The Only Child[*]

Julius Ilciukas[†] Petter Lundborg[§]

Erik Plug[‡] Astrid Würtz Rasmussen[¶]

July 2025

Abstract

We estimate the impact of having siblings on the cognitive and non-cognitive development of first-born children. By exploiting quasi-experimental variation in the sequence of parents' in vitro fertilization (IVF) attempts for a second child, we identify effects for first-born children who would remain only children if all IVF attempts fail. Using Danish administrative registers linked to nationwide school surveys, we find that siblings have little effect on test performance (math and reading), personality traits (conscientiousness, agreeableness, emotional stability), or classroom and school well-being. Overall, we conclude that having siblings neither harms nor spurs the first-borns' cognitive and non-cognitive development.

[*]We thank seminar/conference participants in Aarhus, Amsterdam, Bergen, Bologna, Brno, Berlin, Cologne, Copenhagen, Crieff (Schotland), Edinburgh, Laax (Switzerland), Lausanne, Linz, Liverpool, Lund, Luxembourg, Rotterdam, Sigtuna, Venice, and Wiesbaden for their comments and suggestions. Data in this article are, in part, taken from IVF registers, held by the Danish Health Data Authority (DHDA). DHDA is not responsible for the analysis or interpretation of the data presented. Financial support from the Institute for Evaluation of Labour Market and Education Policy (IFAU), Swedish Research Council for Health, Working Life and Welfare (FORTE), and the Dutch Research Council (NWO) is gratefully acknowledged. The authors declare that they have no relevant or material financial interest that relate to the research described in this paper.

[†]Amsterdam School of Economics, University of Amsterdam. Email: j.ilciukas@uva.nl

[‡]Amsterdam School of Economics, University of Amsterdam. Email: e.j.s.plug@uva.nl

[§]Department of Economics, Lund University. Email: petter.lundborg@nek.lu.se

[¶]Department of Economics and Business Economics, Aarhus University. Email: awr@econ.au.dk

1 Introduction

With collapsing fertility rates worldwide, families with only one child are becoming increasingly common (Bhattacharjee & others., 2024). In the US, for instance, the share of one-child families more than doubled, from 10 percent in 1976 to 22 percent in 2015 (Pew Research Center, 2015). In China, which introduced its one-child policy in 1979, 70 percent of all urban families with children in 2005 were one-child families (National Bureau of Statistics, 2007). And in Denmark, which is the focus country in our study, the share of one-child families is nowadays 21 percent (Statistics Denmark, 2023).[1] Because there are more only children than ever before, we believe it is important to call attention to these children and learn more about their cognitive and non-cognitive development.

In this study, we examine how only children fare in primary and secondary school and explore how they compare to children with siblings. Specifically, we ask whether first-born children (aged 7 to 16) perform better or worse in school depending on whether they have siblings. This question is not so easy to answer, however. On theoretical grounds, any outcome is possible. In Becker's quantity-quality model, children without siblings benefit from undivided parental time and money resources and receive greater investment in their cognitive and non-cognitive development (Becker, 1960; Becker & Lewis, 1973; Becker & Tomes, 1986). In contrast, the confluence model developed by Zajonc & Markus (1975), and later expanded by Zajonc (1976), emphasizes the importance of sibling interactions in shaping development. According to this model, only children, lacking younger siblings, miss out on valuable opportunities for social and cognitive interactions that spur development.

On empirical grounds, existing evidence is limited. Most studies directly compare the cognitive and non-cognitive outcomes of children with and without siblings, which

[1] Several factors help explain the rise of one-child families, including increased divorce rates, which interrupted initial fertility plans; improved educational and labor market opportunities for women, which raised the opportunity costs of childrearing; delayed motherhood, which shortened fertility windows; greater access to modern contraceptives, which enhanced reproductive control; and invasive one-child policies, which limited family size by law. Illustrative studies supporting these explanations include Blake (1985), Fong (2004), Lundberg & Pollak (2007), Feyrer et al. (2008), and Miller (2010). For a comprehensive review on the economics of declining fertility rates, see Doepke et al. (2023) and Goldin (2024).

is not sufficient to make causal claims.[2] For this, we need a natural experiment that generates variation in the likelihood of having siblings that is as good as random.

The natural experiment we exploit is based on in vitro fertilization (IVF) treatments in Denmark. In particular, we study one-child families who undergo IVF in an effort to have a second child and use success at the first IVF attempt (conditional on the parents' age and year of treatment) as a plausibly exogenous source of variation in whether the first-born child gains a sibling. Following Lundborg et al. (2017, 2024), we use this variation as an instrument to estimate the causal effect of having siblings on cognitive and non-cognitive outcomes.[3] For our analyses, we link nationwide school survey data with rich administrative records to construct a sample of first-born children whose parents initiated IVF for a second child.

Like many other studies on only children, we find negative correlations between sibling status and the cognitive and non-cognitive outcomes of first-born children. Compared to those with siblings, only children score lower on math and reading tests, are more neurotic, less conscientious and agreeable, and report lower levels of happiness. These correlations are all sizeable and statistically significant. Once we move from correlations to causal estimates, however, we find that siblings no longer have a significant impact on these outcomes. While our estimates are all statistically insignificant, they are in most cases not precise enough to rule out sizable effects. One possible explanation is that our IV-IVF strategy identifies the sibling effect for a relatively small group of first-born children. These are the *compliers* in our setting and consist of those first-born children who would remain only children if their parents' *first* IVF attempt fails. Most first-born children in our sample, however, go on to have a sibling after a first failed attempt.

Our solution involves the identification of sibling effects for a larger group of first-born children. Following Ilciukas (2024), we exploit quasi-experimental variation in both first and later IVF attempts, adapt our IV-IVF strategy, and identify effects for those first-born children who would remain only children not only if the first attempt fails but also if later attempts fail. We refer to these children as *reliers* as their parents

[2] This is also what Falbo (2012) concludes in one of her only-child reviews: "if we find differences in the outcomes between only children and those with siblings we should be aware that many other factors contribute to these differences, not just the lack of siblings" (Falbo, 2012, p. 47).

[3] Lundborg et al. (2017) introduced this IV-IVF strategy, though in a different context: they use success at the first IVF cycle among childless couples to estimate the causal effect of having children on the careers of women (and men).

rely on IVF for a second child. To make the idea of a larger group of first-born children more concrete, we use the quasi-random success of later IVF attempts to identify the potential outcomes of first-born children with first-attempt IVF siblings in the hypothetical scenario where their parents, after a first failed attempt, continued to experience failures and eventually stopped treatment. The larger group of first-born reliers will, all else equal, translate into more precise estimates and reduce concerns about possible exclusion violations. Using the improved IV-IVF strategy, we continue to find that siblings have no impact on the cognitive and non-cognitive outcomes of first-born children. Moreover, the improved estimates are all close to zero, statistically insignificant, and sufficiently precise to rule out any meaningful effects.

The last question we ask is whether our results generalize to first-born children raised in more representative families. There are several reasons why sibling effects might be weaker in IVF-treated families: they are richer and less bound by financial constraints; they are more educated and better informed on how to raise multiple children; and they express a clear demand for children and, as such, avoid any adverse impact of unwanted siblings. When we zoom in on those first-borns raised in families that more closely resemble representative families (in being less affluent, less educated, and younger), we continue to find small and statistically insignificant sibling effects. While we recognize that IVF and representative families may differ in many observable and unobservable ways, we find little evidence that sibling effects on first-borns vary systematically along the observable dimensions we examine. At the same time, we also recognize that, over time, declining fertility rates and increasing maternal age at first birth will make representative families more comparable to IVF-treated families. In this sense, the sibling effects obtained in IVF families may be increasingly informative for a broader group of families.

Our study relates to several strands of literatures. The literature most closely related to our study exploits China's one-child policy as a natural experiment to estimate the effect of having siblings (H. Li et al., 2008; Rosenzweig & Zhang, 2009; Qian, 2009; Cameron et al., 2013; Liu, 2014; B. Li & Zhang, 2017; Guo et al., 2020; Xiao, 2024). These studies provide mixed results, possibly reflecting differences in how the policy was implemented. China's one-child policy varied across regions and time, differed in enforcement intensity, and included exemptions for certain ethnic groups and families with twins. Using different sources of identifying variation, the estimated sibling effects reported in these studies need not be the same. Regardless, our study differs in that we examine a highly developed country where parents are less reliant on children for old-age support and face weaker resource constraints, both of which could be crucial in shaping siblings effects through parental investments. Moreover, our results may be more generalizable, as the one-child policy reshaped family dynamics and the broader social environment, potentially limiting the relevance of these estimates in settings without such restrictions.

More broadly, our study relates to the larger sibling literature that estimates the causal effect of sibship size on child outcomes using twin births and sibling sex composition as natural experiments. With twin births, researchers exploit the fact that some parents end up with more children than planned; they identify sibship size effects by comparing outcomes of children with younger twin siblings and younger singleton siblings. With sibling sex composition, researchers exploit that parents with same-sex children are more inclined to have another child; they identify sibship size effects by comparing outcomes of same-sex and opposite-sex siblings. While some studies find that having more siblings negatively impacts intermediate school outcomes such as primary and secondary school test scores, grade repetition, and private school choice (Conley & Glauber, 2006; Cáceres-Delpiano, 2006; Goux & Maurin, 2005; Åslund & Grönqvist, 2010), there is little evidence of any sizable sibship size impact on long-run outcomes such as educational attainment and earnings (Black et al., 2005; Åslund & Grönqvist, 2010; Angrist et al., 2010; De Haan, 2010).[4]

Our contribution is twofold. First and foremost, we provide evidence on a previously understudied and possibly more relevant margin of sibship size: having at least one sibling versus being an only child. In contrast, existing studies focus on the effects of having additional siblings beyond the first. This distinction matters for several reasons. One is that the relationship between sibship size and parental investments need not be monotonic (Mogstad & Wiswall, 2016). Since the largest drop in per-child resources occurs when moving from no siblings to any sibling, resource dilution effects may be most pronounced when parents must divide their attention, time, and financial resources for the first time, rather than when an already shared

[4]Our study also connects to those sibling studies that estimate birth order effects (Black et al., 2011; Conley & Glauber, 2006; Gary-Bobo et al., 2006; Kantarevic & Mechoulan, 2006; Kristensen & Bjerkedal, 2007; Booth & Kee, 2009; De Haan, 2010; Houmark, 2023) and sibling spacing effects (Pettersson-Lidbom & Skogman Thoursie, 2009; Buckles & Munnich, 2012). These studies rely on within-family regression designs that exclude only children from their samples and therefore offer no insights into the effects of having no siblings.

pool is divided further. The other reason is that the presence of a sibling may fundamentally alter a child's social environment, introducing peer-like interactions and social learning opportunities that are qualitatively different from the effects of having additional siblings. Moreover, the first-sibling margin is especially relevant in modern low-fertility contexts, where families with more than two children are becoming increasingly rare.

Second, our identification strategy relies on different identifying assumptions. Our core assumption is that IVF success is as good as random, conditional on observables. We provide empirical support for this assumption by showing that success at first and later IVF attempts is uncorrelated with a rich set of parental characteristics and first-born birth outcomes. The twin design, meanwhile, rests on the related assumption that twinning is essentially random. With the rise of IVF treatments, multiple births are becoming less random and more of a choice, turning twin births into a less suitable instrument in contemporary settings (Bhalotra & Clarke, 2019; Monden et al., 2021).

And lastly, our study relates to an expanding literature that uses IVF-based research designs. One strand relies on quasi-experimental variation in parenthood resulting from IVF success to estimate effects on labor market outcomes, mental health, and divorce (Lundborg et al., 2017; Räsänen, 2023; Bensnes et al., 2023; Gallen et al., 2023; Lundborg et al., 2024a; Ilciukas, 2024; Bögl et al., 2024).[5] Another strand uses data on donor-assisted IVF treatments to disentangle nature and nurture influences in the intergenerational transmission of human capital (Lundborg et al., 2024b). We build on these studies by extending our IVF design to a new and understudied question: the effects of growing up as an only child.

The rest of the paper unfolds as follows. Section 2 describes the data, the construction of the estimation sample, and institutions related to IVF treatments. Section 3 introduces the methodology. Section 4 presents results. Section 5 concludes.

[5]Most studies find that IVF success in childless families has, at most, modest effects on labor market outcomes, mental health, and relationship stability. These effects are likely less relevant in families that already have a child and undergo IVF treatment for a second child. If any effects exist, however, they may act as potential mechanisms through which siblings affect the school outcomes of first-born children.

2 IVF Institutions and Data

In our analysis, we combine information from several Danish administrative registers, nationwide school well-being surveys, and school test data. Below, we briefly discuss the relevant IVF institutions, present the main variables, describe our primary sample of children in one-child families who enter IVF treatment for a second child, and assess whether IVF success is as good as random.

2.1 Relevant IVF Institutions

Couples in Denmark are medically classified as infertile after one year of trying to conceive without success. With a referral from a general practitioner, infertile couples qualify for various medical fertility treatments, which vary in intensity, cost, and effectiveness. This study focuses on IVF, an intensive and costly treatment typically considered by couples only after other interventions have failed. IVF involves ovulatory stimulation, egg collection, fertilization of eggs outside the woman's body, and implantation of mostly one fertilized embryo. In Denmark, IVF costs around 28,000 DKK (3,750 euros) per treatment or 52,000 DKK (7,000 euros) for a package of three treatments. The Danish healthcare system covers the first three treatments for childless infertile couples. The couples in our sample already have one child and therefore need to pay for treatment themselves. Exceptions are newly formed couples with one pre-existing child (who do not have to pay) and one-child couples with spare embryos from earlier treatments (who pay considerably less).

IVF is a relatively effective treatment. In our sample, 30 percent of families successfully have a second child directly after their first embryo insertion. Among those for whom the first insertion fails, 65 percent eventually have a second child. It is important to note that not all of these births occur through IVF. Among the couples who end treatment, 20 percent later conceive a second child naturally (Thwaites et al., 2023).

2.2 Main Variables

For the cognitive and noncognitive school outcomes, we rely on nationwide school test data and well-being surveys. From 2010 onward, children in primary and lower secondary education take multiple tests in reading and math. The nationwide school test data contain the test scores for four reading tests (taken in grades 2, 4, 6, and 8) and two math tests (taken in grades 3 and 6). We observe test scores between 2010 and 2021. Most first-born children in our sampling window take a total of four to five tests. We use these test scores to measure the child's cognitive school outcomes. For each test, we calculate the standardized test score based on the mean and standard deviation of the cohort grade-specific test score in the representative sample. We define school outcomes in math and reading test scores as the average of all observed standardized test scores in reading and math.

From 2014 onward, children in public primary and lower secondary education also participate in a yearly nationwide well-being survey. The survey contains many questions, including those that measure three personality traits: agreeableness, conscientiousness, and emotional stability.[6] Children take the survey in grades 4 to 9. We have access to surveys between 2014 and 2021. Each year, the same two to three questions are used to assess each trait, such as "Are you good at collaborating with others?", "If you are interrupted during lessons, can you quickly concentrate again?" and "Do you feel accepted at school?" Children answer these questions on a five-point scale (almost always, often, now and then, rarely, and never). We use these answers to measure the child's personality traits at school.[7] We standardize each trait based on the mean and standard deviation of the specific trait of the cohort grade in the representative sample. We define the student's personality trait as the average of all their grade-specific standardized personality traits. Lastly, the well-being survey also asks students "Are you happy with your class?" and "Are you happy with your school?" We use the two five-point scale answers to measure their overall happiness. Like the personality traits, we take the average of all the cohort-grade-specific standardized happiness answers.

For the endogenous and control variables, we turn to the population register of all

[6]The commonly used five-factor model of personality structure assigns someone's personality to five core domains: agreeableness, conscientiousness, emotional stability, extraversion, and openness to experience. The well-being survey covers three of the five core domains. In the school context of the survey, agreeableness captures the child's tendency to be cooperative and empathetic toward fellow classmates, conscientiousness captures the child's responsibility and ability to work carefully to get things done, and emotional stability captures the child's anxiety and vulnerability to stress (or the absence thereof).

[7]We note that these trait measures were recently validated in a register-matched personality survey by Andersen et al. (2020). The study shows that the measures constructed using the responses to the school survey are strongly correlated with the Big Five personality traits, specifically conscientiousness, agreeableness, and emotional stability, measured using well-established methods.

individuals residing in Denmark. We use this register to select all first-born children, with links to their parents and siblings (if they have any). For the key endogenous variable in our study, we use these links (or the absence thereof) to define whether first-born children have siblings. We use the identifiers of first-born children to collect information from the birth register on their birth weight (in kilograms) and birth length (in centimeters). Using the identifiers of parents, we extract information from the education and tax registers on their educational attainment, labor market status, and annual income. We measure education using dummies for whether each parent completed college. We measure employment status (having positive earnings) and income (in millions of DKK) using values from the year before the birth of the first child.

For the instrumental variable, we use the IVF register provided by the Danish Health Data Authority (Sundhedsdatastyrelsen). The IVF register records all IVF treatments that have taken place in public and private fertility clinics and hospitals since 1994, including information on the main reason for infertility and, for each IVF attempt, information on the type of treatment, the four different stages of treatment (medication, egg and sperm extraction, embryo fertilization, and embryo insertion), treatment success, the date of treatment and, where applicable, the date of birth. To study the causal effect of siblings, we leverage quasi-experimental variation in IVF success after embryo insertion. For each insertion, we construct a success dummy indicating whether the woman had a live birth within 10 months, without further insertions during that period. This success indicator for the first IVF attempt is the instrument in our instrumental variable design.

2.3 Main Estimation Sample

We also use the IVF register to select our estimation sample of all first-born children in families who enter IVF treatment for a second child. To be included in this sample, we select all first-born children (both IVF and non-IVF) in their primary or lower secondary school years, raised in families somewhere treated between 1995 and 2019, with at least one reading test score and no missing control variables.[8] We further

[8]Sample sizes vary across outcomes. We primarily focus on the numbers and statistics for the reading test sample, which is the outcome observed for the largest number of first-born children. We observe reading test scores more than any other outcomes because there were more nationwide tests in reading than in math, and because the nationwide tests were introduced earlier than the well-being surveys.

exclude a small number of first-born children with low birth weight (less than 2,500 grams). Our primary estimation sample contains 10,906 first-born children.

We also construct a secondary estimation sample of first-born children in the general population, applying the same selection criteria used for the IVF sample: children must be in their primary or lower secondary school years, have at least one reading test score, and have no missing control variables. This representative sample includes all first-born children, regardless of whether they have siblings, and allows us to assess the generalizability of our findings. It contains 339,281 first-born children.

Table 1 reports sample means for child and parent characteristics, family size, and school outcomes in both the sample of families who underwent IVF for a second child and the representative sample. We see that child gender and birth weight, which are both strong predictors of cognitive test scores and personality traits (Black et al., 2007; Mueller & Plug, 2006), are nearly identical among first-born children in different families. The likelihood that the first-born child is an IVF child is very different across the different families. This is not surprising. The high IVF rates in families seeking treatment for a second child indicate that fertility problems encountered during the first pregnancy are likely to recur. The low IVF rates in representative families, on the other hand, indicate that most families are unlikely to experience fertility problems requiring IVF treatment. Among all first-born children in the representative sample within our sampling window, only 2 percent are conceived through IVF.

For the parent characteristics across the different families, we see that parents who undergo IVF for a second child are more educated, more likely to be employed, have higher income, and are older, than those who do not. These differences are similar to those found in earlier studies (Bitler, 2012; Lundborg et al., 2017, 2024a; Bensnes et al., 2023). To assess the external validity of our findings, we conduct heterogeneity analyses focusing on first-borns raised in families that more closely resemble the broader population—specifically, those that are less affluent, less educated, and younger.

When examining sibling patterns, we find that IVF-treated families are less likely to have more than one child, and that, when siblings are present, the spacing between the first and second child tends to be longer. Both patterns are consistent with IVF families experiencing fertility problems. Conditional on having a sibling, differences in family size are relatively small. In families with two or more children, first-borns in the IVF sample are one percentage point less likely to have more than one sibling

Table 1: Descriptive statistics for first-born children in different families

	1st-born child in families seeking IVF for 2nd child	1st-born child in representative families	p-values for differences
Characteristics at first child birth			
Female (0/1)	0.48 (0.50)	0.49 (0.50)	0.435
IVF child (0/1)	0.48 (0.50)	0.02 (0.14)	0.000
Child birth weight (kg)	3.51 (0.48)	3.51 (0.46)	0.987
Child birth length (cm)	52.05 (2.24)	52.00 (2.23)	0.023
Mother age	30.42 (4.48)	27.67 (4.56)	0.000
Mother college (0/1)	0.53 (0.50)	0.45 (0.50)	0.000
Mother income (M)	0.21 (0.13)	0.16 (0.12)	0.000
Mother work (0/1)	0.93 (0.26)	0.89 (0.32)	0.000
Father age	32.85 (5.43)	30.36 (5.54)	0.000
Father college (0/1)	0.37 (0.48)	0.30 (0.46)	0.000
Father income (M)	0.28 (0.20)	0.22 (0.16)	0.000
Father work (0/1)	0.91 (0.29)	0.89 (0.32)	0.000
Observations	10,906	339,281	
Characteristics after first child birth			
Any siblings (0/1)	0.75 (0.43)	0.84 (0.37)	0.000
Total siblings (if any)	2.37 (0.56)	2.43 (0.67)	0.000
2+ siblings (0/1, if any)	0.33 (0.47)	0.34 (0.48)	0.005
Birth spacing (if any)	4.93 (2.89)	3.42 (2.01)	0.000
Math score	0.22 (0.91)	0.09 (0.94)	0.000
Reading Score	0.24 (0.88)	0.10 (0.92)	0.000
Agreeableness	0.09 (0.80)	0.02 (0.82)	0.000
Conscientiousness	0.14 (0.83)	0.04 (0.85)	0.000
Emotional stability	0.05 (0.84)	0.00 (0.85)	0.000
School Happiness	0.08 (0.80)	0.02 (0.83)	0.000

Note–The first two columns show sample means for two samples with standard deviations in parentheses: first-born children raised in families who seek IVF treatment for a second child, and first-born children raised in representative families. The third column shows p-values for tests of whether the means in the first two columns are significantly different from each other. We further note that the sample means are reported for the largest sample (which is in our setting the sample with reading test scores as main outcome).

and have 0.06 fewer siblings, on average.

When it comes to school outcomes, first-born children in families who pursue IVF treatment for a second child tend to fare better: they score higher on math and reading tests, exhibit greater agreeableness, conscientiousness, and emotional stability, and not unimportantly report being more satisfied with their overall classroom and school experiences.

2.4 Is IVF Success Conditionally Random?

Our empirical strategy relies on pregnancy after embryo insertion being as good as random. The medical literature indicates that this condition may not hold unconditionally and points to year and age of treatment as the most important factors determining success in IVF treatment.[9] Following this literature, we take year and age at treatment as the primary drivers behind IVF success and test whether, conditional on these factors, treatment success approximates a random process.

Table 2 contains the balancing test results for each embryo insertion, up to the sixth attempt. In particular, it shows the means of baseline characteristics of first-born children (including gender, IVF status, birth weight and length) and their parents (including both parents' education, employment status, and income in the year before the first IVF attempt) in couples with a successful attempt (in columns 1, 4, and 7), the unconditional mean differences between couples with a successful and failed attempt (in columns 2, 5, and 8), and the conditional mean differences after accounting for parental age and treatment year. For consistency with our main estimation approach, described in Section 3, we present the conditional comparison based on inverse probability weights. These weights are constructed using a logistic regression of attempt success on the mother's age and age squared at treatment, the father's age and age squared at treatment, and a full set of treatment year indicators. We present these test results using the largest reading test sample. The results do not meaningfully change when we use a regression framework instead. At the bottom of the difference-in-means columns, we also present p-values for F-tests for whether the child and parent characteristics are jointly statistically significant predictors of treatment success. If IVF success is random within the strata based on year and age

[9] Success rates are lower for older couples due to age-related declines in fertility (Rosenwaks et al., 1995; Templeton et al., 1996; van Loendersloot et al., 2014). Success rates are higher for more recent treatments due to medical innovations (Wang & Sauer, 2006; Niederberger & others., 2018).

at treatment, we expect that the differences in mean characteristics are close to zero and statistically insignificant.

For the first attempt, the unconditional means and differences (in columns 1 and 2) indicate that younger parents and parents who conceived their first child with IVF experience significantly higher success rates. We see little difference on the other characteristics, including well-known predictors of child school performance (including child birth weight and parent education). Nonetheless, all the child and parent characteristics together are statistically significantly related to treatment success, which suggests that success at the first attempt is not random. When we account for parental age and treatment year (in column 3), however, we no longer see a clear relationship between treatment success and the characteristics of children and parents. The differences in means are all near zero. Moreover, when we consider all the child and parent characteristics together, there is no statistically significant difference between successfully and unsuccessfully treated families. These results are consistent with IVF success being as good as random, conditional on parental age and treatment year.[10]

For later attempts, after accounting for parental age and treatment year, there are no substantial differences in the pretreatment characteristics of children or parents between those who succeed and those who do not (columns 3, 6, and 9). While higher-income couples are more likely to continue treatment following repeated failures, treatment success at each stage does not appear selective. At each attempt, joint differences in child and parent characteristics are not statistically significant. We recognize, though, that the smaller samples in later treatments make it less likely to uncover statistically significant differences. Nonetheless, all the balancing results after accounting for parental age and year are as one would expect with conditional random treatment success at the first and later attempts.

[10]Education has also been linked to treatment success (Groes et al., 2024). Table 2 indicates that balancing on year and age at treatment already eliminates any meaningful differences in key pretreatment characteristics including the college indicators we use to measure the parents' education.

Table 2: Relationship between first-born child characteristics and IVF success

	First attempt			Second attempt			Third attempt		
	Succ.	Fail. Diff.	Cond. Diff.	Succ.	Fail. Diff.	Cond. Diff.	Succ.	Fail. Diff.	Cond. Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female child	0.49	0.01	0.01	0.47	-0.01	-0.01	0.48	-0.00	0.01
IVF child	0.50	0.03	0.01	0.52	0.05	0.04	0.50	0.02	0.00
Child weight	3.52	0.01	0.01	3.51	0.01	0.02	3.50	0.01	0.01
Child length	52.05	-0.00	0.03	52.08	0.05	0.10	52.12	0.13	0.07
Mother age	33.41	-1.12	-0.03	33.43	-1.23	-0.03	33.63	-1.20	0.00
Mother college	0.52	-0.00	0.02	0.55	0.00	0.01	0.57	-0.00	0.01
Mother income	0.21	-0.00	0.01	0.22	-0.01	0.00	0.23	-0.00	0.01
Mother work	0.93	0.01	0.01	0.93	0.00	-0.00	0.95	0.02	0.02
Father age	35.87	-1.07	0.00	35.89	-1.14	0.01	35.97	-1.26	0.23
Father college	0.37	-0.01	0.01	0.38	-0.01	0.01	0.43	0.02	0.04
Father income	0.27	-0.00	0.01	0.29	0.01	0.01	0.29	0.00	0.00
Father work	0.91	0.00	-0.00	0.92	0.01	0.00	0.93	0.02	-0.00
Joint p-value		0.00	0.38		0.00	0.17		0.00	0.68
Observations		10,906			6,150			3,498	
Success probability		0.30			0.28			0.25	

	Fourth attempt			Fifth attempt			Sixth attempt		
	Succ.	Fail. Diff.	Cond. Diff.	Succ.	Fail. Diff.	Cond. Diff.	Succ.	Fail. Diff.	Cond. Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female child	0.50	0.02	0.02	0.47	-0.00	0.05	0.50	0.05	0.02
IVF child	0.55	0.06	0.05	0.52	0.01	0.01	0.54	0.03	0.04
Child weight	3.50	0.02	0.03	3.48	0.00	-0.02	3.49	0.03	0.05
Child length	51.90	-0.10	-0.04	51.88	-0.07	-0.06	52.01	0.16	0.18
Mother age	34.03	-0.71	0.08	33.33	-1.45	0.04	33.56	-1.25	-0.13
Mother college	0.64	0.04	0.05	0.65	0.04	0.07	0.60	-0.02	-0.04
Mother income	0.24	0.00	0.00	0.24	0.00	0.01	0.24	-0.00	-0.00
Mother work	0.94	0.00	0.00	0.95	0.02	0.02	0.94	0.01	-0.01
Father age	36.54	-0.55	0.12	35.68	-1.45	-0.25	35.62	-1.41	-0.21
Father college	0.43	-0.00	0.00	0.46	0.00	0.05	0.56	0.11	0.10
Father income	0.29	-0.01	-0.00	0.31	-0.01	-0.01	0.34	0.02	0.02
Father work	0.91	0.00	-0.01	0.92	0.01	0.01	0.96	0.05	0.03
Joint p-value		0.01	0.34		0.00	0.70		0.00	0.79
Observations		1,872			1,121			674	
Success probability		0.22			0.22			0.21	

Note–Columns (1), (4), and (7) present mean characteristics for children whose parents experienced a successful treatment. Columns (2), (5), and (8) present differences in mean characteristics between families with successful and failed treatments. Columns (3), (6), and (9) present differences after accounting for father's age and age squared, mother's age and age squared, and a full set of year-of-treatment indicators measured at each consecutive treatment, using inverse probability weights based on a logistic regression model. Observations refer to the total number of first-born children whose parents underwent the respective procedure for a second child.

3 Identification and Estimation

As the objective of this study is to estimate the effect of having siblings on the school outcomes of first-born children, we start with the standard IV-IVF framework as introduced by Lundborg et al. (2017). We consider all first-born children in one-child families that underwent IVF treatment for a second child and use the following variables in our analysis: Y is the school outcome of the first-born child, S is a sibling indicator that measures whether the first-born child has any siblings or not, and Z_1 is the treatment success indicator that measures whether the first IVF attempt involved embryo implants and led to siblings or not. We define the parents' first attempt Z_1 as our main instrument and show later that it is a strong (and thus relevant) sibling predictor.[11]

3.1 Identification for Complier Children

Using the LATE framework of Angrist & Imbens (1995), we classify children in our sample as either compliers or always takers. The compliers ($C=1$) are children who never end up having siblings after a first failed attempt. The always takers ($C=0$) are children who always end up having siblings, regardless of a first failed attempt. In our setup, there are no never takers and defiers because the first-born children whose parents had a successful first attempt end up having siblings by construction. We define the child's potential school outcomes $Y_z(s)$ indicating what the child's school outcome would be in case the child had parents whose first IVF attempt succeeded or failed ($z = 0, 1$) and the child had any siblings or not ($s = 0, 1$). For each child, there are three potential school outcomes $Y_0(0), Y_0(1)$ and $Y_1(1)$. Only one of these outcomes is realized.

In the standard LATE framework (abstracting from covariates for expositional clarity), we identify the average effect of having siblings for compliers by assuming that the potential outcomes satisfy the following independence and exclusion conditions:

A1 Independence $Y_1(1), Y_0(0), Y_0(1), C \perp Z_1$.

A2 Exclusion (for non-compliers) $E[Y_1(1)|C=0] = E[Y_0(1)|C=0]$.

[11] We use success at first attempt as instrument for two reasons. First, it is more exogenous than success after multiple attempts. Second, it is defined for all one-child families who underwent IVF for a second child.

Independence (**A1**) assumes that success at the first IVF attempt is as good as randomly assigned, that is, independent of child type and potential school outcomes (conditional on the parents' age and age squared at treatment and year of treatment). Table 2 (column 3), which shows that there are no meaningful differences between the pre-treatment characteristics of children and parents (after accounting for parental age and treatment year effects), supports this assumption. Exclusion (**A2**) assumes that first IVF success affects first-born outcomes only through the presence of a sibling, and not through the timing of the sibling's arrival. This assumption would be violated if some couples who fail their first IVF attempt later succeed through additional treatment, in which case the instrument (success at the first attempt) would influence not only the likelihood of having a sibling, but also the timing of the sibling's birth. We refer to this as delayed fertility. Of the two conditions, we consider (**A2**) the more substantive one.

We get the first stage and reduced form by comparing the sibling shares and school outcomes of first born children whose parents experience a first-time IVF success and failure. Under independence, the first stage identifies the complier share

$$E[S|Z_1=1] - E[S|Z_1=0] = \Pr[C=1]. \tag{1}$$

Under independence, the reduced form can be written as

$$\begin{aligned} E[Y|Z_1=1] - E[Y|Z_1=0] \\ &= E[Y_1(1) - Y_0(0)|C=1] \Pr[C=1] \\ &\qquad + E[Y_1(1) - Y_0(1)|C=0] \Pr[C=0]. \end{aligned} \tag{2}$$

The reduced form consists of two terms. The first captures the causal effect of whether siblings arrive for compliers, which is the effect we are after. The second term captures the causal effect of when siblings arrive for always takers, which is in the context of our study the bias term. Under exclusion, this bias term is assumed zero, which means that the reduced form in (2) and first stage in (1) together identify the effect of having siblings for compliers

$$\frac{E[Y|Z_1=1] - E[Y|Z_1=0]}{E[S|Z_1=1] - E[S|Z_1=0]} = E[Y_1(1) - Y_0(0)|C=1]. \tag{3}$$

In the standard LATE framework, we can estimate the effect of having siblings for compliers using either a two-stage least squares regression approach (2SLS) or an inverse probability weights approach (IPW). While we show later that both approaches yield very similar results, we will primarily focus on the IPW method going forward, as it more closely aligns with the approach we propose to enhance statistical precision and mitigate exclusion violations.

3.1.1 2SLS Estimation for Complier Children

In the standard two-stage least squares regression approach (2SLS), the first-stage and second stage regression equations can be written as:

$$S = \alpha X_1 + \beta Z_1 + u, \tag{4}$$

and

$$Y = \gamma X_1 + \delta \widehat{S} + \upsilon. \tag{5}$$

In these two regression equations, all the unobservable factors that are somehow related to having siblings and school outcomes are captured by the econometric errors u and υ. The control variables X_1 contain the parents' ages at first treatment, their squares, and a full set of year at first treatment dummies and ensure that treatment success at first treatment and unobservable factors are assumed unrelated (conditionally on X_1). The other variables Y, S and Z_1 have been defined earlier. The first-stage coefficient β (attached to the instrument) represents the complier share. The second-stage coefficient δ (attached to the first-stage predicted sibling indicator) represents the causal effect of having any siblings on the school outcomes of first-born compliers.

3.1.2 IPW Estimation for Complier Children

Like the 2SLS approach, the IPW approach estimates the impact of having siblings in several steps. We first estimate propensity scores taken from logistic regressions of success at the first attempt on the parents' ages at first attempt, their squares, and a full set of year at first attempt dummies. We define $\widehat{p_1}(X_1)$ as the estimated propensity score and X_1 as the propensity score controls. We second use these predicted propensity scores to balance the first-born children with successfully and unsuccessfully first-time treated parents. Under conditional independence, the first-stage effect is:

$$E\left[\frac{Z_1 S}{\widehat{p_1}(X_1)} - \frac{(1-Z_1)S}{1-\widehat{p_1}(X_1)}\right] \tag{6}$$

Under conditional independence, the reduced form effect is:

$$E\left[\frac{Z_1 Y}{\widehat{p_1}(X_1)} - \frac{(1-Z_1)Y}{1-\widehat{p_1}(X_1)}\right] \tag{7}$$

Under conditional independence and exclusion, the average effect of having siblings for compliers is the reduced form rescaled by the first stage.[12]

3.2 Identification for Relier Children

We next present an alternative approach to quantify the effect of having siblings by exploiting quasi-experimental variation not only at the first IVF attempt but also subsequent IVF attempts. This approach has three advantages over earlier approaches that identify effects in first-born complier samples. First, we identify the effect of having siblings outside the sample of first-born compliers. Second, we get more precise effect estimates. And third, we alleviate concerns about violations of the exclusion condition.

In our earlier analysis, we compared first-born children whose parents' first IVF attempt succeeded (the treatment group) to those whose parents' first attempt failed (the control group). Our new approach introduces an alternative control group and considers first-born children whose parents never conceived a second child via IVF. Unlike the baseline control group, this alternative control group includes first borns whose parents had a second child through means other than IVF, but excludes those whose parents had a second child at later IVF attempts, resulting in a smaller share of always takers.

[12]The first- and second-stage regression models with multivariate control variable X_1, as specified in equations (4) and (5), can be reformulated to use the estimated propensity score $\widehat{p_1}(X_1)$ as a univariate control variable. In this context, the IPW and 2SLS approaches may yield different results: the IPW approach applies uniform weights across observations, whereas 2SLS implicitly assigns complier weights that depend on $\widehat{p_1}(X_1)$ (Angrist and Pischke, 2009, p. 177). In our IVF setup, however, we find that the resulting estimates from both approaches are nearly identical because there is little variation in $\widehat{p_1}(X_1)$.

We acknowledge that the alternative control group is a selective group of first-borns (because their parents' decision to continue treatment after a failed attempt is selective). But if we assume that success at later IVF attempts is also as good as randomly assigned, which appears to be a reasonable assumption, we can use the outcomes of first-born children from unsuccessfully treated couples to infer what the school outcomes of first-born children with later-attempt siblings would have been in a scenario where all IVF attempts had failed. In particular, the randomness at later IVF attempts allows us to focus more on first borns whose parents stopped treatment after each failed attempt and ignore those who had siblings at later attempts. By rebalancing the original treatment group and the new control group, we can identify the effect of having siblings for a broader group of first-born children consisting of compliers and always takers who rely on IVF treatment for siblings.

More formally, we follow Ilciukas (2024) and distinguish first-born children in two dimensions. First, we classify first borns as either reliers or non-reliers (R). The reliers ($R = 1$) are children whose parents rely on IVF treatment for a second child; that is, these are the first-born children who end up with siblings if their parents have a successful first or later treatment but who would not have siblings otherwise. The reliers consist of compliers and always-takers who would get siblings from later IVF attempts but would not get siblings if all attempts failed. In our setup, we exclude those reliers who actually get siblings at later attempts. The non-reliers ($R = 0$) are the first-born children who would have siblings through other means, regardless of IVF success. Most non-relier siblings are conceived naturally. Few non-relier siblings are adopted.[13] Second, we also classify first borns by the number of IVF attempts their parents would undergo for the second child in a scenario where all IVF attempts fail (W). We next define Z_j as the treatment success indicator at IVF attempt j for the second child and Q as the realized number of IVF attempts parents actually undergo for the second child. By definition, we consider parents in IVF treatment until an IVF attempt succeeds ($Z_w = 1, Q = w$) or until they reach the maximum number of IVF attempts they are willing to undergo and quit ($Z_Q = 0, Q = W$).

[13] When we exclude first-born children with siblings from later IVF attempts, the three potential outcomes $Y_0(0), Y_0(1)$ and $Y_1(1)$ are the same as before; that is, $Y_0(0)$ refers to outcomes without siblings and parents whose first to last attempt failed, $Y_0(1)$ refers to outcomes with siblings and parents whose first to last attempt failed, and $Y_1(1)$ refers to outcomes with siblings and parents whose first attempt succeeded.

3.2.1 The Sequential Independence Condition

We next discuss the conditions under which we can identify the average effect of having siblings for reliers. Our key novel condition is the sequential independence assumption:

A3 Sequential independence: $Y_1(1), Y_0(0), Y_0(1), W, R \perp Z_j \mid Q \geq j$.

In words, this condition (**A3**) asserts that, among parents who enter IVF attempt j, success at attempt j is as good as randomly assigned, that is, independent of potential school outcomes and child type (conditional on the parents' age at treatment j and the year of treatment j). If we accept that treatment success at the first attempt is (conditionally) random, the assumption of (conditionally) random treatment success at subsequent attempts is not too difficult to accept. Table 2 further demonstrates balance on relevant child and parent pre-treatment characteristics at each subsequent treatment, which supports the sequential independence condition.

This sequential independence condition gives us the relier share and a modified reduced form for reliers. Consider first-born children whose parents underwent exactly w failed IVF attempts. We know four essential things about these children. First, they are type-w children; otherwise their parents would have either continued treatment or stopped earlier. Second, they are a random subsample of type-w children; under sequential independence, success at attempt w is independent of potential school outcomes and child relier type. Third, we identify the shares of different type-w children from the shares of children whose parents either continue or stop treatment after a random sequence of previous failed attempts.[14] Fourth, we observe their school outcomes, which we can express as the potential outcome average of type-w reliers and non-reliers. For reliers, the potential outcomes are the outcomes for those who never get siblings. For non-reliers, the potential outcomes are the outcomes for those who get siblings though other means.

Putting this together, we can write the average school outcomes for first-born

[14] Let us start with type-1 children. If treatment success at the first attempt is random, the share of type-1 children can be identified from the share of children whose parents decide to end treatment after the first failed attempt ($\Pr[W = 1] = \Pr[Q = 1|Z_1 = 0]$). Next, consider type-2 children. If treatment success is random at the first two attempts, the share of type-2 children can be identified from the share of children whose parents continue treatment after the first failed attempt but stop after the second failed attempt ($\Pr[W=2]=\Pr[W>1]\Pr[Q=2|Q>1, Z_2=0]$). And so on.

children whose parents completed exactly w failed IVF attempts as follows:

$$\begin{aligned} E[Y|Q=w, Z_w=0] = {} & E[Y_0(0)|W=w, R=1] \Pr[R=1|W=w] \\ & +E[Y_0(1)|W=w, R=0] \Pr[R=0|W=w]. \end{aligned} \tag{8}$$

With known shares and outcomes for all type-w children, we can recover the overall average outcome in a scenario where all IVF attempts fail:

$$\begin{aligned} & \textstyle\sum_{j=1}^{\overline{Q}} E[Y|Q=j, Z_j=0] \Pr[W=j] = \\ & \qquad E[Y_0(0)|R=1] \Pr[R=1] + E[Y_0(1)|R=0] \Pr[R=0], \end{aligned}$$

where $\overline{Q}$ represents the highest number of IVF attempts observed in our sample. This is the outcome average for the alternative control group, consisting of first-born children whose parents never conceived a second child via IVF, expressed as potential outcome averages for reliers and non-reliers. When we subtract this average outcome from the average outcome of first-born children in the original treatment group whose parents experience a first-time IVF success ($E[Y|Z_1 = 1]$), we get the new reduced form for reliers:

$$\begin{aligned} & E[Y|Z_1=1] - \textstyle\sum_{j=1}^{\overline{Q}} E[Y|Q=j, Z_j=0] \Pr[W=j] = \\ & \qquad E[Y_1(1) - Y_0(0)|R=1] \Pr[R=1] + E[Y_1(1) - Y_0(1)|R=0] \Pr[R=0]. \end{aligned} \tag{9}$$

The first term captures the effect of whether siblings arrive for reliers. The second term is the bias term and captures the effect of when siblings arrive for non-reliers. Importantly, the new reduced form does not rely on always takers with siblings from later IVF attempts. As a result, the bias term is given much less weight simply because the relier share is much larger than the complier share. In our sample, for example, 53-55 percent of first borns are classified as reliers, compared to only 32-34 percent who are compliers. This shift therefore takes us a long way in addressing exclusion violations that arise from siblings born at later IVF attempts.

3.2.2 Alternative Exclusion Conditions

The sequential independence condition (**A3**), however, does not tackle the exclusion violations caused by the later-born non-IVF siblings. After all, this condition does not assume that the arrival of non-IVF siblings is random. To deal with these later-born non-IVF siblings and the possible exclusion violations they create, we discuss a number of alternative exclusion conditions under which we can identify or bound the impact of having siblings for first-born children.

We first consider the conventional exclusion condition that enables us to identify the effect of having siblings for reliers:

A4 Exclusion (non-reliers): $E[Y_0(1)|R=0] = E[Y_1(1)|R=0]$.

Exclusion for non-reliers (**A4**) is similar to the original exclusion condition (**A2**) in that it asserts that the impact of having siblings early is the same as having siblings late. While this assumption remains substantive, the exclusion condition under (**A4**) is arguably weaker than the one under (**A2**), as it involves relatively fewer first-born children with siblings (because the relier share outweighs the complier share). Under exclusion for non-reliers, we can set the bias term in the modified reduced-form effect for reliers in (9) to zero and identify the effect of having siblings for reliers:

$$E[Y_1(1) - Y_0(0)|R=1].$$

Rather than imposing the standard exclusion condition for reliers, we can replace it with an alternative condition on the (ignorable) impact of non-IVF siblings. This allows us to identify the effect of having siblings for all the first-born children in our sample:

A5 Late sibling exclusion (non-reliers): $E[Y_0(1)|R=0] = E[Y_0(0)|R=0]$.

The late sibling exclusion condition (**A5**) states that the presence of siblings is ignorable for children who gain siblings through means other than IVF. This condition can be justified if sibling effects on first borns are concentrated in their early developmental stages or if siblings arrive after the first-born outcomes are measured. In our sample, first-attempt IVF failure increases the average sibling age gap among non-reliers from 4.1 to 6.3 years, suggesting that many non-reliers have their siblings when they are already in school. Under (**A5**), we effectively treat $E[Y_0(1)|R=0]$ in (9) as if it were $E[Y_0(0)|R=0]$, and identify the effect of having siblings for the full sample of first-born children:

$$E[Y_1(1) - Y_0(0)].$$

We note that conditions (**A4**) and (**A5**) are fundamentally different. Condition (**A4**) assumes that later-born non-IVF siblings have the same impact on the school outcomes of first-born children as siblings conceived at the first-IVF attempt. In contrast, condition (**A5**) assumes that later-born non-IVF siblings have no measurable impact on the school outcomes of first-born children. Both conditions are met if siblings have no impact at all on the school outcomes of first-born children. We consider this an unattractive feature, given that our objective is to estimate the effect of having siblings.

Instead, we can derive informative bounds under alternative assumptions by introducing the following homogeneity and monotonicity conditions:

A6 Homogeneity: $E[Y_1(1) - Y_0(0)|R=1] = E[Y_1(1) - Y_0(0)|R=0]$.

A7 Monotonicity (non-reliers): $0 < \frac{E[Y_0(1)-Y_0(0)|R=0]}{E[Y_1(1)-Y_0(0)|R=0]} < 1$.

The homogeneity condition (**A6**) says that the effects for having siblings are comparable for reliers and non-reliers. This assumption of homogeneous sibling effects is restrictive, but may not be so unrealistic if we believe that most parents in treatment do not expect to have a second child through natural conception. The monotonicity condition (**A7**) allows the effects for having siblings to vary with birth spacing. In particular, it states that the impact of having siblings gets monotonically weaker the later siblings arrive (and the longer first-born children spend their time as only child in the family). In our setup, monotonicity implies that for non-reliers the impact of having siblings early $E[Y_1(1)-Y_0(0)|R=0]$ is stronger than the impact of having them late $E[Y_0(1) - Y_0(0)|R=0]$. This is consistent with the idea that early-life conditions matter most: when siblings arrive early on, they are more likely to shape the later life outcomes of first born children, including their school outcomes. Under conditions (**A6**) and (**A7**), it is rather straightforward to show that the average sibling effects obtained under (**A4**) and (**A5**) can be interpreted as upper and lower bounds on the average effect $E[Y_1(1) - Y_0(0)]$ in a world where the alternative exclusion conditions (**A4**) and (**A5**) do not hold (see the Appendix for details).

3.2.3 IPW Estimation for Relier Children

We can estimate the effect of having siblings using a modified inverse probability weights approach (IPW). For each IVF attempt, we estimate propensity scores taken from logistic regressions of attempt success on the parents' ages at that attempt, their squares, and a full set of attempt year dummies. Let $\widehat{p}_j(X_j)$ be the estimated propensity score at attempt j and X_j the propensity score controls.[15] These propensity scores are then used to balance the first-born children with successfully first-time treated parents (treatment group) and the first-born born children with never successfully treated parents (control group). Under sequential conditional independence, the modified first-stage effect is:

$$E\left[\frac{Z_1 S}{\widehat{p_1}(X)} - \frac{(1-Z_Q)S}{\prod_{j=1}^{Q}(1-\widehat{p_j}(X_j))}\right] \tag{10}$$

Under sequential independence, the modified reduced form effect is:

$$E\left[\frac{Z_1 Y}{\widehat{p_1}(X)} - \frac{(1-Z_Q)Y}{\prod_{j=1}^{Q}(1-\widehat{p_j}(X_j))}\right] \tag{11}$$

Under sequential independence and exclusion, the average effect of having siblings for reliers is the modified reduced form rescaled by the modified first stage. Under sequential independence and late sibling exclusion, the average effect of having siblings for all first born children in our sample is the modified reduced form. Standard errors are obtained using a multiplier bootstrap method.

Before presenting the results, we reiterate the key advantage of our proposed approach over the standard IV-IVF approach. We identify the effect of having siblings for first-born reliers, which consist of one-child families who rely on IVF to conceive a second child. This group is substantially larger than the group of first-born compliers, which consist of one-child families whose second child relies on a first successful IVF attempt.

The larger size of the relier group will, all else equal, translate into more precise estimates and mitigate concerns about potential violations of the exclusion restriction. Intuitively, we know that the noise in instrumental variable estimates gets smaller when reduced-form estimates are divided by larger first-stage estimates. If the variance of the modified and standard reduced forms is more or less comparable, we get a more precise second stage because we divide the reduced form by a larger number. And relatedly, the confounding role of possible exclusion violations gets smaller

[15] To avoid estimating propensities in small samples, we cap the number of attempts used for identification at five and treat any births following more than five attempts as potentially selective. In our sample, about 94 percent of all the parents undergo no more than five attempts. Later, in Table 6, we show that our results are not sensitive to this choice.

when first-stage estimates are larger. If we consider the effect of having siblings as a first-order effect (as opposed to the effect of having siblings later among those with siblings), the modified second stage comes along with a stronger first stage, which in turn assigns less weight to second-order exclusion violations.

4 Results

4.1 Only Child Associations

We start our analysis by just comparing the school outcomes of first-born children with and without siblings in the samples of representative families and families that undergo IVF treatment for a second child. Table 3 reports the unconditional differences in outcomes between only children and children with siblings across the two samples. The outcomes include standardized test scores in math and reading (columns 1 and 2), standardized measures of three personality traits agreeableness, conscientiousness, and emotional stability (columns 3, 4, and 5), and a standardized measure of overall school happiness (column 6). We say that children do better if they have higher test scores, are more agreeable, conscientious, and emotional stable, and report greater happiness in school. Results are presented separately for first-born children in the representative sample (Panel A) and in the IVF sample (Panel B).

The first pattern we observe is that, in both samples, children with siblings do substantially better than only children across all measured outcomes. These positive associations between having siblings and outcomes are consistent with a causal story if children benefit from socializing with younger siblings. The same positive associations, however, are also consistent with a selection story if one-child families are less stable, leading to poorer child outcomes and fewer siblings when this instability is not accounted for. Clearly, simple comparisons between children with and without siblings are not enough to establish a causal link.

The second pattern is that the associations are roughly twice as large in the representative sample as in the IVF sample. One possible explanation is heterogeneity in the effects of having siblings. First-born children in IVF-treated families may respond differently to the arrival of siblings than those first borns in representative families. Another explanation concerns heterogeneity in selection. Most parents in the representative sample have control over their fertility decisions and can choose to have

Table 3: Associations between having siblings and school outcomes

	Math test	Reading test	Agreeable	Consc.	Emotional stability	School happiness
Panel A: Representative sample						
Having siblings	0.171	0.090	0.074	0.122	0.114	0.115
	0.005	*0.004*	*0.005*	*0.005*	*0.005*	*0.005*
Observations	256,992	339,281	224,486	224,571	224,537	224,522
Panel B: IVF sample						
Having siblings	0.088	0.054	0.030	0.051	0.042	0.074
	0.023	*0.020*	*0.021*	*0.022*	*0.022*	*0.021*
Observations	8,308	10,906	7,689	7,666	7,567	7,783

Note–Unconditional outcome differences between first-born children with and without siblings. The outcomes are standardized test scores in math and reading (columns 1 and 2), standardized personality traits—agreeableness, conscientiousness, and emotional stability (columns 3, 4, and 5), and a standardized measure of overall school well-being (column 6). We present unconditional outcome differences for first-born children in representative families (panel A) and families seeking IVF treatment for a second child (panel B). Standard errors are shown in italics.

a second child or not. The parents in the IVF sample differ in two respects. First, all couples want a second child, which limits selection related to fertility preferences. Second, they have only limited control over their fertility, as indicated by the need for IVF, introducing some randomness in fertility. Although this does not eliminate selection (as the number of IVF attempts and natural conceptions may still be selective), it most likely reduces it. Taken at face value, the weaker associations in the IVF sample suggest that the sibling estimates from the representative sample may overstate the benefits of having siblings.

4.2 Only Child Effects in IVF Families

We next estimate the causal impact of having siblings on the school outcomes of first-born children using the IVF sample. As a first step, we need to show that variation in IVF success leads to substantial differences in whether parents go on to have a second child and whether their first-born child ends up with a sibling.

Table 4 presents the first-stage estimates of how parents' first IVF success affects

Table 4: The impact of parents' first IVF succeeding on having siblings (first stage)

	Having at least one sibling					
	Math test sample	Reading test sample	Agreeable sample	Consc. sample	Emotional stability sample	School happiness sample
Panel A: relative to first attempt failing						
Success	0.344	0.338	0.324	0.323	0.323	0.324
	0.005	*0.005*	*0.006*	*0.006*	*0.006*	*0.006*
Observations	8,308	10,906	7,689	7,666	7,567	7,783
Panel B: relative to all attempts failing						
Success	0.555	0.546	0.537	0.538	0.537	0.538
	0.008	*0.008*	*0.010*	*0.010*	*0.011*	*0.010*
Observations	5,961	7,783	5,379	5,350	5,283	5,446

Note–Estimates of the average effect of parents' first IVF attempt succeeding on the likelihood of having at least one sibling. These estimates reflect complier shares (Panel A) and relier shares (Panel B). Standard errors are shown in italics.

the likelihood that the first-born has siblings. Panel A reports the complier first stage. We find a first-stage estimate of 0.34 which implies that 34% of the sample are compliers. This also means that 66% of all first borns whose parents undergo IVF for a second child would have a sibling even if the first attempt fails. These are, in our setup, the always-takers. Panel B shows the relier first stage. We now find a first-stage estimate of 0.55 which means that 55% of all the first-born children in our sample are reliers. The remaining 45% would end up with a sibling even if all IVF attempts failed. We call these first-born children non-reliers. These first-stage results together suggest that our improved approach covers a group that is 1.6 times larger than the standard instrumental variable approach. Since the potential bias from exclusion violations depends on the ratio of always-takers to compliers or non-reliers to reliers, the improved approach reduces the potential for bias by more than half.

Table 5 contains the estimates which are all intended to identify the causal effect of having siblings among children whose parents undergo IVF treatment for a second child. The different panels represent the effect estimates identified under different sets of assumptions.

Table 5: The impact of having siblings on school outcomes in IVF families

	Math test	Reading test	Agreeable	Consc.	Emotional stability	School happiness
Panel A: The average impact for compliers, 2SLS (under A1 & A2)						
Having siblings	-0.054	-0.049	-0.038	-0.025	-0.075	-0.118
	0.069	*0.059*	*0.068*	*0.071*	*0.071*	*0.068*
Observations	8,308	10,906	7,689	7,666	7,567	7,783
Panel B: The average impact for compliers, IPW (under A1 & A2)						
Having siblings	-0.051	-0.035	-0.034	-0.026	-0.076	-0.096
	0.059	*0.046*	*0.060*	*0.061*	*0.075*	*0.064*
Observations	8,308	10,906	7,689	7,666	7,567	7,783
Panel C: The average impact for reliers (under A3 & A4)						
Having siblings	-0.024	-0.006	-0.007	-0.008	-0.016	-0.031
	0.045	*0.035*	*0.045*	*0.047*	*0.049*	*0.042*
Observations	5,961	7,783	5,379	5,350	5,283	5,446
Panel D: The average impact for all children in the sample (under A3 & A5)						
Having siblings	-0.013	-0.003	-0.004	-0.005	-0.008	-0.016
	0.025	*0.019*	*0.024*	*0.025*	*0.026*	*0.023*
Observations	5,961	7,783	5,379	5,350	5,283	5,446

Note–Estimates of the average effect of having siblings on first-born children under different sets of assumptions (described in the main text). The outcomes are standardized test scores in math and reading (columns 1 and 2); standardized personality traits—agreeableness, conscientiousness, and emotional stability (columns 3, 4, and 5); and a standardized measure of overall school well-being (column 6). Standard errors are shown in italics.

Panels A and B present the 2SLS and IPW second-stage estimates. The differences between the two are small and statistically insignificant. If IVF success at the first attempt is a valid instrument, these estimates capture the causal effect of having siblings for complier children (regardless of the estimation strategy used). Compared to the positive associations reported in Table 3, we find that all estimates reverse sign and become statistically insignificant. Although we no longer find evidence that siblings significantly impact the cognitive and noncognitive outcomes of first-born children, the estimates are in most cases not precise enough to draw firm conclusions, leaving open the possibility that siblings may still matter.

Panel C presents the improved second-stage estimates, identified under the sequential independence and nonrelier exclusion assumptions. If both assumptions hold, the estimates capture the causal effect for relier children. These effect estimates remain statistically insignificant, are smaller in magnitude, and are estimated with greater precision than those in Panels A and B. This indicates that siblings do not have a meaningful impact on the school outcomes of first-born children. If any causal effect exists, it is weak at best.

Panel D shows the estimates under the sequential independence and late sibling exclusion assumptions, representing the effect for all children in the sample. All coefficients are near zero, precisely estimated, and statistically insignificant, reinforcing the conclusion that there is no causal relationship between the cognitive and noncognitive school outcomes of first-born children and the presence of siblings.

If we were asked to choose among the different estimates, we would prefer those in Panels C and D. These estimates apply to a larger share of the sample, are substantially more precise, and are less vulnerable to violations of the exclusion restriction than those in Panels A and B. Moreover, the same estimates under alternative, and possibly weaker, homogeneity and monotonicity assumptions serve as lower and upper bounds of the effect of having siblings. Based on these preferred estimates, which are all close to zero, we conclude that having siblings has little to no impact on the cognitive or non-cognitive development of school-aged first-born children. More specifically, based on the corresponding 95% confidence intervals, we can rule out sibling effects larger than 0.1 standard deviations away from zero for most outcomes (or even 0.05 standard deviations away from zero if we consider the estimates in Panel D).

Table 6: The impact of having siblings on school outcomes: Robustness

	Math test	Reading test	Agreeable	Consc.	Emotional stability	School happiness
Panel A: no controls						
Having siblings	-0.037	-0.042	-0.004	-0.024	-0.004	-0.015
	0.024	*0.020*	*0.023*	*0.022*	*0.024*	*0.025*
Observations	5,977	7,800	5,403	5,375	5,315	5,471
Panel B: baseline controls and non-treatment controls						
Having siblings	-0.038	-0.024	-0.010	-0.013	-0.010	-0.018
	0.022	*0.018*	*0.023*	*0.027*	*0.025*	*0.022*
Observations	5,961	7,783	5,379	5,350	5,283	5,446
Panel C: baseline controls, 4 attempts						
Having siblings	-0.013	0.001	-0.009	-0.007	-0.008	-0.015
	0.023	*0.018*	*0.023*	*0.023*	*0.026*	*0.022*
Observations	6,145	8,028	5,569	5,539	5,470	5,636
Panel D: baseline controls, 6 attempts						
Having siblings	0.002	0.004	0.006	-0.000	0.007	-0.000
	0.025	*0.021*	*0.024*	*0.025*	*0.027*	*0.026*
Observations	5,867	7,642	5,286	5,258	5,192	5,352

Note–Results use different modifications of the baseline specification for the average effects under sequential independence and late sibling exclusion assumptions. In panels A and B, we use no-control and full-control inverse probability weights for IVF success. The full-control specification augments the baseline controls with all parental and child characteristics at first child birth from Table 1. In panels C and D, we cap the maximum number of IVF attempts used for identification at 4 and 6. Standard errors are shown in italics.

4.3 Robustness

To probe the robustness of our sibling effect estimates, we explore several alternative specifications. Table 6 reports results when we allow either a narrow or broad set of observables to correlate with IVF success and potential outcomes (Panels A and B), and when we vary the maximum number of IVF attempts used for identification (Panels C and D). We focus on the estimates that leverage the sequential independence and late sibling exclusion assumptions, using the estimates in Table 5, Panel C, as a benchmark.

We begin by examining sensitivity to the inclusion of control variables. The baseline specification adjusts for parental age and calendar time, which are the key medical predictors of IVF success. In Panel A, we omit these controls. Compared to the estimates in Table 5, Panel C, the estimated sibling effects are slightly larger in magnitude but remain statistically indistinguishable from zero.

In Panel B, we expand the standard set of controls to include additional pre-treatment covariates: the first-born's age, gender, birth weight, birth length, and IVF status, along with indicators of parental college education, employment, and pre-attempt income. While these variables may not directly affect outcomes or IVF success, they may proxy for unobserved factors related to both. As with the no-control estimates, the full-control estimates are slightly larger in magnitude but less precise.

It is important to note that adding more controls in IPW settings does not necessarily yield more credible estimates. If the additional covariates are unrelated to IVF success, they may simply introduce noise and reduce precision. Consistent with this, the standard errors in this specification are larger than those in Table 5, Panel C.

Next, we examine how the sibling effect estimates vary with different thresholds for the maximum number of IVF attempts used for identification. Our baseline results are based on the first five attempts. In the sample, 90%, 94%, and 97% of families undergo no more than four, five, and six attempts, respectively. The estimates remain stable when the threshold is lowered to four (Panel C) or raised to six (Panel D).

4.4 Heterogeneity

So far, we have focused on the average effect of having siblings on first-born children's school outcomes in the IVF sample. In our heterogeneity analysis, we explore whether these effects vary across different subgroups of first-borns. We begin by examining dimensions that may influence the relevance of the resource dilution and sibling socialization channels, as well as those that help further address potential exclusion restriction violations. We then consider margins that inform the generalizability of our findings to non-IVF families.

First, we restrict the sample to subgroups defined by child sex, maternal education, and household income. More specifically, we consider boys, girls, children of less educated mothers, and children from low-income couples. Gender may matter if boys and girls differ in how they mature and learn, and therefore in how they respond to siblings. Maternal education could influence the ability to substitute for or reinforce sibling socialization through other forms of support. Finally, resource constraints in low-income families may amplify the role of resource dilution, which could be offset by socialization benefits in more advantaged families.

We also examine subsamples defined by child and maternal age at the time of the first IVF attempt, which helps us further to address potential violations of the exclusion restriction. Older children spend more time as only children, making them better suited for identifying sibling effects under the late-sibling exclusion condition (**A5**). Among those aged three or older at the time of their parents' first IVF attempt, IVF failure increases the average sibling age gap from 6.3 to 8.5 years. Older mothers are less likely to conceive naturally, increasing the share of reliers and thereby strengthening the first stage while reducing bias from natural conceptions. In the subsample of mothers aged 35 and above, the relier share rises to 0.69–0.70, which is substantially higher than the 0.54–0.55 observed in the full sample (Table 5, Panel B).

Another margin considers sibling impacts at the lower end of the outcome distributions. This is motivated by two considerations. First, the relatively high average outcomes in the IVF sample (Table 1) may lead to possible ceiling effects, compressing variation at the top and obscuring meaningful impacts. Second, children in the IVF sample generally outperform those in the representative sample. By focusing on lower-performing children, we may provide estimates that are more informative for the broader population.

And relatedly, we may further improve the relevance for the broader population by estimating average sibling effects in those IVF families that resemble representative families (in being less affluent, less educated, and younger). In addition, we restrict the sample to naturally conceived first-borns, women without diagnosed fertility issues (i.e., where infertility stems from the partner), and second-born singletons. Focusing on these subgroups makes the IVF sample more comparable to the general population, further improving the external validity of our estimates.

Table 7A presents average sibling effect estimates by these subgroups: boys (Panel A), girls (Panel B), children of less educated mothers (Panel C), children from low-income families (Panel D), older first-borns (Panel E), and children of older mothers (Panel F). Table 7B complements this with estimates for the bottom quartile of the outcome distribution (Panel G); average effects in a reweighted sample that matches the representative population on family characteristics at first birth, as reported in

Table 7A: The impact of having siblings on school outcomes: Heterogeneity

	Math test	Reading test	Agreeable	Consc.	Emotional stability	School happiness
Panel A: first-born sons						
Having siblings	0.002	-0.032	-0.024	-0.010	0.003	-0.025
	0.063	*0.052*	*0.063*	*0.066*	*0.061*	*0.060*
Observations	3,013	3,976	2,720	2,701	2,686	2,768
Panel B: first-born daughters						
Having siblings	-0.052	0.040	0.007	-0.003	-0.000	-0.008
	0.059	*0.060*	*0.058*	*0.061*	*0.075*	*0.063*
Observations	2,897	3,776	2,641	2,631	2,583	2,659
Panel C: less educated mothers						
Having siblings	0.024	0.024	0.047	0.056	0.045	0.048
	0.067	*0.055*	*0.066*	*0.060*	*0.065*	*0.056*
Observations	2,941	3,814	2,549	2,540	2,504	2,586
Panel D: low-income couples (bottom 50% of joint income distribution)						
Having siblings	-0.053	-0.048	0.027	0.013	0.022	0.010
	0.063	*0.063*	*0.063*	*0.067*	*0.068*	*0.067*
Observations	3,052	3,986	2,581	2,559	2,533	2,617
Panel E: older first-born children (3 or older at first IVF)						
Having siblings	0.047	0.035	-0.003	0.065	0.007	0.036
	0.061	*0.045*	*0.058*	*0.059*	*0.051*	*0.057*
Observations	3,034	3,947	2,567	2,552	2,507	2,607
Panel F: older mother (35 or older at first IVF)						
Having siblings	-0.016	0.018	-0.017	-0.031	-0.071	-0.013
	0.056	*0.044*	*0.047*	*0.052*	*0.051*	*0.051*
Observations	2,844	3,814	2,522	2,513	2,477	2,552

Note–All panels present results from the baseline specification under sequential independence and late sibling exclusion assumptions. Standard errors are shown in italics.

Table 7B: The impact of having siblings on school outcomes: Heterogeneity

	Math test	Reading test	Agreeable	Consc.	Emotional stability	School happiness
Panel G: effect on first quartile						
Having siblings	0.002	0.005	-0.073	-0.032	-0.011	-0.031
	0.029	*0.031*	*0.033*	*0.061*	*0.057*	*0.048*
Observations	5,961	7,783	5,379	5,350	5,283	5,446
Panel H: re-weighted to match representative sample						
Having siblings	0.005	0.024	-0.027	0.009	0.011	-0.011
	0.032	*0.028*	*0.035*	*0.034*	*0.039*	*0.034*
Observations	5,961	7,783	5,379	5,350	5,283	5,446
Panel I: first child is non-IVF						
Having siblings	0.008	0.031	-0.080	-0.013	0.017	-0.031
	0.054	*0.050*	*0.067*	*0.060*	*0.062*	*0.056*
Observations	3,262	4,182	2,837	2,824	2,778	2,878
Panel J: fathers infertile						
Having siblings	0.093	-0.007	0.022	0.034	0.050	0.042
	0.077	*0.065*	*0.070*	*0.073*	*0.067*	*0.075*
Observations	2,389	3,103	2,174	2,167	2,143	2,204
Panel K: no twins						
Having siblings	-0.013	0.002	-0.030	0.005	-0.024	-0.019
	0.044	*0.040*	*0.044*	*0.044*	*0.052*	*0.050*
Observations	5,515	7,218	4,999	4,973	4,905	5,062

Note–Panel G presents estimates for the effect on the first quartile of the outcome distribution under sequential independence and late sibling exclusion assumptions. Other panels present results from the baseline specification under sequential independence and late sibling exclusion assumptions. Standard errors are shown in italics.

Table 1 (Panel H); and estimates for naturally conceived first-borns (Panel I), children with infertile fathers (Panel J), and families where the second-born child is a singleton (Panel K).

Across subgroups, we find that nearly all sibling effect estimates remain consistently close to zero and statistically indistinguishable from one another. Some estimates are larger in magnitude, but these tend to be imprecise, lack a consistent pattern, or both.

Overall, these findings reaffirm our earlier conclusion: that is, sibling effects are consistently near zero across all outcomes and subgroups. This consistency suggests that having siblings has little impact on both cognitive and non-cognitive outcomes of first-born children. Moreover, the results also speak to the external validity of our conclusions: that is, first-born children who resemble those in the general population (in terms of conception, maternal health, sibling structure, and parental characteristics) also show no meaningful benefit or harm from having siblings.

5 Conclusions

Our study focuses on only children. This is an increasingly important yet often overlooked group of children in the economics literature on sibling composition and child outcomes. Specifically, we assess how the outcomes of only children would change if they were exposed to siblings. By leveraging the apparent randomness in success at first and later IVF treatments in an administrative dataset covering all Danish one-child families who sought IVF for a second child, we estimate the causal effect of being an only child on a range of cognitive and non-cognitive outcomes. This question is particularly relevant today: with fertility rates declining globally, the number of only children is at an all-time high, making it essential to understand the long-term consequences of this demographic shift.

When we consider simple associations, we find that first-born children without siblings perform worse on cognitive tests, are more neurotic, less conscientious and agreeable, and report lower levels of happiness than first-born children with siblings. These only child associations are sizable and statistically significant. However, when we move beyond associations and turn to causal inference, we find that all the effect estimates are close to zero and statistically insignificant, suggesting that being the only child (and thus being deprived of siblings) has no meaningful impact on the cognitive and non-cognitive outcomes of first-born children.

While our findings suggest that first-born children neither benefit nor suffer from siblings, it is important to emphasize that our estimates are realized within Denmark: a country with a generous education system and well-established child care institutions. Suppose, for instance, that sibling effects are negative and driven by resource competition. It is possible that Danish children do not suffer much from siblings because education in Denmark is heavily subsidized and the returns to skill acquisition are relatively low.[16] If parents face fewer constraints and invest less in child skills, Becker's quantity–quality model predicts smaller effects of additional children on child outcomes. In this sense, our findings are comparable to those of Black et al. (2005), who argue that financial constraints are not important and show that children raised in Norway are not worse off with more siblings.

Suppose, on the other hand, that sibling effects are positive and driven by social interactions between siblings. In this setting, it is again possible that only children do not benefit because, in Denmark, children have access to subsidized child care from age one. Only children, although being deprived from siblings, are not isolated and may face ample socialization opportunities in child care from an early age.

The lack of causal effects, combined with the strong negative associations between being an only child and child outcomes, suggests that only children are a negatively selected group. While identifying the precise sources of negative selection into one-child families is beyond the scope of our study, we offer some speculative explanations. One possibility is the optimal stopping rule: if a first-born child exhibits unfavorable outcomes, parents may choose not to have another child. Another possibility relates to relationship instability: if only children are more likely to grow up in divorced families, they may have worse school outcomes (because of disrupted family life) and may be more likely to remain only children (because divorce reduces subsequent fertility).

In closing, our study highlights only children as an increasingly relevant group in the context of global demographic change. By leveraging quasi-random variation in IVF success at first and later treatments, we obtain relatively precise effect estimates of having at least one sibling. We conclude that, in the Danish context, the presence or absence of siblings does not impact the cognitive and non-cognitive development

[16] Harmon et al. (2001) provide a cross-country analysis of the returns to education in Europe and document one of the lowest returns in Denmark.

of first-born children.

Appendix

We start with reduced form for reliers:

$$E[Y_1(1) - Y_0(0)|R{=}1]\,\Pr[R{=}1] + E[Y_1(1) - Y_0(1)|R{=}0]\,\Pr[R{=}0].$$

It consists of two terms. The first term captures the effect of whether siblings arrive for reliers. The second term is the bias term and captures the effect of when siblings arrive for non-reliers.

We next rewrite the bias term for non-reliers:

$$E[Y_1(1) - Y_0(0)|R{=}0]\,\Pr[R{=}0] - E[Y_0(1) - Y_0(0)|R{=}0]\,\Pr[R{=}0],$$

which we then substitute in the relier reduced form:

$$E[Y_1(1) - Y_0(0)] \times \left[1 - \frac{E[Y_0(1) - Y_0(1)|R{=}0]}{E[Y_1(1) - Y_0(1)]}\,\Pr[R{=}0]\right].$$

Under conditions (**A6**) and (**A7**), we can express the modified reduced form as:

$$E[Y_1(1) - Y_0(0)] \times [1 - \theta\,\Pr[R{=}0]],$$

where θ is the ratio between the effects of having siblings late and early for our sample of non-relier first-born children. This ratio is assumed larger than 0 and smaller than 1. This implies that the reduced form for reliers act as an upper bound for the average effect of having siblings because we multiply the average effect of having siblings with something that is less than 1. This also implies that the second stage, which we can write as:

$$E[Y_1(1) - Y_0(0)] \times \left[\frac{1 - \theta\,\Pr[R{=}0]}{1 - \Pr[R{=}0]}\right],$$

acts as a lower bound because we now multiply the average effect of having siblings with something that is larger than 1.

References

Andersen, S. C., Gensowski, M., Ludeke, S. G., & John, O. P. (2020). A stable relationship between personality and academic performance from childhood through adolescence. An original study and replication in hundred-thousand-person samples. *Journal of Personality*, *88*(5), 925–939.

Angrist, J., & Imbens, G. (1995). Identification and estimation of local average treatment effects. *National Bureau of Economic Research*.

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

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

Becker, G. S. (1960). An economic analysis of fertility. In *Demographic and economic change in developed countries.* Columbia University Press.

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

Becker, G. S., & Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics*, *4*(3, Part 2), S1–S39.

Bensnes, S., Huitfeldt, I., & Leuven, E. (2023). Reconciling estimates of the long-term earnings effect of fertility. *IZA Discussion Paper*.

Bhalotra, S., & Clarke, D. (2019). Twin birth and maternal condition. *Review of Economics and Statistics*, *101*(5), 853–864.

Bhattacharjee, N., & others. (2024). Global fertility in 204 countries and territories, 1950–2021, with forecasts to 2100: A comprehensive demographic analysis for the global burden of disease study 2021. *The Lancet*, *403*(10440), 2057–2099.

Bitler, S. L., Marianne P. (2012). Utilization of infertility treatments: The effects of insurance mandates. *Demography*, *49*, 125–149.

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

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2007). From the cradle to the labor market? The effect of birth weight on adult outcomes. *The Quarterly Journal of Economics*, *122*(1), 409–439.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Older and wiser? Birth order and IQ of young men. *Cesifo Economic Studies*, *57*(1), 103–120.

Blake, J. (1985). Number of siblings and educational mobility. *American Sociological Review*, 84–94.

Bögl, S., Moshfegh, J., Persson, P., & Polyakova, M. (2024). The economics of infertility: Evidence from reproductive medicine. *National Bureau of Economic Research*.

Booth, A. L., & Kee, H. J. (2009). Birth order matters: The effect of family size and birth order on educational attainment. *Journal of Population Economics*, *22*(2), 367–397.

Buckles, K. S., & Munnich, E. L. (2012). Birth spacing and sibling outcomes. *Journal of Human Resources*, *47*(3), 613–642.

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

Cameron, L., Erkal, N., Gangadharan, L., & Meng, X. (2013). Little emperors: Behavioral impacts of China's one-child policy. *Science*, *339*(6122), 953–957.

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

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

Doepke, M., Hannusch, A., Kindermann, F., & Tertilt, M. (2023). The economics of fertility: A new era. *Handbook of the Economics of the Family*, *1*(1), 151–254.

Falbo, T. (2012). Only children: An updated review. *Journal of Individual Psychology*(1).

Feyrer, J., Sacerdote, B., & Stern, A. D. (2008). Will the stork return to europe and japan? Understanding fertility within developed nations. *Journal of Economic Perspectives*, *22*(3), 3–22.

Fong, V. L. (2004). *Only hope: Coming of age under China's one-child policy.* Stanford University Press.

Gallen, Y., Joensen, J. S., Johansen, E. R., & Veramendi, G. F. (2023). The labor market returns to delaying pregnancy. *Available at SSRN 4554407*.

Gary-Bobo, R. J., Picard, N., & Prieto, A. (2006). Birth order and sibship sex composition as instruments in the study of education and earnings. *CEPR Discussion Paper*.

Goldin, C. (2024). Babies and the macroeconomy. *National Bureau of Economic Research*.

Goux, D., & Maurin, E. (2005). The effect of overcrowded housing on children's performance at school. *Journal of Public Economics*, *89*(5-6), 797–819.

Groes, F., Houštecká, A., Iorio, D., & Santaeulalia-Llopis, R. (2024). The unequal battle against infertility: Theory and evidence from IVF success. *CEPR Discussion Papers*.

Guo, R., Yi, J., Zhang, J., & Zhang, N. (2020). Rationed fertility: Treatment effect heterogeneity in the child quantity-quality tradeoff. *Unpublished Manuscript*.

Harmon, C., Walker, I., & Westergærd-Nielsen, N. C. (2001). *Education and earnings in europe: A cross country analysis of the returns to education*. Edward Elgar Publishing.

Houmark, M. A. (2023). First among equals? How birth order shapes child development. *Unpublished Manuscript, Aarhus University*.

Ilciukas, J. (2024). Parenthood timing and gender inequality. *Unpublished Manuscript, University of Amsterdam*.

Kantarevic, J., & Mechoulan, S. (2006). Birth order, educational attainment, and earnings: An investigation using the PSID. *Journal of Human Resources*, *41*(4), 755–777.

Kristensen, P., & Bjerkedal, T. (2007). Explaining the relation between birth order and intelligence. *Science*, *316*(5832), 1717–1717.

Li, B., & Zhang, H. (2017). Does population control lead to better child quality? Evidence from China's one-child policy enforcement. *Journal of Comparative Economics*, *45*(2), 246–260.

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

Liu, H. (2014). The quality–quantity trade-off: Evidence from the relaxation of China's one-child policy. *Journal of Population Economics*, *27*, 565–602.

Lundberg, S., & Pollak, R. A. (2007). The American family and family economics. *Journal of Economic Perspectives*, *21*(2), 3–26.

Lundborg, P., Plug, E., & Rasmussen, A. W. (2017). Can women have children and a career? IV evidence from IVF treatments. *American Economic Review*, *107*(6), 1611–1637.

Lundborg, P., Plug, E., & Rasmussen, A. W. (2024a). Is there really a child penalty in the long run? New evidence from IVF treatments. *IZA Discussion Paper*.

Lundborg, P., Plug, E., & Rasmussen, A. W. (2024b, 12). On the family origins of human capital formation: Evidence from donor children. *The Review of Economic Studies*, rdae101. Retrieved from https://doi.org/10.1093/restud/rdae101 doi: 10.1093/restud/rdae101

Miller, G. (2010). Contraception as development? New evidence from family planning in colombia. *The Economic Journal*, *120*(545), 709–736.

Mogstad, M., & Wiswall, M. (2016). Testing the quantity–quality model of fertility: Estimation using unrestricted family size models. *Quantitative Economics*, *7*(1), 157–192.

Monden, C., Pison, G., & Smits, J. (2021). Twin peaks: More twinning in humans than ever before. *Human Reproduction*, *36*(6), 1666–1673.

Mueller, G., & Plug, E. (2006). Estimating the effect of personality on male and female earnings. *Industrial and Labor Relations Review*, *60*(1), 3–22.

Niederberger, C., & others. (2018). Forty years of IVF. *Fertility and Sterility*, *110*(2), 185–324.

Pettersson-Lidbom, P., & Skogman Thoursie, P. (2009). Does child spacing affect children's outcomes? Evidence from a Swedish reform. *Stockholm University Working Paper*(No. 2009: 7.).

Qian, N. (2009). Quantity-quality and the one child policy: The only-child disadvantage in school enrollment in rural china. *National Bureau of Economic Research*.

Räsänen, T. (2023). How does motherhood affect women's careers? Causal estimates using medically induced ovulation treatments. *Unpublished Manuscript, University of Jyväskylä*.

Rosenwaks, Z., Davis, O. K., & Damario, M. A. (1995). The role of maternal age in assisted reproduction. *Human Reproduction*, *10*(suppl_1), 165–173.

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

Templeton, A., Morris, J. K., & Parslow, W. (1996). Factors that affect outcome of in-vitro fertilisation treatment. *The Lancet*, *348*(9039), 1402–1406.

Thwaites, A., Hall, J., Barrett, G., & Stephenson, J. (2023). How common is natural conception in women who have had a livebirth via assisted reproductive technology? Systematic review and meta-analysis. *Human Reproduction*, *38*(8), 1590–1600.

van Loendersloot, L., Repping, S., Bossuyt, P., van der Veen, F., & van Wely, M. (2014). Prediction models in in vitro fertilization; where are we? A mini review. *Journal of Advanced Research*, *5*(3), 295–301.

Wang, J., & Sauer, M. (2006). In vitro fertilization (IVF): A review of 3 decades of clinical innovation and technological advancement. *Therapeutics and Clinical Risk Management*, *2*(4), 355–364.

Xiao, Y. (2024). Heterogeneity in the child quality-quantity trade-off and intergenerational mobility. *CEPR Discussion Papers*.

Zajonc, R. B. (1976). Family configuration and intelligence: Variations in scholastic aptitude scores parallel trends in family size and the spacing of children. *Science*, *192*(4236), 227–236.

Zajonc, R. B., & Markus, G. B. (1975). Birth order and intellectual development. *Psychological Review*, *82*(1), 74.