in explaining firstborn sons' education outcomes. The results are robust, irrespective of which measure of family size is adopted and whether or not birth spacing or pregnancy duration is included. One important interpretation of our results is that the overall impact of sibling gender could be much greater if parents are restricted in their ability to control their total fertility. While we study a particular economy where son preference is strong, it is in regions where son preference is the strongest that we may expect the coexistence of a negative direct effect and a positive indirect effect, driven respectively by gender discrimination among children and the son-preferring stopping rule.

While family size has been taken as an exogenous control in the literature on sibling rivalry and gender bias, we address the endogeneity issues of family size particularly in the context of the son-preferring stopping rule. Since parents' desire for a smaller family is aroused by the presence of a younger son, the negative correlation between that desire and family size leads to an overstatement of the quality–quantity tradeoff (if any). By the same token, since parents' desire for a larger family is aroused by the presence of a younger daughter, the positive correlation between that desire and family size leads to understatement. The direction of omitted variable bias changes with sibling gender because of the son-preferring stopping rule. We find only firstborn females whose next sibling is also female have their educational outcomes hurt by adding another child to the family. For the other gender compositions, the estimated effect of family size is smaller and imprecise.

Our findings on family size effects are related to the quality–quantity tradeoff literature that uses sibling sex composition as an instrument for family size (Angrist and Evans 1998; Black, Devereux, and Salvanes 2005, 2010; Conley and Glauber 2006; Cáceres-Delpiano 2006; Angrist, Lavy, and Schlosser 2010). Most of these studies use data from countries where gender bias within the family is arguably small and high demand for sons is rare. It is likely that the direct effect of sibling gender on child outcomes is negligible (Huber 2015) and that the indirect effect via reduced family size is zero (assuming parents do not follow the son-preferring stopping rule). However, for data from countries where gender bias is present, our empirical findings imply that the gender of the firstborn (or the gender composition of children in general) cannot be used as an instrument for fertility since it affects child outcomes directly and violates exclusion restrictions.

Downloaded from by guest on April 18, 2024. Copyright 2017

As with any study, this paper has some limitations. First, our empirical strategy is only valid when the secondborn's gender is random. However, the randomness of the secondborn's gender may not hold in other regions or for younger cohorts where sex-selective abortions are practiced, as has been shown to be the case in India and China (Kishor and Gupta 2009; Wei and Zhang 2011). Second, using the twins instrument, we find the proportion of compliers is 32–60 percent of the firstborn population. Given that 87 percent of our sample has two or three children, the percentage of compliers in some developing countries is likely lower due to stronger demand for children.

Nevertheless, the estimated magnitude of intrafamily gender bias is important for policy. Unlike the previous evidence of gender bias mostly focusing on infant females, our results show that intrafamily gender bias has a sizable negative impact on firstborn female adolescents' education, although it is mostly indirectly offset by parents' fertility-stopping rules. As China recently started to relax the one-child policy and bring in the two-child policy, our results predict that the overall impact of a next brother (relative to a next sister) on firstborn sons and daughters would be negative because the indirect channel that would have offset the direct negative effect is shutdown.

Downloaded from by guest on April 18, 2024. Copyright 2017

References

Almond, Douglas, Kenneth Y. Chay, and David S. Lee. 2005. "The Costs of Low Birth Weight." *Quarterly Journal of Economics* 120(3):1031–83.

Ananat, Elizabeth O., and Guy Michaels. 2008. "The Effect of Marital Breakup on the Income Distribution of Women with Children." *Journal of Human Resources* 43(3):611–29.

Angrist, Joshua D., and William N. Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variations in Family Size." *American Economic Review* 88(3): 450–77.

Angrist, Joshua D., Victor Lavy, and Analia Schlosser. 2010. "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." *Journal of Labor Economics* 28(4):773–823.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Barcellos, Silvia H., Leandro S. Carvalho, and Adriana Lleras-Muney. 2014. "Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently?" *American Economic Journal: Applied Economics* 6(1):157–89.

Barclay, George W. 1954. *A Report on Taiwan's Population*. Princeton, NJ: Office of Population Research, Princeton University.

Basu, Alaka M. 1989. "Is Discrimination in Food Really Necessary for Explaining Sex Differentials in Childhood Mortality?" *Population Studies* 43(2):193–210.

Bhalotra, Sonia, and Damian Clarke. 2016. "The Twin Instrument." IZA Discussion Paper 10405. Bonn, Germany: IZA.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 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.

———. 2010. "Small Family, Smart Family? Family Size and the IQ Scores of Young Men." *Journal of Human Resources* 45(1):33–58.

Borooah, Vani K. 2004. "Gender Bias among Children in India in Their Diet and Immunisation against Disease." *Social Science & Medicin*e 58(9):1719–31.

Butcher, Kristin F., and Anne Case. 1994. "The Effect of Sibling Sex Composition on Women's Education and Earnings." *Quarterly Journal of Economics* 109(3):531–63.
Cáceres-Delpiano, Julio. 2006. "The Impacts of Family Size on Investment in Child Quality." *Journal of Human Resources* 41(4):738–54.
Chahnazarian, Anouch. 1988. "Determinants of the Sex Ratio at Birth: Review of Recent Literature." *Social Biology* 35(3–4):214–35.
Chen, Lincoln C., Emdadul Huq, and Stan D'Souza. 1981. "Sex Bias in the Family Allocation of Food and Health Care in Rural Bangladesh." *Population and Development Review* 17(1): 55–70.
Chen, Stacey H., Yen-Chien Chen, and Jin-Tan Liu. 2014. "The Impact of Family Composition on Educational Achievement." NBER Working Paper 20443. Cambridge, MA: NBER.
Conley, Dalton, and Rebecca Glauber. 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):722–37.
Currie, Janet. 2009. "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." *Journal of Economic Literature* 47(1):87–122.
Dahl, Gordon, and Enrico Moretti. 2008. "The Demand for Sons." *Review of Economic Studies* 75(4):1085–120.
Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* Baltimore, MD: Johns Hopkins University Press.
———. 2003. "Health, Inequality, and Economic Development." *Journal of Economic Literature* 41(1):113–58.
Duflo, Esther. 2005. "Gender Equality in Development." Bureau for Research in Economic Analysis of Development (BREAD) Policy Paper 011.
Filmer, Deon, Jed Friedman, and Norbert Schady. 2009. "Development, Modernization, and Childbearing: The Role of Family Sex Composition." *World Bank Economic Review* 23(3):371–98.
Ganatra, Bela R., and Siddhi S. Hirve. 1994. "Male Bias in Health Care Utilization for Under Fives in a Rural Community in Western India." *Bulletin of the World Health Organization* 72(1): 101–4.
Garg, Ashish, and Jonathan Morduch. 1998. "Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana." *Journal of Population Economics* 11(4):471–93.
Goodkind, Daniel. 1996. "On Substituting Sex Preference Strategies in East Asia: Does Prenatal Sex Selection Reduce Postnatal Discrimination?" *Population and Development Review* 22(1):111–25.
Greenhalgh, Susan. 1985. "Is Inequality Demographically Induced? The Family Cycle and the Distribution of Income in Taiwan." *American Anthropologist* 87(3):571–94.
Huber, Martin. 2015. "Testing the Validity of the Sibling Sex Ratio Instrument." *Labour* 29(1): 1–14.
Hull, Terence H. 1990. "Recent Trends in Sex Ratios at Birth in China." *Population and Development Review* 16(1):63–83.
Jayachandran, Seema, and Ilyana Kuziemko. 2011. "Why Do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India." *Quarterly Journal of Economics* 126(3):1485–538.
Jensen, Robert. 2005. "Equal Treatment, Unequal Outcomes? Generating Sex Inequality through Fertility Behavior." Unpublished.
Johansson, Sten, and Ola Nygren. 1991. "The Missing Girls of China: A New Demographic Account." *Population and Development Review* 17(1):35–51.
Kaestner, Robert. 1997. "Are Brothers Really Better? Sibling Gender Composition and Educational Achievement Revisited." *Journal of Human Resources* 32(2):250–84.

Downloaded from by guest on April 18, 2024. Copyright 2017

Kishor, Sunita, and Kamla Gupta 2009. "Gender Equality and Womens Empowerment in India. National Family Health Survey (NFHS-3), India, 2005–06." Report. Mumbai, India Mumbai, India: International Institute for Population Sciences.
Lafortune, Jeanne, and Soohyung Lee. 2014. "All for One? Family Size and Children's Educational Distribution under Credit Constraints." *American Economic Review: Papers & Proceedings* 104(5):365–9.
Li, Hongbin, Junsen Zhang, and Yi Zhu. 2008. "The Quantity–Quality Tradeoff of Children in a Developing Country: Identification Using Chinese Twins." *Demography* 45(1):223–43.
Lin, Ming-Jen, Jin-Tan Liu, and Nancy Qian. 2014. "More Missing Women, Fewer Dying Girls: The Impact of Sex Selective Abortion on Sex at Birth and Relative Female Mortality in Taiwan." *Journal of the European Economic Association* 12(4):899–926.
Lundberg, Shelly, and Elaina Rose. 2002. "The Effects of Sons and Daughters on Men's Labor Supply and Wages." *Review of Economics and Statistics* 84(2):251–68.
Morduch, Jonathan. 2000. "Sibling Rivalry in Africa." *American Economic Review* 90(2):405–9.
Parish, William, and Robert Willis. 1993. "Daughters, Education, and Family Budgets: Taiwan Experiences." *Journal of Human Resources* 28(4):863–98.
Ponczek, Vladimir, and Andre Portela Souza. 2012. "New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country." *Journal of Human Resources* 47(1):64–106.
Rose, Elaina. 2000. "Gender Bias, Credit Constraints and Time Allocation in Rural India." *Economic Journal* 110(465):738–58.
Rosenblum, Daniel. 2013. "The Effect of Fertility Decisions on Excess Female Mortality in India." *Journal of Population Economics* 26(1):147–80.
Rosenzweig, Mark R., and Kenneth I. Wolpin. 1980. "Testing the Quantity–Quality Fertility Model: The Use of Twins as a Natural Experiment." *Econometrica* 48(1):227–40.
———. 2000. "Natural 'Natural Experiments' in Economics." *Journal of Economic Literature* 38(4):827–74.
Rosenzweig, Mark R., and Junsen Zhang. 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):1149–74.
Rudd, Jeremy B. 1993. "Boy–Girl Discrimination in Taiwan: Evidence from Expenditure Data." Princeton, NJ: Research Program in Development Studies, Princeton University.
Sen, Amartya. 1990. "More than 100 Million Women Are Missing." *New York Review of Books* 37 (20), December 20.
Singh, Ashish. 2012. "Gender Based Within-Household Inequality in Childhood Immunization in India: Changes over Time and across Regions." *PLOS One* 7(4):e35045. doi:10.1371/journal.pone.0035045.
Steelman, Lala Carr, Brian Powell, Regina Werum, and Scott Carter. 2002. "Reconsidering the Effects of Sibling Configuration: Recent Advances and Challenges." *Annual Review of Sociology* 28:243–69.
VanderWeele, Tyler J. 2013. "A Three-Way Decomposition of a Total Effect into Direct, Indirect and Interactive Effects." *Epidemiology* 24(2):224–32.
———. 2014. "A Unification of Mediation and Interaction: A 4-Way Decomposition." *Epidemiology* 25(5):749–61.
Vogl, Tom S. 2013. "Marriage Institutions and Sibling Competition: Evidence from South Asia." *Quarterly Journal of Economics* 128(3):1017–72.
Waldron, Ingrid. 1983. "Sex Differences in Human Mortality: The Role of Genetic Factors." *Social Science and Medicine* 17(6):321–33.
Wei, Shang-Jin, and Xiaobo Zhang, 2011. "The Competitive Saving Motive: Evidence from Rising Sex Ratios and Savings Rates in China." *Journal of Political Economy* 119(3):511–64.
Yamaguchi, Kazuo. 1989. "A Formal Theory for Male-Preferring Stopping Rules of Childbearing: Sex Differences in Birth Order and the Number of Siblings." *Demography* 26(3):451–65.

Downloaded from by guest on April 18, 2024. Copyright 2017