The only child.

Julius Ilciukas, Petter Lundborg, Erik Plug, Astrid Würtz Rasmussen*

January 2025

Abstract

We estimate the impact of having siblings on school outcomes of first-born children. By leveraging exogenous variation in first and later IVF treatments, we construct an improved instrumental variable estimator that tackles exclusion violations and identifies causal effects for *compliers* and *always takers* with siblings from later treatments. With nationwide school surveys linked to administrative records, we find that first-born children with and without siblings perform equally well on nationwide reading and math tests, are equally conscientious, agreeable, and emotionally stable, and report the same levels of school well-being. We conclude that the cognitive and non-cognitive outcomes of school-aged first-born children neither benefit nor suffer much from having siblings.

*Ilciukas: Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (email: j.ilciukas@uva.nl). Lundborg: Department of Economics, Lund University, Tycho Brahes Väg 1, 223 63 Lund, Sweden (email: petter.lundborg@nek.lu.se). Plug: Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (email: e.j.s.plug@uva.nl). Rasmussen: Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus, Denmark (email: awr@econ.au.dk). We thank seminar/conference participants in Aarhus, Amsterdam, Bergen, Bologna, Brno, Berlin, Cologne, Copenhagen, Crieff (Schotland), Edinburgh, Laax (Switzerland), 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.

1 Introduction

With collapsing fertility rates worldwide, families with only one child are becoming increasingly common (Bhattachee et al. 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 are one-child families (National Bureau of Statistics 2007). And in Denmark, which is the country we 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 think 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. In particular, we ask whether first-born children (aged 7 to 16) do better or worse in school in case they have siblings. This question is not so easy to answer, however. On theoretical grounds, all answers are possible. Becker (1960), and later Becker and Lewis (1973) and Becker and Tomes (1976), for instance, argue that children without siblings do better in all those skills that parents consider desirable. In their quantity-quality model, they formalize that only children do not have to share the family resources with siblings, meaning that parents have more time and money to invest in their only-child's cognitive and non-cognitive development. In contrast, Zajonc and Markus (1975), and later Zajonc (1976), argue that only children may do better had they one or two siblings. In their confluence model, they stress that only children have no younger siblings to socialize with, which slows down their cognitive and non-cognitive development.[2] On empirical grounds, answers are generally not so informative. Most existing studies just compare the cognitive and non-cognitive outcomes of children with

[1] There are several explanations for the rise of one-child families including increased divorce rates which interrupted initial fertility plans, improved educational and labor market opportunities for women which increased the opportunity costs of raising children, delayed motherhood which shortened fertility windows, increased access to modern contraceptives which gave women greater reproductive control, and invasive one-child policies which forced parents to have just one child. Example studies for each explanation include Blake (1985), Fong (2004), Lundberg and Pollak (2007), Feyrer, Sacerdote and Stern (2008), and Miller (2010). For a more comprehensive review on the economics of falling fertility rates, we refer to Doepke et al. (2023) and Goldin (2024).

[2] At the same time, the confluence model also allows only children to do worse with siblings but only if the adverse impact of later-born siblings on the overall cognitive and non-cognitive family environment is strong enough.

and without siblings, which is not enough to make causal claims.[3] For this, we need a natural experiment that generates as-good-as-random variation in the likelihood that some children have siblings and others not.

The natural experiment we exploit is based on in vitro fertilization (IVF) treatments. We study all one-child families that undergo IVF treatment for a second child in Denmark. Because treatment success at the first IVF attempt is effectively random, we take the approach of Lundborg, Plug, and Rasmussen (2017, 2024) as point of departure and use success at the first IVF attempt as an instrumental variable (IV) to estimate the impact of having siblings on a wide range of school outcomes of first-born children.[4]

There are two concerns about these IV-IVF estimates. One is that the estimates capture causal effects but only if the exclusion condition holds. This condition says that success at the first IVF attempt only affects the school outccomes of first-born children through its impact of having a sibling. This condition gets violated when, for instance, the development of first-born children depends not only on whether but also on when siblings arrive, which happens when couples decide to continue treatment after a first failed attempt and get a second-born child at later attempts. The second one is that the estimates capture causal effects but only for *complier* children. These children are a subset of first-born children who would never have siblings in case the first treatment fails.

In this study, we propose a novel IVF-based method to address both concerns. In particular, we exploit the randomness at later IVF attempts to recover what the school outcomes for children with later-born siblings would have been had they remained an only child. These potential outcomes allows us, first, to get around possible exclusion violations caused by first-born children with siblings from later treatments and, second, to identify the causal effects of having siblings for a larger group of *relier* children. These are all the first-born children

[3]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" (2012 p. 47). We review the limited literature providing causal evidence in Section 2.

[4]Lundborg, Plug, and Rasmussen (2017, 2024) are the first to introduce this IV-IVF strategy albeit in another context: that is, they exploit IVF treatment success at the first IVF cycle among childless couples that undergo IVF treatment to estimate the causal effect of having children on the career of women (and men). Their main findings can be summarized as follows: first, IVF-induced fertility variation is arguably exogenous and unrelated to observable pre-treatment characteristics that are strong female career predictors, and second, women who are successfully treated by IVF experience a sizable earnings penalty shortly after the birth of the first child, but this earnings penalty fades out and is effectively zero for adolescent children.

who rely on either first or later treatments for siblings.

Our empirical analysis relies on nationwide school survey data combined with detailed administrative records. We first construct our primary sample of first-born children in families undergoing IVF treatment for a second child and demonstrate that IVF success at first and later attempts generate enough exogenous variation in the likelihood to have siblings. We then take our improved IV-IVF strategy to estimate the causal effect of being the only child on school performance, personality traits, and school happiness. We find that first-born children with and without siblings (for exogenous reasons) do equally well in terms of school test scores in math and reading, in terms of personalty traits agreeableness, conscientiousness, and emotional stability, and in terms of overall school happiness.

In sum, our study has three novel features. First, our focus is on only children; this is an increasingly important yet often neglected group of children. Second, our econometric framework identifies the average effect of being an only child for all children whose parents rely on both first and later IVF treatments for a second child; the effect estimates are valid, not just for compliers, but for a larger group of first-born children we call reliers. And third, our findings are quite intriguing; we show that, at least in the Danish context, the presence or absence of siblings does not have any substantial causal effect on the cognitive and non-cognitive development of first-born children.

The rest of the paper unfolds as follows. Section 2 embeds our study into the existing literature on siblings and fertility treatments. Section 3 describes the data, how we construct the estimation sample, and offers some relevant institutions regarding IVF treatments. Section 4 introduces the improved IV-IVF strategy to estimate only child effects. Section 5 presents our main set of results. And Section 6 concludes.

2 Previous literature

Sibling studies

Our study connects to a large empirical literature on four related sibling topics: sibling size, sibling order, sibling spacing, and siblinglessness. We provide a short review, focus on those studies that (aim to) estimate the causal effects of sibling size, birth order, sibling spacing, and lack of siblings, and discuss our

contributions.[5]

The first literature is concerned with estimating the causal relationship between sibling size and child quality (usually measured by cognitive test scores, educational attainment or earnings) using twin births and sibling sex composition as natural experiments. With twin births, researchers exploit that some parents end up with more children than planned; they identify sibling size effects from comparing outcomes of children with younger twin siblings and younger singleton siblings. With sibling sex composition (in combination with parental preferences for opposite-sex siblings), researchers exploit that parents with same-sex children are more inclined to have another child; they identify sibling size effects from 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 and Glauber 2006; Caceres-Delpiano 2006; Goux and Maurin 2005; Åslund and Grönqvist 2010), there is little evidence of any sizable sibling size impact on long-run outcomes such as educational attainment and earnings (Black, Devereux and Salvanes 2005; Åslund and Grönqvist 2010; Angrist, Lavy and Schlosser 2010; and De Haan 2010).[6] A recent study (Mogstad and Wiswall 2018) calls these negligible sibling size impacts into question and suggests that the causal relationship between sibling size and child outcomes need not be linear or even monotone. Our study offers new insights on sibling size effects measured at a different and possibly a more relevant margin; we estimate the effect of having siblings (moving from zero to one sibling) as opposed to the estimated effect of having one more sibling (moving from two to three, or three to four siblings) in the twin and same-sex sibling studies.

The second literature estimates birth order effects on child quality using sibling-fixed effects specifications. With sibling-fixed effects, researchers estimate birth-order effects within families, controlling for any observable and unobservable family-specific characteristics that are affecting all siblings in the

[5]There is much empirical work that estimate associations between child outcomes, sibling size, birth order, spacing, and being the only child. Because these associations do not make a distinction between selection and causation, their interpretation remains unclear. For references on these sibling studies, we refer to literature surveys of Blake (1989) and Falbo and Polit (1986, 1987).

[6]These results, obtained with data from the US, France, Norway, and Sweden, may not hold in a more developing context. Comparable twin-IV studies indeed find negative sibling size impacts in India (Rosenzweig and Wolpin 1980), Turkey (Dayioglu, Kirdar, and Tansel 2009), and Brazil (Ponczek and Portela-Souza 2014).

same way. These studies generally find that first-born children do much better than later-born children in terms of cognitive development, educational attainment, and earnings (Black, Devereux and Salvanes 2005 2011; Conley and Glauber 2006; Gary-Bobo, Picard and Pietro 2006; Kantarevic and Mechoulan 2006; Kristensen and Bjerkedal 2007; Booth and Kee 2009; De Haan 2010).[7] Interestingly, similar birth order effects are observed in a recent study using the same data sources and outcomes as us, that is, first-born children perform better in both reading and math, are more conscientious, agreeable, and emotionally stable, and they report higher well-being in school (Houmark 2023). While only children are also first-born children, these sibling-fixed-effects studies are not informative about only children; that is, birth-order effects in sibling-fixed-effects specifications can only be estimated on samples of children with siblings. Our study is different in that we contrast the outcomes of first-born children that have siblings or not (for exogenous reasons).

The third literature estimates sibling spacing effects using parental benefit reforms and miscarriages as natural experiments (Pettersson-Lidbom and Skogman-Thoursy 2009; Buckles and Munnich 2012). These studies find that child cognition benefits from having siblings later. Like the previous birth order studies, these studies ignore only children; that is, these studies exploit exogenous variation in the timing of sibling births in samples of children with siblings. In contrast, our study shifts the focus from when siblings are born to whether siblings are born at all.

The fourth literature, which is the one most closely connected to our study, estimates the only child effect on various outcomes using China's one-child policy as natural experiment. This policy, which was introduced in 1979, enforced families to have only one child by, among other things, levying fines on families that had a second child. These fines, which were sometimes as high as several years of family income (Ebenstein 2010), were set differently for families belonging to different ethnicity groups in different regions at different times. There were also exemptions. Families with twins, for instance, were spared from fines. With China's one-child policy, researchers have exploited twin, ethnicity, regional or time-induced policy variation to identify the only child effect. In the context of

[7] Again, these results unlikely hold in a developing context. Comparable sibling-fixed-effects studies, for instance, find that first-born children do worse in the Philippines (Ejrnaes and Pörtner 2004), Nepal (Edmonds 2006), Nicaragua and Guatemala (Dammert 2010), and Ecuador (De Haan, Plug, and Rosero 2014). A likely reason for these opposite birth order effects is that poor parents send their first-born children out to work as soon as they can, so they can contribute to the family resources and lift the education constraints for later-born children.

(cognitive) school outcomes, the available studies produce mixed results; that is, the only child effects are sometimes positive (Li, Zhang, and Zhu 2008; Rosenzweig and Zhang 2009; Li and Zhang 2017), sometimes weak and only positive for some children (Liu 2014; Xiao 2021), and sometimes negative (Qian 2009).[8] In the context of non-cognitive outcomes, only one causal study estimates the impact of siblings on, among other things, personality traits. Cameron et al. (2013) find that first-born children born under China's one-child policy are significantly more pessimistic, less conscientious, and possibly more neurotic. We recognize, however, that these only child results from China may not generalize to more developed countries where families do not rely (or rely much less) on their children for their pensions, social security, and health care. In this sense, our study is the first to estimate causal only child effects on cognitive and non-cognitive outcomes in the context of a developed country.

IVF studies

Our study also connects to a vast literature on IVF treatments. Given our use of IVF treatments as instrumental variable in IV strategies, we organize this literature with the standard IV conditions in mind. The first one is relevance, which says that success at the first IVF attempt is a strong fertility predictor. This condition is met almost by construction given that IVF is the leading medical intervention to help infertile couples to get pregnant and conceive children. The second one is independence, which asserts that treatment success does not depend on the pretreatment characteristics of IVF treated women. It is reasonably well known that this condition does not hold unconditionally. One success dimension, for instance, is age, where successfully treated women are typically younger than unsuccessfully treated women. This is in line with a vast medical literature that points to age as the single most important factor determining success in IVF treatments (Rosenwaks et al. 1995, Templeton et al. 1996, van Loendersloot et al. 2014). Another dimension in which treated women sometimes differ is education (Groes et al. 2023). We will argue that treatment success is conditionally random and show later that accounting for the parents'

[8] Rosenzweig and Zhang (2009) estimate only child effects by comparing, among other things, the school outcomes of first-birth singletons and first-birth twins. This approach is problematic, as twins and singletons are inherently different. Bharadwaj, Lundborg, and Rooth (2018) consider neonatal differences and show that twins have much lower birth weights (and that low birth weight adversely impacts long-run earnings). Bhalotra and Clarke (2019) consider background differences and show that healthier and better educated mothers are more likely to give birth to twins.

age and year of treatment already eliminates any meaningful difference in what we consider the most relevant parent and child pre-treatment characteristics.

The last IV condition is exclusion, which asserts that treatment success impacts the school outcomes of first-born children only through the impact of having siblings. Again, this condition may not always hold if we think of treatment success, or the absence thereof, as a possible driver for depression (affecting the parenting skills of mothers), divorce (changing the rearing environment of children), and delayed fertility (postponing the arrival of siblings). Regarding IVF-induced depression and divorce, we show in earlier work that IVF treatment success among IVF treated couples in Denmark has negligible impacts on antidepressant medication and divorce (Lundborg et al. 2017). Recent studies from Sweden, Norway, Finland, and The Netherlands using IVF-based identification methodologies also find that mental health effects are small and short-lived (Bensnes et al. 2023, Räsänen 2023, Bögl et al. 2024, and Ilciukas 2024).[9] This is also in line with a vast medical literature. With self-assessed depression scores, most survey studies find that treated women adjust well to unsuccessful IVF treatments (Verhaak et al. 2007). And with medical depression indicators (antidepressant medication, psychiatry visits, or hospitalizations for mental disorders), register-based studies find that depression rates are generally low among IVF treated women (Agerbo et al. 2013, Baldur-Feskov et al. 2013, Pedro et al. 2019, Yli-Kuha et al. 2010). Regarding having siblings later, we refer to the sibling spacing literature reviewed above.

For tackling possible IVF-induced exclusion violations in child penalty studies, we conclude this review with four recent studies that relax the exclusion condition in different ways. Lundborg et al. (2024) take a long-run perspective and estimate child penalties 25 years after treatment arguing that depression, divorce, and delayed fertility among IVF-treated women cause foremost short-run violations. Bensnes et al. (2023), as well as Gallen et al. (2023), estimate child penalties by age of the child and get around delayed fertility violations by imposing parametric conditions on how child penalties vary by the mother's age at first treatment and the time elapsed since first treatment. Ilciukas (2024) estimates bounds on child penalties in the presence of delayed fertility violations

[9] Bögl et al. (2024) also look at women who remained childless for five years following their first IVF failure. While they find a notable negative impact on mental health after IVF failure, the specific nature of this sample complicates its interpretation. These women were selected based on both initial and subsequent failed attempts. Since the decision to pursue further treatment after any failed attempt is endogenously determined, this group of childless women, five years after their first attempt, may not be the most suitable comparison to those who succeeded after their first IVF attempt.

by exploiting (conditional) independence at later IVF attempts. Our approach differs in terms of focus and approach. Our focus is on first-born children and how they respond to the arrival of siblings. Our approach follows Ilciukas in that we exploit independence at later IVF attempts but differs in that we point-identify only child effects under arguably weaker conditions than the restrictive exclusion condition.

3 IVF institutions and data

In our analysis we combine information from several Danish administrative registers, and nationwide school well-being survey and school test data. Below we briefly discuss the relevant IVF institutions, present the main variables under study drawn from several data sources, describe our primary sample of children in one-child families that enter IVF treatment for a second child, and evaluate the independence properties of first and later IVF treatments.

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 3 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 (they don't have to pay) and one-child couples with spare embryos from earlier treatments (they pay considerably less).

IVF is a relatively effective treatment, with an average first-treatment success rate of 32-35 percent for the one-child couples in our study. Two factors are noteworthy here. First, the decision to stop treatment after a failed attempt tends to be selective; among those who continue treatment we expect to see couples with more resources and/or a stronger desire for a second child. Second,

the one-child couple's infertility status is not always an irreversible condition; among those who stop treatment we see that about 20 percent of the couples that ended treatment eventually conceive naturally of a second child (Thwaites et al. 2023).

Main variables

For the cognitive and non-cognitive school outcomes, we rely on nationwide school test data and well-being surveys. From 2010 onwards, children in primary and lower secondary education take multiple tests in reading and math. The nationwide school test data contain the test scores for 4 reading tests (taken in grades 2, 4, 6, and 8) and 2 math test (taken in grades 3 and 6). We observe test scores between 2010 and 2021. Most first-born children in our sampling window take in total 4 to 5 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 cohort-grade-specific test score mean and standard deviation in the representative sample of first-born children that take the test. We define the school outcomes in math and reading test scores as the average of all available standardized test scores in reading and math.

From 2014 onwards, 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.[10] Children take the survey in grades 4 to 9. We have access to the 2014-2021 surveys. Each year each trait is assessed by the same 2 to 3 questions 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 5 point scale (running from almost always, often, now and then, rarely, to never). We use these answers to measure the child's personality traits in school.[11] We standardize each trait based on the answer average and stan-

[10]The commonly used five-factor model of personality structure assigns someone's personality to 5 core domains: agreeableness, conscientiousness, emotional stability, extraversion, and openness to experience. The well-being survey covers 3 of the 5 core domains. In the school context of the survey, agreeableness captures the child's tendency to be cooperative and empathetic towards 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).

[11]We note that these trait measures have been recently validated in a register-matched personality survey by Andersen et al. (2020). This study demonstrates that the answers to these questions correlate strongly with well-established big five personality measures, with fac-

dard deviation in the representative sample of first-born children that answer at least one trait question. We define the student's personality trait as the average of all their grade-specific standardized personality traits. And 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 5 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 with identifiers of all individuals residing in Denmark. We use this register to isolate all 7-16 aged first-born children, with links to their parents and siblings (in case 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 or not. With the identifiers of first-born children, we use the birth register to collect additional information on their birth weight (in 1,000 grams) and birth length (in centimeters). With the identifiers of parents, we use the education and tax registers to extract additional information on the parents' educational levels, labor market status, and annual income. We measure education by a college education dummy indicating whether parents spent at least two years in college. We measure pre-treatment employment status and income by taking averages over the annual time being employed and annual income (in 1,000,000 DKK) observed in the 4 years preceding the treatment year.

For the instrumental variables, we use the IVF register provided by the Danish Health Data Authority (Sundhedsdatastyrelsen). The IVF register records all IVF treatments taking 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 treatment stages (medication, egg and sperm extraction, embryo fertilization, and embryo insertion), treatment success, the date of treatment, and where applicable the date of birth. For each IVF attempt, we construct a success dummy indicating whether the treatment has been successful and lead to a live child birth at most 10 months after treatment. This has been validated with medical records in Lundborg et al. (2017). This success indicator for the first IVF attempt is the instrument in our instrumental variable design.

tor loadings confirming that the answers effectively capture core aspects of conscientiousness, agreeableness, and emotional stability.

Main estimation sample

We also use the IVF register to select our estimation sample of all first-born children raised in families that enter IVF treatment for a second child. To be included in this sample, we select all first-born (IVF and non-IVF) children in their primary and lower secondary school years, raised in families that were treated somewhere between 1995 and 2005, with at least one math test score, or one reading test score, or one response to the satisfaction question, or one trait-specific response in one of the nationwide survey waves, and non-missing control variables. We further exclude a small number of first-born children with low birth weight (less than 2,500 grams). Our primary estimation sample contains 10,785 first-born children.

In addition, we use the other registers under the same selection rules to compile a secondary estimation sample of first-born children raised in representative families. This sample is based on the population sample of all families and representative of all first-born children who may or may not have second-born siblings. We use the secondary sample of first-born children raised in representative families to assess the wider generalizability of our results obtained with the primary sample of first-born children raised in families seeking IVF treatment for a second child. Our secondary estimation sample contains 332,883 first-born children.

Table 1 provides sample means for the school outcomes of first-born children (which we study below) and for the control variables (which we use to compare first-born children in IVF-treated and representative families). We make four informative comparisons regarding pre-treatment child characteristics, parent characteristics, the sibling treatment, and the post-treatment school outcomes.

For the first-born child characteristics across the different families, we see that child gender and birth weight, both being strong predictors for cognitive test scores and personality traits (Black et al. 2007, Mueller and Plug 2006), are nearly identical. The likelihood that the first-born child is an IVF child is very different in the different families. This is not so surprising, though. The high IVF rates in families seeking treatment for a second child just indicate that the fertility problems families encounter at their first pregnancy are likely to recur at their second pregnancy. The low IVF rates in representative families, on the other hand, indicate that most families are unlikely to experience fertility problems requiring IVF treatment. Of all the first-born children in the representative sample (in our sampling window), only 2 percent are conceived

through IVF.

For the parent characteristics across the different families, we see that parents who rely on IVF for a second child are, on average, more educated, more employed, richer, and older than those parents who do not rely on IVF. These differences are similar to those found in earlier studies (Bitler and Schmidt 2012, Lundborg et al. 2017 2024, Bensnes et al. 2023). For the sibling outcomes, which is the treatment we evaluate in this study, we see that first-born children in IVF treated families are less likely to have siblings, and those with siblings have their siblings later. Both patterns are consistent with IVF families experiencing fertility problems.

And for the school outcomes of first-born children, we see that first-born children raised in families that seek IVF treatment for a second child do significantly better; that is, they perform better on math and reading tests, they are more agreeable, more conscientious, more emotionally stable, and not unimportantly they also express to be happier with their education. As we identify only child effects in samples of first-born children in families that seek IVF treatment for a second child, these differences suggest that some caution is warranted when generalizing the only child effects we estimate in this study to other first-born children.

Is IVF treatment success conditionally random?

Our empirical strategy to identify only child effects strongly relies on IVF treatment success in first and later treatments being the outcome of a random process. As we mentioned earlier, this condition may not hold unconditionally. IVF treatment success depends on the year of treatment (i.e., more recent treatments are more successful because of medical innovations) and the parents' age at treatment (i.e., treatments given to older couples are less successful because of the age gradient in fertility problems). If these medical indicators (year and age at treatment) are the primary drivers behind treatment success, as the medical studies we referred to earlier emphasize, we should see that treatment success becomes as good as random conditional on year and age at treatment.

Table 2 contains the balancing test results for each IVF attempt (up to the sixth attempt). In particular, we show 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 couples with a successful attempt (in columns 1, 4, and 7), the un-

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)	0.366
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)	0.789
Child birth length (cm)	52.05 (2.25)	52.00 (2.22)	0.045
Mother age at birth	30.42 (4.47)	27.68 (4.54)	0.000
Mother college (0/1)	0.53 (0.50)	0.45 (0.50)	0.000
Mother income (M)	0.23 (0.14)	0.18 (0.13)	0.000
Mother work (0/1)	0.93 (0.26)	0.90 (0.31)	0.000
Father age at birth	32.84 (5.40)	30.34 (5.51)	0.000
Father college (0/1)	0.37 (0.48)	0.30 (0.46)	0.000
Father income (M)	0.30 (0.21)	0.25 (0.17)	0.000
Father work (0/1)	0.91 (0.28)	0.91 (0.29)	0.017
Observations	10,785	332,883	
Characteristics after first child birth			
Siblings (0/1)	0.75 (0.43)	0.84 (0.37)	0.000
Birth spacing (if siblings)	4.91 (2.87)	3.41 (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 (0.81)	0.000
Consciousness	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 whether the means in the first two columns are significantly different from each other (low values indicate significant differences). 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).

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

	first attempt			**second attempt**			**third attempt**		
	succ. (1)	fail. diff. (2)	diff. ipw (3)	succ. (4)	fail. diff. (5)	diff. ipw (6)	succ. (7)	fail. diff. (8)	diff. ipw (9)
Female child	0.49	0.01	0.01	0.47	-0.01	-0.01	0.47	-0.01	0.00
IVF child	0.50	0.03	0.01	0.52	0.05	0.04	0.50	0.02	0.01
Child weight	3.52	0.01	0.01	3.51	0.01	0.02	3.50	0.01	0.01
Child length	52.04	-0.01	0.02	52.07	0.04	0.09	52.12	0.13	0.07
Mother age	33.40	-1.11	-0.03	33.42	-1.22	-0.03	33.61	-1.20	0.01
Mother college	0.52	-0.00	0.02	0.55	-0.00	0.01	0.57	-0.01	0.00
Mother income	0.26	-0.01	0.01	0.26	-0.01	-0.01	0.27	-0.01	0.00
Mother work	0.91	-0.01	-0.00	0.92	-0.00	-0.01	0.95	0.02	0.02
Father age	35.85	-1.06	-0.00	35.88	-1.12	0.00	35.95	-1.23	0.23
Father college	0.37	-0.01	0.01	0.38	-0.01	0.01	0.42	0.01	0.04
Father income	0.36	-0.01	0.01	0.38	0.01	0.01	0.38	0.00	0.00
Father work	0.90	0.01	0.00	0.91	0.01	0.00	0.92	0.02	-0.00
Joint p-value		0.00	0.18		0.00	0.11		0.00	0.44
Observations		10,785			6,086			3,462	

	fourth attempt			**fifth attempt**			**sixth attempt**		
	succ. (1)	fail. diff. (2)	diff. ipw (3)	succ. (4)	fail. diff. (5)	diff. ipw (6)	succ. (7)	fail. diff. (8)	diff. ipw (9)
Female child	0.49	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	0.02	0.54	0.03	0.03
Child weight	3.50	0.03	0.03	3.48	0.01	-0.02	3.49	0.03	0.05
Child length	51.91	-0.09	-0.04	51.88	-0.06	-0.06	52.01	0.17	0.18
Mother age	34.03	-0.69	0.06	33.34	-1.43	0.02	33.56	-1.23	-0.12
Mother college	0.64	0.04	0.05	0.65	0.04	0.07	0.60	-0.02	-0.04
Mother income	0.29	-0.01	-0.00	0.29	-0.01	-0.01	0.30	0.00	0.01
Mother work	0.95	0.02	0.00	0.95	0.03	0.02	0.93	0.01	0.02
Father age	36.53	-0.52	0.09	35.68	-1.40	-0.26	35.62	-1.36	-0.19
Father college	0.43	-0.00	0.00	0.46	0.01	0.05	0.56	0.11	0.10
Father income	0.38	-0.01	-0.01	0.43	0.03	0.03	0.45	0.07	0.07
Father work	0.89	-0.01	-0.00	0.94	0.06	0.05	0.94	0.07	0.07
Joint p-value		0.01	0.32		0.00	0.20		0.00	0.20
Observations		1,855			1,112			669	

Note–The columns (1), (4), and (7) present the mean characteristics for children whose parents experienced a successful treatment. The columns (2), (5), and (8) present the difference in mean characteristics between families with successful and failed treatment. The columns (3), (6), and (9) present the inverse probability weighted differences. The inverse probability weighted differences are based on father age and age squared, mother age and age squared, and a full set of year of treatment indicators measured at each consecutive treatment. Observations refer to the total number of firstborn children whose parents underwent the respective procedure for a second child.

conditional mean differences of baseline characteristics between couples with a successful and failed attempt (in columns 2, 5, and 8), and the conditional differences in samples weighted by the inverse probability of attempt success. The weights are extracted from logistic regressions of having a successful attempt on the mother age and age squared at treatment, father age and age squared at treatment, and a full set of year of treatment indicators.[12] At the bottom of the difference-in-means columns, we also present p-values for joint F-tests for whether the child and parent characteristics are jointly statistically significant predictors of treatment success. Random treatment success within the medically-driven matching strata based on year and age of treatment would predict that the differences in mean characteristics are close to zero and that the characteristics together are jointly statistically insignificant.

For the first attempt, the unconditional means and differences (in columns 1 and 2) indicate that mostly younger parents and parents who conceived their first child with IVF experience significantly higher success rates. We see little differences on the other characteristics, including well-known predictors of child school performance (such as child birth weight and parent education). Nonetheless, all the child and parent characteristics together are statistically significantly related to treatment success, which suggest that success at the first attempt is not the outcome of a random process. When we use the inverse probability weights to balance the sample on the mothers' age, fathers' age, and treatment year (in column 3), we no longer see a clear relationship between treatment success and the characteristics of children and parents. The inverse probability weighted 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 mothers (p-value is 0.18). The random assignment of treatment success should have produced very similar results, on average.

For later attempts after accounting for parental age and treatment year (as displayed in columns 3, 6, and 9), there are no substantial differences in the pretreatment characteristics of children or parents between those who succeed and those who do not. While we see that mostly higher income couples are more inclined to continue treatment after (a sequence of) treatment failures,

[12] We present these test results using the largest sample with reading test scores as outcome. 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 nationwide well-being surveys.

treatment success itself does not seem to be selective at each of the treatments itself. At each treatment, the p-value for all the child and parent characteristics together are (mostly much) larger than 0.10. 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 matching are as one would expect with conditional random treatment success at the first and later attempts.

4 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 has been successful 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.

Identification for complier children

Using the local average treatment effects (LATE) framework (Imbens and Angrist 1994), we classify the children in our sample as either compliers or non-compliers (or always takers). The compliers ($C=1$) are children who never end up having siblings after a first failed attempt. The non-compliers ($C=0$) are children who always end up having siblings, regardless of a first failed attempt. There are no never takers and defiers because, in our IV-IVF setup, the first-born children whose parents had a successful first attempt always end up having siblings. We next define the child's potential school outcomes $Y_z(s)$ indicating what the child's school outcome would be in case the child had any siblings or not ($s=0,1$) and in case the child had parents whose first IVF attempt succeeded or failed ($z=0,1$). 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, we can 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) $E[Y_1(1)|C=0] = E[Y_0(1)|C=0]$.

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 similar potential outcome averages for non-compliers with first-attempt and later-born siblings, that is, the child's outcome response to siblings does not depend on when siblings arrive, or on how many IVF attempts it takes for the IVF-treated parents to achieve success. This assumption is the more substantive one.

To identify the causal effect, we start with the reduced form and compare the school outcomes of children whose parents experience a first-time IVF success of failure. The reduced form under independence is given by

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

It 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 non-compliers, which is in the context of our study the bias term.

Under exclusion, the potential outcome averages are similar for non-compliers who get siblings at first and later attempts, which means that the second term in (1) is zero. Under independence, the complier share is similar among parents whose first attempt either failed or succeeded, which means that the complier share is identified from the share of first-born children without siblings in the sample of parents whose first attempt failed ($\Pr[C=1] = \Pr[S=0|Z_1=0]$). Taken together, we identify the effect of having siblings for compliers from:

$$\beta_C = E[Y_1(1) - Y_0(0)|C=1] = \frac{\theta_C}{\Pr[S=0|Z_1=0]}.$$

Identification for relier children

The IV-IVF identification strategy has two limitations. First, the corresponding estimates capture causal effects only if the exclusion condition holds, which may fail if first-born children's outcomes depend not only on whether but also on when siblings arrive. Second, the estimates capture causal effects only for complier children, a subset of first-borns who never have siblings if the first treatment fails.

To address these limitations, we extend our focus to later IVF attempts. By exploiting randomness not only at the first attempt but also at later attempts (after a sequence of failed attempts), we can identify average outcomes for all children in a scenario where they never receive an IVF sibling. This approach mitigates exclusion violations caused by first-born children with siblings from later treatments and allows us to identify the causal effect of having siblings for a larger group of first-born children (including compliers and always-takers who get their siblings at later attempts but would not get any siblings if all attempts failed).

To formally present the improved IV-IVF approach, we expand the LATE framework. Following Ilciukas (2024), we reclassify subpopulations of first-born children. We label children 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. As such, 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. 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.

We also classify children by the number of IVF attempts their parents are "willing" to undergo for the second child. W is an integer and represents the total number of IVF attempts parents would undergo for a second child in a scenario where all attempts fail. 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 undergo for the second child. By definition, parents undergo IVF until an attempt succeeds ($Z_w = 1, Q = w$) or until they reach the maximum number of attempts they are willing to undergo and quit ($Z_Q = 0, Q = W$). The child types R and W are observed for some but not all first-born

children.

Sequential Independence Condition

We next discuss the conditions under which we 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.

The sequential independence condition works as follows. Consider first-born children whose parents completed exactly w failed IVF attempts. We know three 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, it is for any type-w child in our sample effectively random to end up with parents who never experience a successful treatment. And third, we observe their school outcomes. Together, this means that their average outcomes represent the average outcomes for all type-w children in a scenario where all treatments fail. For reliers, these are the outcomes for those who never get siblings. For non-reliers, these are the outcomes for those who get siblings though other means. We can therefore write the average school outcomes for first-born children whose parents completed exactly w failed IVF attempts as follows:

$$\begin{aligned} E[Y|Q=w, Z_w=0] = {} & E[Y_0(0)|W=w, R=1]\Pr[R=1|W=w] \\ & + E[Y_0(1)|W=w, R=0]\Pr[R=0|W=w]. \end{aligned} \tag{2}$$

Following a similar argument, we can identify the shares of different type-w children using the shares of children whose parents either continue or stop treatment

after a sequence of previous failed attempts.[13] For a formal proof, we refer to Ilciukas (2024).

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

$$\sum_{j=1}^{\overline{Q}} E[Y|Q=w, Z_w=0]\Pr[W=j] = \\ E[Y_0(0)|R=1]\Pr[R=1] + E[Y_0(1)|R=0]\Pr[R=0],$$

where $\overline{Q}$ represents the highest number of IVF attempts observed in our sample. When we subtract this average outcome from the average outcome of children whose parents experience a first-time IVF success ($E[Y|Z_1=1]$), we get the new reduced form for reliers:

$$\theta_R = E[Y_1(1) - Y_0(0)|R=1]\Pr[R=1] \\ + E[Y_1(1) - Y_0(1)|R=0]\Pr[R=0]. \tag{3}$$

The first term captures the effect of whether siblings arrive for reliers. The second term 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. This implies that, compared to the baseline reduced form for compliers in (1), greater weight is given to the effect of having siblings, simply because the relier share is (mechanically) larger than the complier share.

Exclusion and Ignorable Sibling Response Conditions

With the modified reduced-form effect for reliers in (3), we are able to identify the average effect of having siblings for either reliers or the full sample of first-born children using conditions that mirror those in the standard IV framework but are now modified to accommodate non-reliers rather than always-takers:

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

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

Exclusion for non-reliers (**A4**) asserts that the impact of having siblings later for those children with non-IVF siblings is ignorable (after accounting for parental

[13] 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.

age and treatment year effects). While this assumption remains substantive, the exclusion condition under (**A4**) is arguably weaker than the one under (**A2**), as it involves far fewer first-born children. Under exclusion for non-reliers, we can set the bias term in (3) to zero and identify the effect of having siblings for reliers:[14]

$$\beta_R = E[Y_1(1) - Y_0(0)|R=1] = \frac{\theta_R}{\sum_{j=1}^{\overline{Q}} \Pr[S=0|Q=j, Z_j=0] \Pr[W=j]}.$$

The ignorable sibling response condition (**A5**), on the other hand, states that the impact of having siblings is ignorable for children who get siblings through means other than IVF. These children had, on average, their siblings much later than others in our sample. If the time spent as an only child is the primary reason for differences in their school outcomes, the potential school outcome $Y_0(1)$ for these older first-born children with young siblings likely captures a large fraction of their lives without siblings ($Y_0(0)$). Under the ignorable sibling response condition, we effectively treat $E[Y_0(1)|R=0]$ in (3) 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:

$$\beta = E[Y_1(1) - Y_0(0)] = \theta_R.$$

We close with three remarks about the identifying conditions (**A4**) and (**A5**). First, these conditions may not hold universally. However, we suspect that the reduced-form estimate in (3) is exposed to less bias than the one in (1). With a much smaller share of always takers responsible for the exclusion bias, we effectively estimate a stronger first-stage effect and, as a result, obtain a more precisely estimated second-stage effect with less bias. Second, these conditions do not rule out other potential exclusion violations, such as depression and divorce following unsuccessful treatments. However, with little evidence that these violations matter for childless families seeking treatment for a first child (Lundborg et al. 2017, 2024; Bensnes et al. 2023; Bögl et al. 2024; Ilciukas 2024), we suspect that these violations will matter less for one-child families seeking treatment for a second child. Third, these conditions may yield different estimates even without any condition violation. Under the different conditions, we identify the effect of having siblings for different groups of first-born children;

[14] We identify, under sequential independence, the share of reliers among children whose parents are willing to undergo exactly w attempts from the share of parents having only one child after undergoing exactly w attempts in the observed scenario that all attempts fail. We recover the overall relier share by taking the weighted average over all type-w children.

that is, we estimate average effects for compliers (under independence and non-complier exclusion), for reliers (under sequential independence and non-relier exclusion), and for all first-born children (under sequential independence and ignorable sibling response).

Estimation

Based on different combinations of assumptions, we use three different effect estimators of sibling exposure on the school outcomes of first-born children. Table 3 displays the different (sequential) inverse probability weighted estimators (in column 1), under which assumptions the estimators identify causal effects (in column 2), for which type of first-born children the estimator identifies an average effect (in column 3), and the estimation sample used to construct the estimators (in column 4). For details, we refer to Ilciukas (2024).

The different estimators in Table 3 rely on the estimated weights $\widehat{\pi}_j(X_j)$. These weights are extracted from logistic regressions of Z_j on X_j, run on samples of couples at their jth IVF attempt (after having experienced j–1 failures), where Z_j is the indicator for having a successful treatment at the jth attempt, and where X_j is a small set of controls including the mother age and age squared at attempt j, father age and age squared at attempt j, and a full set of year of attempt j indicators.

While our estimators rely on weights extracted from success rates from first to last attempts, one practical problem is that there are fewer families at later attempts. Of the 10,785 treated families, for instance, there are only 85 families left at the 10th attempt (after 9 consecutive failures). In our sample, about 94 percent of all families undergo no more than 5 attempts. To avoid the problem of small sample weights at later attempts, we cap the maximum number of attempts at 5 and treat those first-born children who end up with siblings from the 6th or later IVF attempts as if they are non-reliers. In estimation, we have experimented with different caps. Later, in Table 6, we show that our results are not sensitive to these choices.

Table 3
Three estimators under different assumptions, for different populations, with different samples

estimator	assumptions	population	sample
$\beta_C = \left\{ E\left[\frac{YZ_1}{\widehat{\pi}_1(X_1)}\right] - E\left[\frac{Y(1-Z_1)}{1-\widehat{\pi}_1(X_1)}\right]\right\} \Big/ \left\{ E\left[\frac{SZ_1}{\widehat{\pi}_1(X_1)}\right] - E\left[\frac{S(1-Z_1)}{1-\widehat{\pi}_1(X_1)}\right]\right\}$	independence (**A1**), exclusion for compliers (**A2**)	compliers ($C=1$)	all first-born children
$\beta_R = \left\{ E\left[\frac{YZ_1}{\widehat{\pi}_1(X_1)}\right] - E\left[\frac{Y(1-Z_Q)}{\prod_{j=1}^{Q}(1-\widehat{\pi}_j(X_j))}\right]\right\} \Big/ \left\{ E\left[\frac{SZ_1}{\widehat{\pi}_1(X_1)}\right] - E\left[\frac{S(1-Z_Q)}{\prod_{j=1}^{Q}(1-\widehat{\pi}_j(X_j))}\right]\right\}$	sequential independence (**A3**), exclusion for reliers (**A4**)	reliers ($R=1$)	first-born children excluding those with IVF-induced siblings from later treatments
$\beta = E\left[\frac{YZ_1}{\widehat{\pi}_1(X_z)}\right] - E\left[\frac{Y(1-Z_Q)}{\prod_{j=1}^{Q}(1-\widehat{\pi}_j(X_j))}\right]$	sequential independence (**A3**), ignorable sibling response (**A5**)	all first-born children ($R=1$, $R=0$)	first-born children excluding those with IVF-induced siblings from later treatments

Note–For the construction of the three different estimators, we need three other variables: Z_Q is the success indicators for the last IVF treatment families undergo for a second child, $\widehat{\pi}_j(X_j)$ is the predicted success probability at attempt j taken from logistic regressions of having a successful attempt j on the mother age and age squared at treatment j, father age and age squared at treatment j, and a full set of year of treatment j indicators in sample undergoing attempt j.For the first estimator, we classify children as compliers and non-compliers (or non-compliant always takers). The compliers ($C=1$) are children who never end up having siblings after a first failed attempt. The non-compliers ($C=0$) are children who always end up having siblings, regardless of a first failed attempt. For the two other estimators, we classify children as reliers or non-reliers (or non-reliant always takers). The reliers ($R=1$) are children whose parents rely on either first or later IVF treatments to get a second child, consisting of compliers and always takers who would get siblings from later IVF attempts, regardless of a first failed attempt but not if all attempts failed. The non-reliers ($R=0$) are the first-born children who would get siblings through other means, regardless of IVF success.

5 Results

Only child associations in representative and IVF families

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 4 contains the unconditional differences between the school outcomes of first-born children with and without siblings in the two samples. The school outcomes we consider 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 happiness (column 6). We say that children do better in school if they perform better on test scores, if they are more agreeable, more conscientious, and more emotionally stable, and if they report to be happier in school. We present results for first-born children in the representative sample (panel A) and in the IVF sample (panel B).

When we examine the estimated differences between first-born children with and without siblings, we see two clear patterns. First, children with siblings do substantially better than children without siblings on all the school outcomes we measure. Second, the estimates for first-born children raised in representative families are about twice as large as those for first-born children in IVF families.

We recognize that these patterns are merely suggestive. The two patterns are consistent with a causal story in case children benefit from socializing with their younger siblings. The same two patterns are also consistent with a selection story in case a divorce in one-child families leads to poorer child outcomes and fewer siblings, or parents with a poor quality first child (in the representative sample) more often decide against a second child. Obviously, these simple comparisons between children with and without siblings are not enough to establish a causal link running from having siblings to child outcomes.

Only child effects in IVF families

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

Table 5 (panel A) presents the second-stage effect estimates using treatment

Table 4
Associations between having siblings and outcomes

	math test	**reading test**	**agreeable**	**consc.**	**emotional stability**	**school happiness**
Panel A: Representative sample						
having siblings	0.174	0.094	0.074	0.122	0.114	0.115
	0.005	*0.005*	*0.005*	*0.005*	*0.005*	*0.005*
Obs. (1000s)	252.3	332.9	220.2	220.2	220.2	220.2
Panel A: IVF sample						
having siblings	0.086	0.052	0.029	0.051	0.043	0.077
	0.023	*0.020*	*0.021*	*0.022*	*0.022*	*0.021*
Observations	8,216	10,785	7,604	7,582	7,482	7,697

Note–We report unconditional outcome differences between first-born children with and without siblings. The outcomes we consider 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). Robust standard errors are shown in italics.

success at the first attempt as the instrumental variable. If the standard independence and exclusion conditions hold, treatment success at the first attempt is a valid instrument, and the corresponding estimates represent the causal effect of having siblings in the sample of complier children. When we compare the estimates in panel A to the positive ones reported in Table 4, we find that all the sibling effect estimates change sign and turn statistically insignificant. If we accept treatment success at the first attempt as a valid instrument, it seems that, for all the school dimensions we measure, first-born children do worse, but not significantly worse, in the presence of siblings.

We note that these insignificant sibling effect estimates are not the result of a weak first-stage relationship between IVF success at first treatment and having a sibling. Table 5A (panel A), which presents the first-stage estimates for the second-stage estimates in Table 5 (panel A), shows that treatment success has a strong and statistically significant impact on the likelihood to have a second child (with corresponding F statistics that are way larger than the typical rules-of-thumb values for strong instruments). We find that first-born children, whose parents had a successful first attempt, are about 32-34 percentage points more likely to have a sibling. If treatment success at the first IVF attempt is conditionally random, these first-stage estimates indicate that

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

	math test	**reading test**	**agreeable**	**consc.**	**emotional stability**	**school happiness**
Panel A: The average impact for compliers (under A1 & A2)						
having siblings	-0.055	-0.045	-0.033	-0.022	-0.073	-0.098
	0.072	*0.059*	*0.061*	*0.070*	*0.068*	*0.064*
Observations	8,216	10,785	7,604	7,582	7,482	7,697
Panel B: The average impact for reliers (under A3 & A4)						
having siblings	-0.029	-0.012	-0.005	-0.004	-0.011	-0.031
	0.049	*0.039*	*0.043*	*0.045*	*0.045*	*0.045*
Observations	5,892	7,691	5,310	5,282	5,215	5,377
Panel C: The average impact for all children in the sample (under A3 & A5)						
having siblings	-0.016	-0.007	-0.003	-0.002	-0.006	-0.016
	0.027	*0.021*	*0.023*	*0.024*	*0.024*	*0.024*
Observations	5,892	7,691	5,310	5,282	5,215	5,377

Note–The estimates represent the average causal effect of having siblings for first-born children derived under different sets of assumptions (explained in the main text). The effects are estimated by inverse propensity weighting. The weights are taken from logit regression models predicting the probabilities of having a successful first treatment (panel A) and of having a successful treatment at each treatment (panels B and C). The control variables in the logit regression models are a full set of dummies for year of treatment and for the parents' age and age squared at treatment (in panels A, B, and C). The outcomes we consider 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). Robust standard errors are shown in italics.

Table 5A
The impact of successful 1st attempt on having siblings (first stage results)

	math test	reading test	agreeable	consc.	emotional stability	school happiness
Panel A: The effect of first successful attempt on having siblings						
having siblings	0.344	0.337	0.322	0.322	0.322	0.323
	0.008	*0.005*	*0.007*	*0.006*	*0.007*	*0.006*
Observations	6,185	10,785	7,589	7,583	7,471	7,702
Panel B: The effect of first attempt succeeding, relative to all failing, on having siblings						
having siblings	0.551	0.544	0.535	0.536	0.535	0.536
	0.012	*0.009*	*0.010*	*0.010*	*0.011*	*0.010*
Observations	4,457	7,691	5,300	5,287	5,223	5,379

Note–The first-stage estimates represent the average causal effect of a successful first attempt on the likelihood having siblings. The first-stage effects reflect the complier shares (panel A) and relier shares (panel B). The first-stage effects are estimated on different sample depending on the main outcomes we use: 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 happiness (column 6). Robust standard errors are shown in italics; * indicates significance at 10 percent level, ** indicates significance at 5 percent level, and *** at 1 percent level.

32-34 percent of all the first-born children are complier children. The other children are the non-complier (or always-taker) children. The corresponding non-complier-to-complier ratios, which are close to 2, suggest that the confounding role of possible exclusion violations can be sizable.

Of the standard instrument validity conditions, we consider the exclusion condition the most substantive one. More specifically, we are concerned that treatment failure may influence the outcomes of first-born children through its influence on sibling spacing. If we expand the independence condition to hold for all subsequent IVF attempts, which seems reasonable to us, we can tackle this issue and replace the more restrictive exclusion condition for always takers (including those who get their siblings either at later IVF attempts or through other means) with weaker potential outcome conditions using fewer always takers (excluding those who get their siblings at later IVF attempts but keeping those who get siblings later through other means).

Table 5 (panel B) presents the modified second-stage effect estimates under the sequential independence and non-relier exclusion conditions. If these conditions hold, we can remove first-born children who get siblings at later IVF

attempts from the sample, continue to treat success at the first attempt as a valid instrument, and interpret the modified estimates as the causal effect of having siblings for both complier children and always-taker children who would get siblings at later attempts. We find that the modified second-stage estimates are considerably smaller than the ones in panel A. In fact, all the effect estimates are now much closer to zero and statistically insignificant, suggesting that the relationship between the school outcomes of first-born children and having any siblings is at best weak.

When we compare the modified first-stage estimates in panel B (in Table 5A) to those in panel A, we find that the first-stage estimates are, as expected, much larger as we removed a large fraction of always takers. In particular, the modified estimates suggest that 53 to 55 percent of all children are relier children (under the sequential independence condition). The larger first-stage estimates also imply a weaker influence of any possible exclusion violation. Its influence, as measured by the modified non-relier-to-relier ratios, are now close to 0.8 in our sample.

Table 5 (panel C) presents the effect estimates under the sequential independence and ignorable sibling response conditions. If these conditions hold, the corresponding effect estimates are actually the reduced-form estimates for the modified second-stage estimates in the previous panel, and represent the average causal effect of having siblings for all the first-born children in our sample. We find that all the effect estimates are near zero and statistically insignificant, again suggesting that there is no relationship between the school outcomes of first-born children and having any siblings.

It is worth noting that, once we assume sequential independence, our findings appear quite insensitive to the way we treat the counterfactual outcomes of first-born children who get their siblings later through other means than IVF. Whether we think of them as first-born children who never get siblings (because they get siblings later and spent a longer part of their young lives as only child) or as first-born children who get siblings at the first IVF attempt (because they get siblings too but later), we always find that the arrival of siblings has a negligible impact on their school outcomes. The latter is also consistent with a negligible impact of possible exclusion violations (otherwise we would have found different effect estimates).

If we were asked to choose among the different estimates, we would pick those shown in panels B and C. One reason for preferring these estimates is that the other estimates (in panel A) are derived under a more substantive

condition. And the other reason is that these estimates tell the same story and, as such, do not seem to depend much on the more reasonable, yet different, conditions. Based on these preferred estimates, which are all close to zero and precise enough that the 95 percent confidence intervals exclude sibling effect estimates more than 0.1 standard deviations away from zero (or 0.05 standard deviations away from zero if we consider the estimates in panel C), we can draw one immediate conclusion: the cognitive and non-cognitive development of school-aged first-born children neither benefits nor suffers much from having siblings.

Robustness

To probe the robustness of the sibling effect estimates, we consider a number of alternative specifications and alternative samples. Table 6 contains alternative sibling effect estimates when we apply a narrow and wide set of matching controls (in panels A and B), and when we impose a more and less stringent maximum number of IVF attempts (in panels C and D). We focus on the modified second-stage effect estimates obtained under the sequential independence and ignorable sibling response conditions (using the sibling effect estimates in Table 5 panel C as point of reference).

We first present results for the narrowest and widest set of matching controls possible. Recall that we have constructed our sibling effect estimates using inverse probability weights taken from treatment success logistic regressions with controls based on both parents' age at treatment and treatment year. In panel A we estimate the sibling estimates without using any controls; that is, we use inverse probability weights based on unconditional treatment success probabilities. When we compare the estimates to the ones we present in Table 5 (panel C), we find that sibling effects are somewhat larger in size, but not large enough to be statistically different from zero. The estimate for reading appears an exception, showing a statistically significant test score loss of 0.076 standard deviation due to having siblings.

In panel B we construct the sibling estimates using a much wider set of controls; that is, we estimate the inverse probability weights from treatment success logistic regressions and augment the set of medically-driven controls with age, gender, birth weight, birth length and IVF status of the first-born child, and college education indicators, pretreatment employment indicators and pretreatment annual incomes of both parents. Similar to no-control estimates,

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

	math test	**reading test**	**agreeable**	**consc.**	**emotional stability**	**school happiness**
Panel A: no controls						
having siblings	-0.067	-0.076	-0.006	-0.038	-0.002	-0.026
	0.043	*0.035*	*0.044*	*0.041*	*0.042*	*0.041*
Observations	5,908	7,708	5,334	5,307	5,246	5,402
Panel B: baseline controls and non-treatment controls						
having siblings	-0.072	-0.049	-0.016	-0.016	-0.011	-0.034
	0.044	*0.037*	*0.046*	*0.043*	*0.044*	*0.047*
Observations	5,892	7,691	5,310	5,282	5,215	5,377
Panel C: baseline controls, 4 attempts						
having siblings	-0.029	-0.005	-0.015	-0.009	-0.012	-0.029
	0.047	*0.042*	*0.043*	*0.044*	*0.054*	*0.049*
Observations	6,075	7,935	5,499	5,470	5,401	5,566
Panel D: baseline controls, 6 attempts						
having siblings	-0.001	0.002	0.013	0.004	0.019	0.000
	0.045	*0.037*	*0.044*	*0.046*	*0.042*	*0.042*
Observations	5,798	7,550	5,217	5,190	5,124	5,283

Note–Results are from our baseline specification (under A3 & A5). We use inverse probability weights based on the matching controls parents' age at treatment and treatment year. We set 5 as the maximum number of IVF attempts. In panels A and B we use no-control and full-control inverse probability weights. The full-control specification augments the matching controls with age, gender, birth weight, birth length and IVF status of the first-born child, and college education indicators, pretreatment employment indicators and pretreatment annual incomes of both parents. In panels C and D, we cap the maximum number of IVF attempts at 4 and 6 attempts. Standard errors are shown in italics.

we find that the full-control estimates are somewhat larger but less precise. It is important to emphasize that adding more controls does not always lead to more credible estimates. If the additional controls are uninformative about treatment success, they only make the estimates more noisy. When we compare standard errors, we find that all are larger than those reported in Table 5 (panel C).

We next present sibling effect estimates imposing different maximum numbers of IVF attempts. Our baseline estimates are constructed under a maximum of 5 attempts, which covers about 94 percent of all families. If we would work with a maximum of 4 or 6, we would cover about 90 or 97 of all families, respectively. We find that our near zero sibling effect estimates hardly change when we set the maximum number of IVF attempts to either 4 (in panel C) or 6 (in panel D).

Heterogeneity

We have so far focused on the average effect of having siblings on first-born children's school outcomes. In our heterogeneity analysis, we examine whether the effects of having siblings are different for different groups of first-born children.

We first restrict the sample to boys, girls, less educated mothers, and low-income couples. These margins may matter in theory if boys and girls mature and learn differently and, as a result, respond differently to siblings; if less educated mothers struggle more with planning, misallocate time and money, and overinvest in their first-born children (or underinvest in their second-born children); or if financial constraints hinder the development of children and lead to stronger sibling effects in more financially constrained families.

We also consider older children (3 and older at first treatment) and older mothers (35 and older at first treatment). These groups may tackle some of the internal validity concerns we mentioned earlier. Older children, for instance, spend more time as the only child and, as a result, form a more appropriate group to identify the true effect of having siblings under the ignorable sibling response condition. To get an idea about the age at which these children get their siblings, we look at two groups: successfully first-time treated parents and never-successfully treated parents who get their second child through other means. In the sample of first-born children of 3 and older at first treatment, we find that the average age when siblings arrive is 6.4 and 8.4 years old, respectively. On the other hand, older mothers likely show a stronger first stage relationship and, as a result, reduce the confounding influence of exclusion vio-

lations. In the sample restricted to mothers of 35 and older at first treatment, we find relier shares of 0.67-0.69 that are substantially larger than the relier shares of 0.53-0.55 observed in the full sample (see Table 5A, panel B).

And lastly, we restrict the sample to naturally conceived first-born children, healthier women whose fertility problem is on their partner side, and second-born singletons. These margins, shared with first-born children raised in representative families, may ensure that the sibling effects we identify in IVF families have a wider generalizability.

Table 7 contains the alternative sibling effect estimates when we look at boys (in panel A) girls (in panel B), less educated mothers (in panel C), low-income couples (in panel D), older children (in panel E), older mothers (in panels F), naturally conceived children (in panel G), infertile fathers (in panel H), and second-born singletons (in panel I). Regardless of how we restrict the sample, we find that almost all sibling effect estimates remain remarkably close to zero and statistically indistinguishable from one another. Some effect estimates seem large (such as the negative effect on math test scores for children of less-educated mothers, the positive effects on conscientiousness and happiness scores for children in low-income families, and the negative effect on agreeableness scores for naturally conceived children), but they are not very precise and lack a clear and consistent pattern.

Overall, these findings reaffirm our earlier conclusion; that is, the consistency of the near-zero sibling effects across the different groups of first-born children for all the school outcomes we measure suggests again that having siblings has little impact on the cognitive and noncognitive school outcomes of first-born children. Moreover, they also speak to the external validity of our conclusion; that is, the first-born children who resemble those in representative families (in being naturally conceived, raised by healthy mothers, and having no twin siblings) do not benefit nor suffer from siblings either.

6 Conclusions

Our study revolves around only children, an increasingly important group of children that is often neglected in the economics literature examining the relationship between sibling composition and child outcomes. In particular, 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

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

	math test	reading test	agreeable	consc.	emotional stability	school happiness
Panel A: first-born sons						
having siblings	-0.005	-0.043	-0.027	-0.012	0.007	-0.026
	0.061	*0.051*	*0.065*	*0.061*	*0.070*	*0.062*
Observations	2,982	3,934	2,688	2,670	2,654	2,736
Panel B: first-born daughters						
having siblings	-0.054	0.037	0.012	0.011	0.007	-0.007
	0.058	*0.057*	*0.058*	*0.069*	*0.070*	*0.060*
Observations	2,862	3,727	2,606	2,596	2,549	2,624
Panel C: less educated mothers						
having siblings	0.025	0.026	0.051	0.069	0.056	0.053
	0.056	*0.052*	*0.052*	*0.067*	*0.063*	*0.060*
Observations	2,910	3,766	2,519	2,511	2,475	2,556
Panel D: low-income couples (bottom 50% joint income)						
having siblings	-0.070	-0.018	-0.030	-0.098	-0.104	-0.105
	0.062	*0.056*	*0.060*	*0.059*	*0.065*	*0.059*
Observations	2,824	3,713	2,777	2,764	2,730	2,800
Panel E: older first-born child (3+)						
having siblings	0.040	0.029	-0.004	0.075	0.021	0.040
	0.055	*0.050*	*0.060*	*0.065*	*0.054*	*0.060*
Observations	2,990	3,886	2,525	2,511	2,466	2,565
Panel F: older mother (35+)						
having siblings	-0.021	0.007	-0.021	-0.031	-0.070	-0.017
	0.051	*0.041*	*0.048*	*0.055*	*0.055*	*0.047*
Observations	2,798	3,753	2,478	2,469	2,434	2,508
Panel G: first child is non-IVF						
having siblings	0.003	0.026	-0.079	-0.005	0.028	-0.028
	0.057	*0.054*	*0.057*	*0.070*	*0.063*	*0.065*
Observations	3,214	4,117	2,790	2,778	2,732	2,831
Panel H: fathers infertile						
having siblings	0.082	-0.023	0.020	0.026	0.052	0.047
	0.073	*0.066*	*0.066*	*0.066*	*0.072*	*0.067*
Observations	2,364	3,067	2,151	2,144	2,120	2,181
Panel I: no twins						
having siblings	-0.019	-0.004	-0.027	0.007	-0.020	-0.020
	0.043	*0.043*	*0.043*	*0.047*	*0.053*	*0.045*
Observations	5,449	7,129	4,933	4,908	4,840	4,996

Note–Results from baseline specification (under A3 & A5). Robust standard errors are shown in italics.

IVF treatments in an administrative dataset of all Danish one-child families seeking IVF for a second child, we can identify the causal effect of being an only child on a range of cognitive and non-cognitive outcomes of first-born children. This is particularly relevant today; with collapsing fertility rates worldwide, the number of only children is at an all-time high, making it crucial to better understand the long-term consequences of this demographic shift.

When we look at 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 when compared to first-born children with siblings. These only child associations are all 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 (or being deprived from siblings) has no meaningful effect on the cognitive and non-cognitive outcomes of first-born children.

While our findings clearly suggest that first-born children neither benefit nor suffer from siblings, it is important to note that our effect estimates are realized within Denmark: a country with a generous educational system and well-established child care institutions. If, for instance, sibling effects are negative and driven by siblings sharing limited resources, it is possible that we do not see that children suffer much from siblings because, in Denmark, education is heavily subsidized and skill returns are low.[15] If parents invest less in child skills (because skill returns are low) and face fewer constraints (because education costs are low), Becker's quantity-quality model predicts that the outcomes of first-born children should not be affected much by siblings. 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 in general, raised in Norway, are not worse off with more siblings. If, on the other hand, sibling effects are positive and driven by siblings socializing each other, it is also possible that we do not see that only children benefit from siblings either because, in Denmark, children have easy access to subsidized child care from ages 1 to 6. Only children, although being deprived from siblings, are not isolated and face ample socialization opportunities in child care from an early age.

Our findings, based on simple associations, also suggest that only children are a negatively selected group of children. While identifying the exact cause of

[15] 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.

negative selection into one-child families is beyond the scope of our study, we can offer some speculative explanations. One possibility is the optimal stopping rule: after a first-born child with unfavorable outcomes, parents may decide not to have a second child. Another one could be linked to divorce: children from divorced families are not only worse off (because divorce negatively impacts children's outcomes) but also more likely to end up as only child (because divorce distorts original fertility plans).

In closing, we highlight once again the three novel features of our study. First, our focus is on only children; this is an increasingly important yet often neglected group of children. Second, our econometric framework identifies the average effect of being an only child for all children whose parents rely on both first and later IVF treatments for a second child; the effect estimates are valid for reliers, not just compliers. And third, our findings are quite intriguing; we show that, at least in the Danish context, the presence or absence of siblings does not have any substantial causal effect on the cognitive and non-cognitive outcomes of first-born children.

References

Agerbo, Esben, Preben Bo Mortensen, and Trine Munk-Olsen. 2012. "Childlessness, Parental Mortality and Psychiatric Illness: A Natural Experiment based on In Vitro Fertility Treatment and Adoption," *Journal of Epidemiology and Community Health.*

Andersen, Simon, Miriam Gensowski, Steven Ludeke, and Oliver John. 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), pp. 925-939.

Angrist, Joshua, Victor Lavy, and Analia Schlosser. 2010. "Multiple Experiments for the Causal Link between the Quantity and Quality of Children," *Journal of Labor Economics,* 28(4), pp. 773-824.

Åslund, Olof, and Erik Grönqvist. 2010. "Family Size and Child Outcomes: Is There Really No Trade-Off?" *Labour Economics,* 17(1), pp. 130-139.

Baldur-Felskov, Birgitte et al. 2013. "Psychiatric Disorders in Women with Fertility Problems: Results from a Large Danish Register-Based Cohort Study," *Human Reproduction,* 28(3), pp. 683-690.

Becker, Gary. 1960. "An Economic Analysis of Fertility," In *Universities-National Bureau Committee for Economic Research, Demographic and Economic Change in Developed Countries.* Princeton: Princeton University Press

Becker, Gary, and Gregg Lewis. 1973. "On the Interaction between the Quantity and Quality of Children," *Journal of Political Economy*, 81(2), Part 2, pp. S279-S288.

Becker, Gary, and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families," *Journal of Labor Economics,* 4(3), Part 2, pp. S1-S39.

Bensnes, Simon, Ingrid Huitfeldt, and Edwin Leuven. 2023. "Reconciling Estimates of the Long-Term Earnings Effect of Fertility," IZA Discussion Papers 16174, Institute of Labor Economics (IZA).

Bhalotra, Sonia, and Damian Clarke. 2019. "Twin Birth and Maternal Condition," *Review of Economics and Statistics,* 101(5), pp. 853-864.

Bharadwaj, Prashant, Petter Lundborg, and Dan-Olof Rooth. 2018. "Birth Weight in the Long Run," *Journal of Human Resources,* 53(1), pp. 189-231.

Bhattacharjee, Natalia at al. 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," *Lancet*, 403 (05), pp. 2057-2099.

Bitler, Marianne, and Lucie Schmidt. 2012. "Utilization of Infertility Treatments: The Effects of Insurance Mandates," *Demography,* 49(1), pp. 125-149

Black, Sandra, Paul Devereux, and Kjell Salvanes. 2005. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *Quarterly Journal of Economics,* 120(2), pp. 669-700.

Black, Sandra, Paul Devereux, and Kjell Salvanes. 2005. "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," *Quarterly Journal of Economics*, 122(1), pp. 409-439.

Black, Sandra, Paul Devereux, and Kjell Salvanes. 2011. "Older and Wiser? Birth Order and IQ of Young Men," *CESifo Economic Studies,* 57(1), pp. 103-20.

Blake, Judith. 1985. "Number of Siblings and Educational Mobility," *American Sociological Review,* 1985(1), pp. 84-94.

Blake, Judith. 1989. *Family Size and Achievement.* Berkeley: University of California Press.

Bögl, Sarah, Jasmin Moshfegh, Petra Persson, and Maria Polyakova. 2024. "The Economics of Infertility: Evidence from Reproductive Medicine," NBER Working Paper 32445, National Bureau of Economic Research.

Booth, Alison, and Hiau Joo Kee. 2009. "Birth Order Matters: The Effect of Family Size and Birth Order on Educational Attainment," *Journal of Population Economics,* 22(2), pp. 367-397.

Buckles, Kasey, and Elizabeth Munnich. 2012. "Birth Spacing and Sibling Outcomes," *Journal of Human Resources,* 47(3), pp. 613-642.

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

Cameron, Lisa, Nisvan, Erkal, Lata Gangadharan, and Xin Meng. 2013. "Little Emperors: Behavioral Impacts of China's One-Child Policy." *Science* 339 (6122), pp. 953-957.

Conley, Dalton, and Rebecca Glauber. 2006. "Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility," *Journal of Human Resources,* 41(4), pp. 722-737.

Dammert, Ana. 2010. "Siblings, Child Labor, and Schooling in Nicaragua and Guatemala," *Journal of Population Economics,* 23(1), pp. 199-224.

Dayioglu, Meltem, Murat Kirdar, and Aysit Tansel. 2009. "Impact of Sibship Size, Birth Order and Sex Composition on School Enrolment in Urban Turkey," *Oxford Bulletin of Economics and Statistics,* 71(3), pp. 399-426.

De Haan, Monique. 2010. "Birth Order, Family Size and Educational Attainment," *Economics of Education Review,* 29(4), pp. 576-588.

De Haan, Monique, Erik Plug, and José Rosero. 2014. "Birth Order and Human Capital Development: Evidence from Ecuador," *Journal of Human Resources,* 49(2), pp. 359-392.

Doepke, Matthias, Anne Hannusch, Fabian Kindermann, and Michèle Tertilt. "The Economics of Fertility: A New Era," in Shelly Lundberg and

Alessandra Voena, eds., *Handbook of the Economics of the Family.* Vol. 1 (1), Elsevier, 2023, pp. 151-254.

Edmonds, Eric. 2006. "Understanding Sibling Differences in Child Labor," *Journal of Population Economics,* 19(4), pp. 795-821.

Ebenstein, Avraham. 2010. "The "Missing Girls" of China and the Unintended Consequences of the One Child Policy," *Journal of Human Resources,* 45(1), pp. 87-115.

Ejrnaes, Mette, and Claus Pörtner. 2004. "Birth Order and the Intrahousehold Allocation of Time and Education," *Review of Economics and Statistics,* 86(4), pp. 1008-1019.

Falbo, Toni. 2012. "Only Children: An Updated Review," *Journal of Individual Psychology*, 68(1), pp. 38-49.

Falbo, Toni, and Denise Polit. 1986. "Quantitative Review of the Only Child Literature: Research Evidence and Theory Development," *Psychological Bulletin,* 100(2) pp. 176-189.

Falbo, Toni, and Denise Polit. 1987. "Only Children and Personality Development: A Quantitative Review," *Journal of Marriage and the Family,* 49(2), pp. 309-325.

Feyrer, James, Bruce Sacerdote, and Ariel Stern. 2008. "Will the Stork Return to Europe and Japan? Understanding Fertility Within Developed Nations," *Journal of Economic Perspectives,* 22(3), pp. 3-22.

Frangakis, Constantine, and Donald Rubin. 2002. "Principal Stratification in Causal Inference," *Biometrics,* 58(1), pp. 21-29.

Fong, Vanessa. 2004. *Only Hope: Coming of Age under China's One-Child Policy.* Stanford University Press.

Gallen, Yana, Juanna Schrøter Joensen, Eva Rye Johansen, and Gregory Veramendi. 2023. "The Labor Market Returns to Delaying Pregnancy," Unpublished Manuscript, University of Chicago.

Gary-Bobo, Robert, Natalie Picard, and Ana Prieto. 2006. Birth Order and Sibship Sex Composition as Instruments in the Study of Education and Earnings. CEPR Discussion Paper No. 5514.

Goldin, Claudia. 2024. "Babies and the Macroeconomy," NBER Working Paper 33311, National Bureau of Economic Research.

Goux, Dominique, and Eric Maurin. 2005. "The Effect of Overcrowded Housing on Children's Performance at School," *Journal of Public Economics,* 89(5) pp. 797-819.

Groes, Fane, Anna Houštecká, Daniela Iorio, and Raül Santaeulàlia Llopis. 2024. "The Unequal Battle Against Infertility: Theory and Evidence from IVF Success," Centre for Economic Policy Research.

Harmon, Colm, Ian Walker, and Niels Westergaard-Nielsen. 2001. *Education and Earnings in Europe: A Cross Country Analysis of the Returns to Education.* Edward Elgar Publishing.

Houmark, Mikkel Aagaard. 2023. "First Among Equals? How Birth Order Shapes Child Development." Unpublished Manuscript, Aarhus University.

Ilciukas, Julius. 2024. "Parenthood, Timing, and Gender Inequality," Unpublished Manuscript, University of Amsterdam.

Kantarevic, Jasmin, and Stéphane Mechoulan. 2006. "Birth Order, Educational Attainment, and Earnings: An Investigation Using the PSID," *Journal of Human Resources,* 41(4), pp. 755-777.

Kristensen, Petter, and Tor Bjerkedal. 2007. "Explaining the Relation Between Birth Order and Intelligence," *Science,* 316(5832). pp. 1717.

Li, Hongbin, Junsen Zhang, and Yi Zhu. 2008. "The Quantity-Quality Trade-Off of Children in a Developing Country: Identification using Chinese Twins," *Demography,* 45(1), pp. 223-243.

Li, Bingjing, and Hongliang Zhang. 2017. "Does Population Control Lead to Better Child Quality? Evidence from China's One-Child Policy Enforcement," *Journal of Comparative Economics,* 45(2), pp. 246-260.

Liu, Haoming. 2014. "The Quality-Quantity Trade-Off: Evidence from the Relaxation of China's One-Child Policy," *Journal of Population Economics,* 27(2), pp.: 565-602.

Lundberg, Shelly, and Robert Pollak. 2007. "The American Family and Family Economics," *Journal of Economic Perspectives,* 21(2), pp. 3-26.

Lundborg, Petter, Erik Plug and Astrid Würtz, Rasmussen. 2017. "Can Women have Children and a Career? IV Evidence from IVF Treatments," *American Economic Review,* 107(6), pp. 1611-1637.

Lundborg, Petter, Erik Plug and Astrid Würtz, Rasmussen. 2023. On the

Family Origins of Human Capital Formation: Evidence from Donor Children. *Review of Economic Studies,* forthcoming.

Lundborg, Petter, Erik Plug and Astrid Würtz, Rasmussen. 2024. "Is There Really a Child Penalty in the Long Run? New Evidence from IVF Treatments," IZA working paper series IZA DP 15824.

Miller, Grant. 2010. "Contraception as Development? New Evidence from Family Planning in Colombia," *Economic Journal,* 120(545), pp. 709-736.

Magne Mogstad and Matthew Wiswall. 2016. "Testing the Quantity-Quality Model of Fertility: Estimation using Unrestricted Family Size Models," *Quantitative Economics,* 7(1), pp. 157-192.

Pew Research Center. 2015. *Parenting in America*.

Pedro, Juliana et al., 2019. "Infertility-Related Stress and the Risk of Antidepressants Prescription in Women: A 10-year Register Study," *Human Reproduction,* 34(8), pp. 1505-1513.

Pettersson-Lidbom, Per, and Peter Skogman Thoursie. 2009. "Does Child Spacing Affect Children's Outcomes? Evidence from a Swedish Reform. Stockholm University Working Paper No. 2009: 7.

Ponczek, Vladimir, and Andre Portela Souza. "New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country," *Journal of Human Resources,* 47(1), pp. 64-106.

Qian, Nancy. 2009. "Quantity-Quality and the One Child Policy: The Only-Child Disadvantage in School Enrollment in Rural China," NBER Working Paper Series No. 14973.

Rosenzweig, Mark, and Kenneth Wolpin. 1980. "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica,* 48(1), pp. 227-240.

Rosenwaks, Zev, Owen Davis, and Mark Damario. 1995. "The Role of Maternal Age in Assisted Reproduction," *Human Reproduction,* 10(1), pp. 165-173.

Rosenzweig, Mark, and Junsen Zhang. 2009. "Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's "One-Child" Policy," *Review of Economic Studies,* 76(3), pp. 1149-1174.

Templeton, Allan, Joan Morris, and William Parslow. 1996. "Factors that Affect Outcome of In-Vitro Fertilisation Treatment," *Lancet,* 348(9039), pp. 1402-1406.

Thwaites, Anette, Jennifer Hall, Geraldine Barrett, annd Judith Stephenson. 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), pp. 1590-1600.

Van Loendersloot, Laura, Sjoerd Repping, Patrick Bossuyt, Fulco van der Veen, and Madelon van Wely. 2014. "Prediction Models in In Vitro Fertilization; Where Are We? A Mini Review," *Journal of Advanced Research*, 5(3), pp. 295-301.

Verhaak, Chris et al. 2007. "Women's Emotional Adjustment to IVF: A Systematic Review of 25 Years of Research," *Human Reproduction Update,* 13(1), pp. 27-36.

Xiao, Yun. 2024. "Heterogeneity in the Child Quality-Quantity Trade-off and Intergenerational Mobility," CEPR Discussion Paper No. 18961

Yli-Kuha, Anna-Niina et al. 2010. "Psychiatric Disorders Leading to Hospitalization Before and After Infertility Treatments," *Human Reproduction,* 25(8), pp. 2018-2023.

Zajonc, Robert. 1976. "Family Configuration and Intelligence: Variations in Scholastic Aptitude Scores Parallel Trends in Family Size and the Spacing of Children," *Science*, 192(4236), pp. 227-236.

Zajonc, Robert, and Gregory Markus. 1975. "Birth Order and Intellectual Development," *Psychological Review,* 82(1), pp. 74.