The Effect of a Sibling's Gender on Earnings, Education and Family Formation

Peter , Noemi ; Lundborg, Petter; Mikkelsen, Sara; Webbink , Dinand

2018

Working Paper 2018:3

Department of Economics
School of Economics and Management

The Effect of a Sibling's Gender on Earnings, Education and Family Formation

Noemi Peter
Petter Lundborg
Sara Mikkelsen
Dinand Webbink

February 2018

The effect of a sibling's gender on earnings, education and family formation*

Noemi Peter[a,⋆], Petter Lundborg[b,d], Sara Mikkelsen[b], Dinand Webbink[c,d,e]

[a]Faculty of Economics and Business, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands
[b]Department of Economics, Lund University, P.O. Box 7082, SE-220 07 Lund, Sweden
[c]Erasmus School of Economics Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands
[d]Institute of Labor Economics (IZA), Germany
[e]Tinbergen Institute, The Netherlands
⋆Corresponding author.
Email addresses: a.n.peter@rug.nl (N. Peter), petter.lundborg@nek.lu.se (P. Lundborg), sara.mikkelsen@nek.lu.se (S. Mikkelsen), webbink@ese.eur.nl (D. Webbink)

This version: February 2018

Abstract

We examine how the gender of a sibling affects earnings, education and family formation. Identification is complicated by parental preferences: if parents prefer certain sex compositions over others, children's gender affects not only the outcomes of other children but also the existence of potential additional children. We employ two empirical strategies that both address this problem. First, we look at a sample of dizygotic (i.e. non-identical) twins. Second, we use a large sample of singletons to estimate whether first-borns are affected by the gender of their second-born sibling. We find that a same-sex sibling increases men's earnings and family formation outcomes (marriage and number of children), as compared to an opposite-sex sibling. Women with a same-sex sibling also earn more and are somewhat more likely to form a family in the singleton sample. A large part of the positive effect on men's income can be explained by competition among brothers. Women on the other hand seem to benefit from sisters because of shared labor market networks. The effects on family formation might stem from differential parental treatment for men, and from competition between sisters for women.

Keywords: sibling gender, sex composition, twins, income, schooling, fertility

JEL codes: J00, J16, J24, J13

*We are grateful to Hessel Oosterbeek, Erik Plug, Thomas Buser, Paul J. Devereux, Zsombor Cseres-Gergely and to seminar participants in Amsterdam, at the 2012 Summer Workshop of the Institute of Economics, Hungarian Academy of Sciences and at the ESPE Annual Conference 2014 for their comments and suggestions. Part of this research was carried out at the University of Amsterdam during the PhD studies of Noemi Peter. Lundborg acknowledges support from the Swedish Research Council for Health, Working Life, and Welfare, FORTE [grant numbers 2010-0745]. Part of the data used in this paper comes from the Swedish Interdisciplinary Panel (SIP), administered by the Centre for Economic Demography, Lund University, Sweden. The authors have no other funding sources or interests to disclose.

1 Introduction

Siblings play a significant role in most people's lives. They change the environment in which children grow up and often remain important figures in adulthood. They provide reference points to each other and also alter interactions with parents (Adams, 1999; Keim et al., 2009; McHale et al., 2003). While these observations suggest that siblings have a big potential to have an impact, evidence on causal effects in sibling relationships is still scarce.[1] We try to shed more light on this by studying the role of one particular factor: we examine how the gender of a sibling affects individuals' earnings, education and family formation.

We can expect gender to be a relevant factor because having a brother implies a different family environment and peer relationship than having a sister. Research on family conditions shows that parents are more likely to gender-differentiate their parenting in case of opposite-sex children than in case of same-sex children (McHale et al., 2003). Differential parental treatment is also present in adulthood. In particular, parents support the family formation of daughters more than the family formation of sons, for example by providing more informal childcare (Danielsbacka et al., 2011; Pollet et al., 2009). In addition to the parental treatment channel, siblings can have direct effects as well. These effects can differ by the gender composition of the sibship as same-sex sibships may be more competitive (Conley, 2000). Brothers and sisters also provide different reference points to their siblings because of gender differences in general. For example, women typically marry and have children at a younger age, work in different type of jobs and have lower earnings than men on average. In sum, those with a sister are exposed to different influences than those with a brother. We examine whether this has an impact on individuals' labor market outcomes and family formation. We focus on these outcome variables as many gender differences relate to these domains.

We investigate this question with two different empirical strategies. First, we look at twins and analyze how the gender of a co-twin affects the outcomes of the other twin. We examine men and women separately, thus, we compare men (women) who have a co-twin brother to men (women) who have a co-twin sister.[2] For this strategy we make use of data from the Swedish Twin Registry which contains information on twins' zygosity. This information is important for our purposes because only the sex of dizygotic (i.e. non-identical) co-twins is random; the sex of monozygotic (i.e. identical) twins is always the same as the sex of their co-twin (see Section 2). Therefore, estimates on the sample of all twins could suffer from "zygosity bias". That is, the coefficient of the co-twin's gender could pick up potential differences between dizygotic and monozygotic twins. We avoid this problem by restricting the estimation sample to dizygotic twins.

Our second empirical strategy is to look at singletons who have at least one younger sibling.

[1]This literature is small but emerging, with studies such as Joensen and Nielsen (2018), Black et al. (2017), Altonji et al. (2016), Nicoletti and Rabe (2014), Dahl et al. (2014) and Breining (2014). Another related literature is the one on birth order effects (see e.g. Black et al., 2005; Booth and Kee, 2009; de Haan, 2010; Breining et al., 2017 and Breining and Doyle, 2015).

[2]Separating the analysis of men and women follows the tradition in the literature (see references two paragraphs later).

Specifically, we analyze how the gender of the second-born sibling affects the outcomes of the first-born. Thus, we compare first-born men (women) who have a second-born brother to first-born men (women) who have a second-born sister. For this strategy we make use of the Swedish Multigenerational Register which contains family links for the entire population of Sweden. This enables us to create a large sample of first-born singletons (around 1 million observations for men and women altogether). We also have access to background variables which we use to validate our identification assumption empirically, that is, that the sex of the sibling is as good as randomly assigned. In addition, we can investigate not only the main effects but also several potential underlying mechanisms with high-quality data from various registers.

The literature on the effect of siblings' gender is small but recently expanding. The findings and their interpretations are both mixed.[3] The pioneering study of Butcher and Case (1994) found that women with any sisters attained lower education than women with only brothers, but their results could not be replicated by others (Kaestner, 1997; Hauser and Kuo, 1998; Conley, 2000).[4] A recent paper by Brenøe (2017) focuses specifically on participation in Science, Technology, Engineering and Mathematics (STEM) education and finds that brothers decrease the likelihood of choosing a STEM field. However, this effect persists into occupational choice only for women. She also finds that brothers increase men's earnings while they decrease women's earnings. A few other recent studies focus on labor market outcomes. Gielen et al. (2016) examine the earnings of twins and closely spaced singleton sibling pairs. They find that those with a same-sex sibling have higher earnings than those with an opposite-sex sibling, except in the case of closely spaced singleton men.[5] Rao and Chatterjee (2018) and Cools and Patacchini (2017) both look at self-reported wages from the US. The two studies use two different surveys and reach different conclusions. Rao and Chatterjee (2018) finds that the wages of women do not change by sibling sex composition, while the wages of men are increasing in the proportion of siblings who are brothers. They also find that according to a job search questionnaire, same-sex siblings serve more often as contacts to get a job. They argue that this might explain their finding on men's wages. Cools and Patacchini (2017) find that the wages of men do not change by sibling sex composition, but women with brothers earn less. Their sample is too young to measure actual family formation, but self-reported intentions show that women with brothers are more family-centric. The authors conclude that these family-focused intentions could partly explain their findings.

[3]We focus here on results from Western countries. For results on other populations, see e.g. Parish and Willis (1993), Morduch (2000), Chen et al. (2017) and Jayachandran and Pande (2017).

[4]Similarly, a draft by Pettersson-Lidbom et al. (2008) finds little evidence of an effect on various educational and labor market outcomes.

[5]Gielen et al. (2016) do not investigate specific social mechanisms behind their results as they focus on the relation between a potential biological factor and wages. We will discuss their approach and its differences from our strategy in Section 4.

In terms of empirical approach, the above studies typically include both older and younger siblings in their main estimation sample.[6] However, since parents may prefer certain sex compositions over others (Angrist and Evans, 1998), the gender of older children can affect subsequent fertility decisions. This means that the gender of an earlier born child influences the selection of a potential later born child into the sample. As we explain in Section 2 and show empirically in Section 4, this can lead to biased estimates since parental preferences may affect children's outcomes as well.

Both of our empirical strategies circumvent this problem. The twin strategy exploits the fact that twins are born at the same time, so parents cannot make decisions about one twin based on the gender of the other twin.[7] The singleton strategy also avoids the selection problem since the gender of the second-born child cannot affect the existence of the first-born child retroactively. By executing these two empirical strategies we aim to obtain clear evidence on the effect of a sibling's gender. As we use two different samples for the two strategies, our results can help to understand the robustness of the effects to different circumstances such as family constellation.

Our paper provides several other contributions as well. We give a more comprehensive picture of the effects as we also analyze family formation thoroughly. Our sample of singletons is substantially larger than the samples of the other studies. We have high quality register data not only on important outcome variables like earnings, but also on several additional variables that can be used to analyze the underlying channels systematically.

We find that in both samples, a same-sex sibling increases men's earnings, their probability of ever getting married and having kids, and their number of children. In the singleton sample, women with a same-sex sibling also earn more and are somewhat more likely to form a family. We explore several potential mechanisms behind the results. We find that a large part of the positive effect on men's income can be explained by same-sex competition: men with brothers tend to choose higher paying occupations. There is no such effect for women, and the gender of the sibling does not affect the likelihood of working in a gender-typical occupation for either gender. Instead, the positive effect on women's earnings comes largely from lower unemployment. The analysis suggests that women benefit from sisters because women can use a sister's labor market network more efficiently than the network of a brother. For family formation, our analysis suggests that differential parental treatment can be an explanation for the results on men, while competition between sisters might explain the effects on women.

The paper unfolds as follows. Section 2 explains our empirical approach in more detail and Section 3 describes the data. We present the results in Section 4 and discuss potential explanations in Section 5. Finally, Section 6 concludes.

[6] Brenøe (2017) and a draft by Pettersson-Lidbom et al. (2008) are exceptions. Rao and Chatterjee (2018) and Cools and Patacchini (2017) do a robustness check where they only use the gender of the next younger sibling as treatment.

[7] Selective abortion and IVF techniques were not available in the time period that we examine.

2 Empirical strategy

Our aim is to estimate the causal effect of a sibling's gender. The traditional approach to this is to compare those who have a brother to those who have a sister. However, this approach can lead to biased results when the estimates are based on all kinds of siblings. The problem arises because the gender of older children can influence whether parents proceed to have a next child.

For example, in developed countries a typical empirical finding is that parents of two boys or two girls are more likely to have additional children than parents of a boy and a girl (Angrist and Evans, 1998; Hank and Kohler, 2000). Clearly, this indicates parental preferences for children's sex composition. At the same time, the exact nature of these preferences is less obvious. One possibility is that this pattern arises because some families prefer a gender mix. However, another possibility is that this is not about preferring a mix, but some families prefer boys while others prefer girls. According to this explanation, some parents with same-sex kids proceed to a next child because their first two children are not from the preferred sex. Families with a boy and a girl do not proceed, since they have a child from the preferred gender for sure. Note that for this explanation to be true, it is not necessary to observe differential progression after the first child. If the number of boy-preferring families equals the number of girl-preferring families, the same number of families will proceed to a second child after a first-born boy as after a first-born girl.[8]

As this example demonstrates, parental preferences can be of various types. Thus, parents who proceed to a next child after a boy may have different preferences than parents who proceed after a girl. If people with different preferences raise their children differently, comparing individuals with an older brother to individuals with an older sister will lead to biased estimates. As we show in Appendix A.1, this intuition can be formalized as a selection bias problem, as in Heckman (1979).

To avoid the selection bias we need a setting where the sibling's gender cannot affect parents' decisions. We look at two such settings in this paper. First, we analyze a sample of dizygotic (DZ) twins.[9] This strategy circumvents the problem because twins are born at the same time, and hence parents cannot make decisions about one twin based on the sex of the other twin. Our second strategy looks at first-born singletons who have a second-born singleton sibling (and they might also have additional younger siblings). We investigate how the older child is affected by the gender of the second-born sibling. This strategy is also "clean" (i.e. free from the selection bias), since the gender of the second-born child cannot affect the existence of the first-born child retroactively. By executing these two strategies we aim to obtain strong evidence on the causal effect of a sibling's gender.

[8]Andersson et al. (2006) investigate births in Nordic countries between 1961-1999. They find that differences in parity progression ratios fluctuate over time, so there is no imbalance in second-birth probability that would be stable over the entire period. This pattern is consistent with heterogeneous parental preferences that exhibit changing frequencies in the population. Our own data for Swedish cohorts 1938-1977 shows similar fluctuations: in 1938-1949 and in the 60's, parents are slightly more likely to have a second child after a first-born boy than after a first-born girl, but the pattern is reversed in 1970-1977 and there is no imbalance in the 50's.

[9]Unlike monozygotic (MZ) twins, dizygotic (DZ) twin pairs can be opposite-sex or same-sex; see further discussion at the end of this section.

We analyze men and women separately both in the twin and in the singleton setting. Thus, altogether we conduct four comparisons: 1) we compare men who have a co-twin brother to men who have a co-twin sister; 2) we compare women who have a co-twin brother to women who have a co-twin sister; 3) we compare first-born singleton men who have a second-born brother to first-born singleton men who have a second-born sister; 4) we compare first-born singleton women who have a second-born brother to first-born singleton women who have a second-born sister. All four comparisons have the same general structure, so our empirical strategy can be summarized by a simple OLS regression that is estimated on four different samples. Specifically, we estimate

$$Y_i = \alpha + \beta G_i + \gamma X_i + U_i \tag{1}$$

where Y_i denotes the outcome of individual i (who is a twin man in comparison 1, a twin woman in comparison 2, a first-born singleton man in comparison 3, and a first-born singleton woman in comparison 4)[10] and G_i denotes the gender of the sibling (the gender of the co-twin in comparisons 1 and 2, and the gender of the second-born singleton in comparisons 3 and 4). X_i denotes our control variables: in the twin samples, we control for the presence of older siblings and for birth year fixed effects; in the singleton samples, only first-borns are included and hence we only control for birth year fixed effects.[11]

Since the gender of existing children can affect the number of younger siblings, the estimates have to be interpreted as a total impact that includes the family size channel. Controlling for family size would lead to bad control bias (Angrist and Pischke, 2009), so instead we will investigate the family size channel separately in Section 5.2. The analysis in that section indicates that the effects that we find are not driven by this channel.

Our key identification assumption is that sex is random, which is in line with the established view in biology.[12] In subsection 3.2 we look at empirical patterns in the singleton data and find that they are consistent with this assumption: the spacing of births and parental background cannot explain variation in G_i.[13] In the twin setting, zygosity could be an additional factor that threatens identification, so we need to address this issue as well. By zygosity we refer to the fact that twinning can occur in two ways (Phillips, 1993). Dizygotic (DZ) twins arise from two different zygotes, which

[10]In the twin samples, individuals can come from the same family (e.g. we can obtain two observations from a male-male twin pair). Therefore we will estimate standard errors such that we take into account potential clustering at the level of the family of origin.

[11]We handle older siblings differently in the twin and singleton samples because of sample size considerations. Twinning is a rare event and twinning at first parity is even more rare. To avoid sample size problems, we do not restrict the twin sample to those with no older siblings, but control for the presence of older siblings instead. This is controlling for already realized fertility that cannot be affected by the sex composition of the twin pair. In case of the singletons there are a substantial number of births and hence we can restrict the sample to first-borns for an easier interpretation of the effect.

[12]See e.g. Reece et al. (2010). Note that medical technologies for sex selection were unavailable in the time period that we examine. As far as natural sex selection is concerned, Wilcox et al. (1995) found that the timing of intercourse in relation to ovulation has no significant effect. Similarly, Gray et al. (1998) found no evidence for the hypothesis that maternal hormones influence sex selection.

[13]We do not have access to such background variables in the twin sample (see section 3 for details about our data).

result from the separate fertilization of two different ova by two different sperms. Their sex is determined the same way as the sex of singletons: the offspring will be male (female) if the sperm bears a Y (X) chromosome. In contrast, monozygotic (MZ) twins arise when one zygote splits into two genetically identical units, so MZ twin pairs are always same-sex. This means that conditional on own sex, the sex of the co-twin correlates with zygosity. Thus, if we could not differentiate twins by zygosity, differences between opposite-sex and same-sex twins could reflect not only the impact of the sibling's gender, but also the impact of having a genetically identical sibling or other potential differences between MZ and DZ twins. To make sure that our estimates are free from this "zygosity bias", we sought out a dataset which contains zygosity information and base our twin strategy solely on DZ twins.

3 Data

3.1 Twin data

Our twin analysis is based on a sample of dizygotic twins who were born in Sweden in 1926-1958. The data comes from the SALT project (Screening Across the Lifespan Twin Study) of the Swedish Twin Registry (STR), which surveyed twins irrespective of their sex composition (Lichtenstein et al., 2002). The STR had records of every twin birth from the national birth register and they contacted all available twins for a phone interview.[14] The process started with a pilot in 1996-1997 and then full scale data collection took place in 1998-2002. The oldest cohorts were surveyed first, followed by younger cohorts.[15] The interviewers tried to get as many responses as possible; e.g. people were called back at least five times if they were unavailable. This led to the remarkably high response rate of 74 percent.

The STR determined zygosity based on questions about intrapair similarity in childhood. This classification was validated in the pilot: 13 DNA-markers were analyzed in blood samples. Zygosity assignment proved correct in 99% of the cases (see Lichtenstein et al., 2002).

The survey provided information on marital status, children and siblings as well. In case of marriage, we use the dummy "ever married" that takes one if the individual has ever been married or cohabited and zero otherwise. The questions on children and siblings refer to biological children and siblings (other than the co-twin) who are still alive. We look at the number of children to analyze

[14] The SALT contains twins from cohorts 1886-1925 as well, but information on these cohorts was compiled differently. Specifically, whether twins from these cohorts entered the SALT project depended on their participation in earlier surveys, which mainly targeted same-sex pairs. In turn, the SALT contains disproportionately few opposite-sex twins from cohorts 1886-1925. We do not use these early cohorts in our research, only the complete cohorts of 1926-1958.

[15] Of course "younger" is meant in relative terms; participants are 41-74 years old. The STR focused on these ages because they wanted to screen for some diseases. A population above 40 is actually quite fortunate for our purposes, since fertility is typically completed by that age.

fertility.[16] To analyze timing, we calculate age at first birth from the birth year of children.

For the analysis of education and income we obtained registry data. The education data comes from the education register (utbildningsregistret, UREG 1990-2007); in a few cases when that is unavailable, we use information from the survey. Years of schooling is calculated from obtained degree. The income data comes from the national register on taxable income (IoT, Inkomst- och taxeringsregistret). This is *not* based on individuals' self-reports but on information collected from employers and other agencies providing taxable benefits. We use records from years 1971-2007 (all converted to 2010 prices).[17] The income variable measures total taxable labor income in the given year: it includes wages, income from self-employment, pensions, sickness benefits and other taxable benefits.[18] From any given year, we use the available income records of those who are aged between 25 and 64 in that year. The main source of income in this age range is typically the labor market.

Since we have income records from many years, we have several observations per person. We take their average to get a permanent income measure for each individual. We use the log of this measure, ln(permanent income) as outcome variable. Note that people may have zero income in one year or another; however, almost nobody has zero income in all years so taking the logarithm leads to hardly any missing observations.[19] On the other hand, we averaged over different number of observations in different cohorts. This implies potential heteroskedasticity, even though cohort fixed effects are included in the regressions. We will use robust standard errors to take this into account.

To balance the sample across the main outcomes, we drop individuals whose education, income measure, marital status or number of children is missing.[20] Summary statistics are shown in Table 1. The first column shows statistics for dizygotic twins whose co-twin is of the opposite sex, while the second column shows statistics for dizygotic twins whose co-twin is of the same sex. Recall that the SALT surveyed individuals, not twin pairs, so inclusion in the sample is not conditional on the

[16]We imputed data for childless women. This is because the SALT survey was built up such that before the questions on siblings and children, there was a separate section for women only, where they were asked about pregnancy and giving birth. The questions on children were subsequently not asked from women who never gave birth. These women got 'missing' assigned for the children variables, even though their number of children should be zero. We do not have access to the variable on childbirth so we do not know which observations are truly missing and which ones should be recorded as zero. However, the questions on children follow directly the ones on siblings, so we inferred who could have answered them. In particular, a woman is assumed to have zero children if she answered the question on siblings but not the question on children. Since data on men is complete, we can analyze that to check the accuracy of our imputation strategy. The results are reassuring: 99.7% of men who answered the question on siblings also answered the question on children, so it seems safe to assume that those who answered the former would have also answered the latter question.

[17]In fact, we have income data for years 1968-2007. However, after inspecting the data, we had concerns about sample selection in the first 3 years. In 1968-1970, 16.24% of the observations are missing, whereas data is almost complete after that period. Therefore we use only the 1971-2007 data in the analysis.

[18]Many benefits became taxable in 1974 and therefore they are included in the income measure since then. This means that our variable was somewhat more restrictive in the pre-1974 years. Nonetheless, recent work by Björklund, Jäntti, and Lindquist (2009) suggests that this is not a major concern, as benefits constitute a very small share of total earnings. In any case, the estimates are similar if we restrict income data to 1974-2007.

[19]We lose only 0.08% of women. We do not lose any men.

[20]Since age at first birth cannot be calculated for individuals who do not have children, we do not balance the sample with respect to this variable.

participation of the co-twin.

These statistics provide a first impression of the effect of having a brother or a sister for twins. Although the differences are quite small, we observe that male twins with a brother are more likely to get married and to have kids. They also have more kids and earn somewhat more. Female twins with sisters are younger at first birth and obtain less schooling. The next two columns show differences between singletons. Again, the differences are small but mostly consistent with what we observed for the twins.

3.2 Singleton data

Our dataset on singletons comes solely from administrative registers. We identify first-born singletons born 1938-1977 with a second-born full singleton sibling using family links in the Multigenerational register. We take 1938 as the first cohort as we want to select individuals without older siblings and the Multigenerational register does not contain information on siblings before 1932. The 6 years cutoff is reasonable given that it is atypical to space the birth of siblings further apart.[21] For the same reason, we restrict the sample to individuals whose second-born sibling was born at most 6 years apart. In addition to the second-born sibling, people in our sample might have more younger siblings as well.

We add information from several registers in order to create variables for labor market and family formation outcomes that follow the definitions used for the twins as closely as possible. We derive years of schooling from highest education registered in the education register (UREG 1990-2011), and when that is not available, we use educational information from the 1970 census. Permanent income is defined exactly the same way as for twins, i.e. the natural logarithm of average income between ages 25-64, but for singletons we can use records from 1971-2011. Family formation variables are also defined the same way as for twins, however cohabitation is not included in "ever married" for singletons because it is only observed in the registers if a cohabiting couple has children together. We observe children born to the singletons up to 2012; thus, our sample is at least 35 years old when we measure number of children. We balance the sample across the main outcomes just like we do for the twin sample.

We present the summary statistics for the singleton sample in Column 3 and 4 of Table 1. Column 3 shows statistics for first-born singletons whose second-born sibling is of the opposite sex, while column 4 shows statistics for first-born singletons whose second-born sibling is of the same sex. Since the sex ratio in the general population is slightly imbalanced (about 51% males, 49% females), the share of same-sex siblings is 51% for men and 49% for women. These shares are fifty-fifty for our

[21] In almost ninty percent of the cases, the first-born and the second-born have an age difference of at most 6 years. Starting the sample from a later cohort does not change much in this statistics.

Table 1: Summary statistics

	Dizygotic twins (cohorts: 1926-1958)		First-born singletons (cohorts: 1938-1977)	
	OS co-twin	SS co-twin	OS SB sibl.	SS SB sibl.
Panel 0: Men				
Ever married[a]	0.875	0.889	0.670	0.679
	(0.330)	(0.315)	(0.470)	(0.467)
Any kids	0.817	0.830	0.792	0.797
	(0.387)	(0.376)	(0.406)	(0.402)
Number of kids	1.865	1.908	1.775	1.792
	(1.239)	(1.234)	(1.236)	(1.235)
Age at first birth	27.882	27.854	28.901	28.790
	(5.154)	(5.229)	(5.431)	(5.419)
Number of siblings[b]	2.006	1.992	1.752	1.833
	(1.890)	(1.815)	(1.108)	(1.113)
Years of schooling	10.778	10.781	12.158	12.168
	(3.120)	(3.131)	(2.679)	(2.677)
Ln(permanent income)	12.434	12.447	12.455	12.463
	(0.397)	(0.385)	(0.566)	(0.561)
Number of obs.	6759	6668	257174	271432
Panel 1: Women				
Ever married[a]	0.896	0.896	0.745	0.750
	(0.305)	(0.305)	(0.436)	(0.433)
Any kids	0.875	0.871	0.872	0.874
	(0.330)	(0.335)	(0.334)	(0.332)
Number of kids	1.984	1.996	1.983	1.985
	(1.170)	(1.183)	(1.140)	(1.134)
Age at first birth	25.064	24.822	26.448	26.362
	(4.774)	(4.725)	(5.154)	(5.138)
Number of siblings[b]	2.048	2.031	1.756	1.841
	(1.867)	(1.858)	(1.111)	(1.134)
Years of schooling	10.839	10.662	12.571	12.563
	(2.976)	(2.981)	(2.579)	(2.583)
Ln(permanent income)	11.935	11.914	12.104	12.109
	(0.557)	(0.603)	(0.579)	(0.576)
Number of obs.	7401	7341	256788	241953

Note: OS means opposite-sex, SS means same-sex, SB means second-born. Standard deviations in parentheses. The singleton sample consists of individuals whose second-born sibling was born at most 6 years later. [a] Marriage includes cohabitation in case of the twin sample, but not in case of the singleton sample. [b] Number of siblings does not include the co-twin.

dizygotic twins, which is in line with previous research on twins.[22] The pattern in the twin sample is consistent with the notion that male fetuses are more vulnerable and hence a male co-twin can die in utero more easily (Hansen et al., 1999, Mizuno, 2000).[23] From Table 1 we can also see that in spite of this, there are somewhat more women in the twin sample. This is probably because women are more likely to respond to surveys (Singer et al., 2000).[24]

For the singleton sample we have more rich data available from the registers than for the SALT sample, which we can exploit in additional ways. First, we observe educational level not only for the singletons themselves but also for their parents. We utilize this data to get a measure of the socio-economic background of the individuals. For this we first determine whether the parents have more or less years of schooling than the median among parents who had their first child in the same year, separately for mothers and fathers. Based on their parent's education, we subsequently create four categories for the singletons: "High parental education" (23%) means that both parents are higher educated than the median; "Mixed parental education" (27%) means that one of the parents has more years of schooling than the median while the other's education is at or below the median; "Low parental education" (38%) means that both parents have at most median years of schooling; "Missing parental education" (12%) denotes that either parent has missing years of schooling.[25] In Section 5 we will use this classification of socio-economic background to examine the potential mechanisms behind our results. In addition, we can now use the parental educational categories together with additional background variables (age difference between the first-born and second-born sibling in months and the mother's and father's age at the first birth in years) to verify that the gender of the second-born sibling is as good as randomly assigned. More specifically, we run regressions of second-born sex on these background variables, separately for men and women. The results are shown in Table 2. As we can see, the background variables do not explain variation in the gender of the second-born, which supports our identification assumption.[26]

[22] The fifty percent share is so widely accepted that it is even called a "rule", the so-called Weinberg's differential rule (WDR) (Benirschke et al., 2012). Although James (1979) raised questions about this, several studies found that it is supported empirically (Vlietinck et al., 1988; Husby et al., 1991; Fellman and Eriksson, 2006). Thus, Elston et al. (2002) conclude that this is a reasonable approximation and actual deviations are small.

[23] If fetal vulnerability is related to later health, brothers in the twin sample might be healthier on average than brothers in the singleton sample, which might lead to observing more positive effects of brothers in the twin sample than in the singleton sample.

[24] Note that we analyze men and women separately, so this will simply mean that the analysis of the latter will be based on more observations. The response rate is presumably unrelated to the sex of the co-twin since the fifty percent share of sisters is in line with earlier research on twins (see footnote 22).

[25] There is no difference in the distribution of parental education categories between first born singleton men and women.

[26] After presenting our main results in Section 4, we will also do a robustness check and verify that our results are not sensitive to including these variables as controls in the regressions.

Table 2: Randomness of the sex of the second-born sibling

	Second-born is female
Panel 0: Men	
Birth spacing	-0.000
	(0.000)
Age 1st birth mother	-0.000
	(0.000)
Age 1st birth father	0.000**
	(0.000)
High parental educ.	0.002
	(0.002)
Mixed parental educ.	0.002
	(0.002)
Missing parental educ.	0.001
	(0.003)
Observations	528,606
R-squared	0.000
Panel 1: Women	
Birth spacing	0.000
	(0.000)
Age 1st birth mother	0.000
	(0.000)
Age 1st birth father	-0.000
	(0.000)
High parental educ.	-0.002
	(0.002)
Mixed parental educ.	-0.001
	(0.002)
Missing parental educ.	0.003
	(0.003)
Observations	498,741
R-squared	0.000

Note: The sample consists of first-born singletons born 1938-1977 whose second-born sibling was born at most 6 years later. Birth spacing means the age difference between the first-born and the second-born child (in months). The regressions control for cohort fixed effects. Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

In addition to parental background, we also have a substantial amount of additional register data for our singleton sample, such as occupational choice, work industry and unemployment, which we will use later to explore potential mechanisms behind the effects of the sibling's gender. The variables used for this additional analysis will be described in more detail in Section 5 (also see Table A5 in Appendix A.4).

4 Results

In this section we show results on the effect of a sibling's gender. As explained in the empirical strategy section, we analyze men and women separately in both the twin and the singleton samples. For the interpretation of the effects, it is useful to let the sibling gender dummy G_i indicate a brother for the male samples and a sister for the female samples. This enables us to look at the estimates as the effect of having a same-sex sibling, for both men and women. Note that the coefficients capture a total impact which can result from several underlying mechanisms; we will discuss these after the main findings are presented.

The main results for twins are reported in Table 3. For men, having a brother instead of a sister has no impact on education, which is consistent with the findings of previous literature.[27] On permanent income we see a positive impact (significant at the 10 percent level). The gender of the co-twin affects family formation as well: those with brothers are more likely to get married and have children, and they also have more children.

The results on women show a different pattern. Women with sisters have lower education, which is in line with the early study of Butcher and Case (1994) but not with later studies. There is no significant effect on income. For family formation, the estimates are small and insignificant. However, the estimated coefficient for age at first birth suggests that there is an effect on timing: women who get children do so at a younger age if they have a sister compared to a brother.[28]

The magnitude of the effects is small but not trivial. For instance, the estimated effect of the earnings premium for men with brothers is one percent over the whole life cycle. This is also similar to previous findings on other dimensions of family environment, such as birth order effects (Black et al., 2005). Moreover, males with a brother have at least 0.04 more children. This is equivalent with at least 2 percent more children.

As we explained in the empirical strategy section, these estimates are free from zygosity bias as they are based solely on DZ twins. Comparing these estimates to results on all twins (including MZs) can be helpful, since zygosity information is often difficult to obtain for researchers. We investigate this issue in Appendix A.2 and find that zygosity bias can indeed distort most of the results, with the

[27] See e.g. Butcher and Case (1994) or Brenøe (2017). Also, Ahrenfeldt et al. (2015) find no effect on test scores for twins.

[28] Age at first birth is conditional on having any children, which decreases the sample size for this outcome and warrants an indicative interpretation of the coefficients.

Table 3: Main results for dizygotic twins

	Years of schooling	Ln(permanent income)	Ever married	Any kids	Number of kids	Age at first birth
Panel 0: Men						
SS co-twin	-0.013	0.012*	0.013**	0.013*	0.044**	-0.035
	(0.055)	(0.007)	(0.006)	(0.007)	(0.022)	(0.101)
Observations	13,427	13,427	13,427	13,427	13,427	10,978
Outcome mean	[10.78]	[12.44]	[0.88]	[0.82]	[1.89]	[27.87]
Panel 1:Women						
SS co-twin	-0.122**	-0.014	-0.000	-0.004	0.013	-0.207**
	(0.049)	(0.010)	(0.005)	(0.006)	(0.020)	(0.086)
Observations	14,742	14,742	14,742	14,742	14,742	12,836
Outcome mean	[10.75]	[11.92]	[0.90]	[0.87]	[1.99]	[24.94]

Note: The sample consists of dizygotic twins born 1926-1958. “SS co-twin” is a dummy which takes 1 if the co-twin is of the same sex as the individual, and 0 if the co-twin is of the opposite sex. “Ever married” is a dummy which takes 1 if the individual has ever been married or cohabited, 0 otherwise. Permanent income is average income in years 1971-2007 (if subject is aged 25-64 in the given year). All regressions control for older sibling, missing sibling information and cohort fixed effects. The standard errors shown in parentheses are robust and clustered by the family of origin. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

severity of the distortion varying by outcome and sample.

In our twin analysis we included dummy variables for the presence of, and missing information on, older siblings.[29] With these variables we aim to control for variation in outcomes that arises because of differences in family constellation. However, the impact of the co-twin's gender itself may differ depending on the presence of older siblings. Table 4 shows results from including interactions between the variables on older siblings and having a same-sex co-twin. The first row of each panel shows the estimated effects for twins who reported zero older siblings. For men, we see that the results on the income and family formation of these twins are in line with the main results in Table 3. Although the estimated effects are not statistically significant due to the inclusion of the interaction terms, they are of similar magnitude as in Table 3. For women, on the other hand, the negative effect on education appears to be driven by those with older siblings.

The results for first-born singletons are shown in Table 5. Recall that for singletons we estimate the effect of the second-born's gender on outcomes of first-born singletons who have at least one younger sibling. For men, the results are in line with what we observed for male twins. Men with brothers have a higher permanent income compared to men with sisters. They also have a higher probability to ever be married and have kids. For singleton women, the estimated effects of having a

[29] Almost sixty percent of the twins report having olders siblings. Missing sibling information means that the person did not answer the questions on the (birth year of) siblings (about eight percent of the sample).

Table 4: Main results for dizygotic twins by family structure

	Years of schooling	Ln(permanent income)	Ever married	Any kids	Number of kids	Age at first birth
Panel 0: Men						
SS co-twin	-0.018	0.015	0.009	0.016	0.052	0.068
	(0.102)	(0.013)	(0.010)	(0.012)	(0.037)	(0.177)
Older sib.*SS	0.021	-0.004	0.003	-0.005	-0.010	-0.179
	(0.124)	(0.016)	(0.012)	(0.015)	(0.046)	(0.219)
Missing sib.*SS	-0.089	-0.011	0.027	-0.003	-0.023	0.053
	(0.186)	(0.025)	(0.024)	(0.026)	(0.094)	(0.408)
Observations	13,427	13,427	13,427	13,427	13,427	10,978
Outcome mean	[10.78]	[12.44]	[0.88]	[0.82]	[1.89]	[27.87]
Panel 1:Women						
SS co-twin	-0.039	-0.014	-0.006	-0.008	0.000	-0.178
	(0.090)	(0.017)	(0.009)	(0.010)	(0.035)	(0.152)
Older sib.*SS	-0.160	0.004	0.010	0.006	0.025	-0.107
	(0.109)	(0.021)	(0.011)	(0.012)	(0.043)	(0.187)
Missing sib.*SS	0.223	-0.055	0.003	-0.000	-0.035	0.571
	(0.193)	(0.047)	(0.024)	(0.023)	(0.087)	(0.366)
Observations	14,742	14,742	14,742	14,742	14,742	12,836
Outcome mean	[10.75]	[11.92]	[0.90]	[0.87]	[1.99]	[24.94]

Note: The sample consists of dizygotic twins born 1926-1958. “SS co-twin” is a dummy which takes 1 if the co-twin is of the same sex as the individual, and 0 if the co-twin is of the opposite sex. “Ever married” is a dummy which takes 1 if the individual has ever been married or cohabited, 0 otherwise. Permanent income is average income in years 1971-2007 (if subject is aged 25-64 in the given year). All regressions control for older sibling, missing sibling information and cohort fixed effects. The standard errors shown in parentheses are robust and clustered by the family of origin. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Main results for first-born singletons

	Years of schooling	Ln(permanent income)	Ever married	Any kids	Number of kids	Age at first birth
Panel 0: Men						
SS sibling	0.006	0.007***	0.009***	0.006***	0.018***	-0.124***
	(0.007)	(0.002)	(0.001)	(0.001)	(0.003)	(0.016)
Observations	528,606	528,606	528,606	528,606	528,606	419,902
Outcome mean	[11.93]	[12.46]	[0.70]	[0.80]	[1.78]	[28.39]
Panel 1: Women						
SS sibling	-0.003	0.005***	0.005***	0.002**	0.001	-0.074***
	(0.007)	(0.002)	(0.001)	(0.001)	(0.003)	(0.015)
Observations	498,741	498,741	498,741	498,741	498,741	435,395
Outcome mean	[12.27]	[12.09]	[0.76]	[0.88]	[1.91]	[26.23]

Note: The sample consists of first-born singletons born 1938-1977 whose second-born sibling was born at most 6 years later. “SS sibling” is a dummy which takes 1 if the second-born sibling is of the same sex as the first-born and 0 if the second-born sibling is of the opposite sex. “Ever married” is a dummy which takes 1 if the individual has ever been married, 0 otherwise. Permanent income is average income in years 1971-2011 (if subject is aged 25-64 in the given year). All regressions control for cohort fixed effects. Robust standard errors are shown in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

same-sex sibling are similar to the results on men, although the magnitudes are smaller. On income, there is a small positive effect of having a sister. Furthermore, the family formation estimates suggest that women with sisters have a slightly higher likelihood of getting married and have children at a younger age than women with brothers. The results are robust to the inclusion of additional controls, as can be seen from Table A3 in Appendix A.3.[30]

Since we observe income over several years (1971-2011) we can also check whether the income effect differs during the life-cycle. Figure A1 in Appendix A.3 shows results from running separate regressions for income at different ages. The outcome is the natural logarithm of a five-year moving income average. Men with brothers have a higher average income than men with sisters already in their mid 20s and the premium is increasing in age up to the mid 50s. This suggests that men with brothers have a steeper income profile than men with sisters. For women, a same-sex sibling does not seem to alter the slope of the income profile. The small estimated income effect is more or less constant from the mid 30s and onwards.

Recall that our identification strategy circumvents the selection bias. As we argued in Section 2

[30] Also note that using a probit or logit model to estimate the impact on the probability of ever getting married or having kids does not alter the results for twins or singletons (untabulated results). As another robustness check, we also looked at pooling ln(income) from different years, instead of calculating permanent income over the years. In addition to the usual controls, these pooled regressions on ln(income) also include year fixed effects and use weights such that every person is represented equally. We find that the results on income are qualitatively the same as in the main specifications, in both the twin and singleton samples (untabulated results).

(and Appendix A.1), this does not hold for alternative strategies such as comparing individuals with older brothers to individuals with older sisters, since these two types of individuals may come from families with different types of gender preferences. In Table A4 of Appendix A.3 we show estimated effects of having an older sibling of the same sex for second-born singletons. These estimates confirm the notion of selection bias as they are completely different from the results of Table 5 for a number of outcomes. For example, the coefficients are zero for men's income and marriage, and the estimates on any kids and number of kids are negative for both men and women. These findings support our argument that one should not include younger siblings in the estimation sample.

Taking the results from our two estimation strategies (twins and first-born singletons) together, we get strong evidence that a sibling's gender has a positive impact on men's earnings, probability of family formation and number of children. The effects on women are less clear cut. The effect on the schooling of twin women is not robust to family composition and is not observed in the singleton sample. The effects on women's income and probability of family formation are only observed in the singleton sample and they are smaller than the corresponding effects on men. However, women with sisters have their first child at an earlier age in both samples. In Figure A2 in Appendix A.3 we examine the result on timing further by plotting the effect of a same-sex sibling on the probability of being a mother by a certain age. In the singleton sample, we find that women with sisters are more likely to have children already at the end of their teenage years. Nonetheless, most of the effect comes from women's fertility decisions in their 20s, in both the singleton and the twin sample. In the twin sample women with brothers eventually catch up, while in the singleton sample this happens only partly, leading to the small increase in the probability of having children that we observed already in Table 5.

Differences in results for women might come from the fact that the twin sample consists of older cohorts. Women's educational and labor market position improved substantially over the years, and family planning also became easier.[31] Indeed, when we look at differences by cohort, we find that the same-sex coefficient on income and education is most negative in the early cohorts (untabulated results). However, this does not apply for the results on family formation.

Another reason why results between the two samples might differ is spacing. To explore this possibility we look at results from a sample of singletons who are closely spaced. This exercise can be informative not only about the persistence of the estimates but also in understanding the potential mechanisms at play (we will come back to this second point in the next section). Table 6 shows results on first-born singletons whose younger siblings are born within 24 months. For men, the estimates become larger in magnitude compared to the main results in Table 5. It is also notable that the estimated effect on men's income is very similar to that of twins. For women, the magnitude of

[31] Social attitudes towards the use of contraceptives and abortion in Sweden became more liberal in the 1960's. In the mid-1970s important legal changes and policy efforts started. A new abortion law was enacted in 1974 making abortion free on request. Contraceptive services were expanded throughout the country in the same years (see Oláh and E.M. Bernhardt, 2008; Swedish Institute, 1994).

Table 6: Main results for closely spaced singletons (at most 24 months spacing)

	Years of schooling	Ln(permanent income)	Ever married	Any kids	Number of kids	Age at first birth
Panel 0: Men						
SS sibling	0.030**	0.010***	0.013***	0.007***	0.025***	-0.134***
	(0.015)	(0.003)	(0.002)	(0.002)	(0.007)	(0.034)
Observations	129,492	129,492	129,492	129,492	129,492	102,205
Outcome mean	[11.94]	[12.42]	[0.68]	[0.79]	[1.82]	[28.53]
Panel 1: Women						
SS sibling	0.017	0.011***	0.001	0.001	-0.007	-0.037
	(0.014)	(0.003)	(0.002)	(0.002)	(0.007)	(0.031)
Observations	122,398	122,398	122,398	122,398	122,398	106,912
Outcome mean	[12.33]	[12.07]	[0.75]	[0.87]	[2.04]	[25.89]

Note: The sample consists of first-born singletons born 1938-1977 whose second-born sibling was born at most 24 months later. "SS sibling" is a dummy which takes 1 if the second-born sibling is of the same sex as the first-born and 0 if the second-born sibling is of the opposite sex. "Ever married" is a dummy which takes 1 if the individual has ever been married, 0 otherwise. Permanent income is average income in years 1971-2011 (if subject is aged 25-64 in the given year). All regressions control for cohort fixed effects. Robust standard errors are shown in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

the income effect increases while the effects on family formation disappear. Thus, a closer spacing might explain why we did not find effects on the family formation of twin women.

The approach of comparing the effect of a sibling's gender for twins and closely spaced singletons has been used by Gielen et al. (2016) to examine the relation between a biological factor and wages. In particular, their study is based on the Twin Testosterone Transfer (TTT) hypothesis, which is the assumption that testosterone can transfer between twins in utero. Under this assumption, those with a male co-twin would be exposed to higher prenatal testosterone than those with a female co-twin.[32] The main result of Gielen et al. (2016) is that same-sex male twins earn significantly more than opposite-sex male twins while closely spaced male singletons are unaffected by the gender of their sibling. The authors interpret their finding such that prenatal testosterone increases the wages of men.

[32]The inspiration for the TTT hypothesis comes from animal studies which found that testosterone can transfer between littermates (e.g. vom Saal and Bronson, 1980; Even et al., 1992). The impetus to the TTT hypothesis was given by Miller (1994), who speculated that testosterone may transfer in case of humans as well and argued that human twins should be studied to uncover potential signs of testosterone transfer. Several papers examined the outcomes of twins in this vein - for reviews, see Cohen-Bendahan et al. (2005a) and Tapp et al. (2011). The reviews find that the results of this literature lack consistency.

Our paper is not specifically focused on the TTT hypothesis, partly because of the lack of clear evidence behind this assumption.[33] Nonetheless, given the related context it is noteworthy that in our paper the effect on men's income is also significant for singletons, which would be difficult to explain with testosterone transfer.[34] There are at least two important differences between the empirical strategy of Gielen et al. (2016) and our paper which could explain the difference in findings. First, we focus on first-born singletons whereas Gielen et al. (2016) include singletons irrespective of birth order. As we demonstrated with Table A4, estimates of the effect of an older sibling's gender on younger siblings suffers from selection bias, which distorts the results on men's income towards zero. Second, when we look at closely spaced singletons we use a cutoff of 24 months, which is a common spacing (more than 20 percent of second-borns are spaced within this range in our data). In contrast, Gielen et al. (2016) choose a cutoff spacing of 12 months in their main specification. This is a very atypical sample: only 1 percent of second-borns are spaced within this range in our data, and less than 4 percent of siblings in the data of Gielen et al. (2016). Since pregnancies normally last about 9 months, there might be special reasons for a birth spacing of at most 12 months (e.g. religious beliefs, unplanned pregnancies), which could also alter sibling interactions or how they are raised. To examine this issue empirically, we run the estimates on men's income on the sample of first-borns whose second-born sibling is at most 12 months apart. We found that the coefficient is not only insignificant but actually turns negative: it is -0.012, with a standard error of 0.015. Thus, this small sample of atypically shortly spaced siblings seems to be under the influence of special factors, which obscure the effect that is identified in the main sample of first-borns and in the sample of 24-month-spaced first-borns.[35]

[33]To the best of our knowledge, the only direct test of this hypothesis is provided by the data of Abeliovich et al. (1984) which do not support the TTT assumption. In particular, Abeliovich et al. (1984) measured testosterone levels in the amniotic fluid of twin fetuses. Their data show that testosterone levels are not elevated for twins with a male co-twin. Other studies that measure testosterone are also not supportive of the assumption that it can transfer from a fetus to others. In particular, several studies failed to find a difference in maternal testosterone levels by the sex of the fetus (see Glass and Klein, 1981; Rodeck et al., 1985; van de Beek et al., 2004; Cohen-Bendahan et al., 2005b; Toriola et al., 2011; only Meulenberg and Hofman, 1991 is an exception) and those that measured testosterone in both amniotic fluid and maternal serum found no correlation between the two measures (Rodeck et al., 1985; van de Beek et al., 2004).

[34]For women, we can rule out the TTT channel in an additional way. Animal studies have shown that females enter puberty later if they develop between male fetuses instead of female fetuses (Ryan and Vandenbergh, 2002). In the human setting this would imply that women with co-twin brothers would enter puberty later than women with co-twin sisters. Thus, we estimate the effect of a co-twin sister on age at menarche (first menstruation) and find a statistically insignificant positive coefficient, which is inconsistent with the predictions of the hypothesis.

[35]As a robustness check, Gielen et al. (2016) also look at 24-months spacing, but they do so using older and younger siblings as well. As we pointed out above, such estimates can suffer from the selection bias.

5 Potential mechanisms

Our results for men show that having a sibling of the same sex increases earnings and the probability of family formation. The pattern is similar for women but less strong and less clear because of differences in results between twins and singletons. In this section, we evaluate potential mechanisms behind these results. In subsection 5.1, we consider explanations related to differential parental treatment and to direct sibling-to-sibling influences. In subsection 5.2, we investigate two mechanisms related to constraints: economies of scale and family size.

5.1 Differential parental treatment and direct sibling-to-sibling influences

A sibling's gender may affect labor market outcomes and family formation indirectly via differential parental treatment and directly via sibling-to-sibling influences. We start this section by reviewing these potential mechanisms, so that their predictions become clear. For the ease of the reader we also give an overview of these predictions in Table 7 and summarize our results as well in the same table. After having reviewed the mechanisms and their predictions, we investigate them empirically with the help of additional register data that we have on singletons (see Table A5 for detailed descriptions of the additional variables).

Differential parental treatment occurs when parents treat boys and girls differently. McHale et al. (2003) reviews research on the family contexts of gender development and finds that such parental behavior occurs more frequently in case of opposite-sex children than in case of same-sex children. As a result, individuals with opposite-sex siblings might be more likely to make gender stereotypical choices than those with same-sex siblings. Brenøe (2017) shows that this is the case for the choice of educational field. In particular, she finds that women with brothers are less likely to complete an education in Science, Technology, Engineering and Mathematics (STEM) compared to women with sisters. She also provides evidence that parents of opposite-sex children gender-specialize their parenting more than parents of same sex children, which could reinforce gender-conformity in the children. Rao and Chatterjee (2018) and Cools and Patacchini (2017) report survey findings that are also in line with this mechanism: women's responses to attitude questions suggest that those with brothers hold more traditional, family-oriented views.

Differential parental treatment is present in adulthood as well. In particular, Pollet et al. (2009) and Danielsbacka et al. (2011) find that parents support the family formation of daughters more than the family formation of sons. For example, they provide more informal childcare and are more likely to provide essentials, gifts and extras for the baby. This implies that there is a substitution effect from sisters: as compared to those with brothers, those with sisters get less support from their parents for their own family formation. This effect goes in the same direction as the gender-stereotyping influences from childhood. Thus, the predictions from differential parental treatment are that brothers have a negative impact on the labor market outcomes of their sibling, while sisters negatively affect

Table 7: Comparison of predictions from mechanisms and results

	Labor market		Family formation	
	Men	Women	Men	Women
Panel A: Predictions from mechanisms for the effect of a same-sex sibling				
Differential parental treatment	negative	positive	positive	negative
Competition	positive	unclear	unclear	positive
Job search network	positive	positive		
Panel B: Our results on the effect of a same-sex sibling				
Twins	positive	insignificant	positive	insignificant
Singletons	positive	positive	positive	positive

the family formation of their sibling.

In addition to the indirect influence via parents, siblings might affect each other directly. One reason for this is that siblings provide reference points to each other (Adams, 1999; Keim et al., 2009). This may lead to competition, as people do not want to lag behind their siblings (corresponding to the notion of loss aversion by Kahneman and Tversky, 1984). In line with this, Kuegler (2009) finds that individuals are less satisfied with their life if their sibling earns more than they do. Joensen and Nielsen (2018) shows that students are more likely to chose advanced math-science courses if their older sibling was exogenously induced to do so, and that this result can be explained by competition.

The gender of the sibling is relevant for two reasons. First, competition may be stronger among same-sex siblings as the reference point is more salient when the siblings are of the same sex. Indeed, Joensen and Nielsen (2018) find that their result is led by same-sex siblings (brothers in particular), and same-sex sibships are more competitive according to Conley (2000). A second reason for gender to matter is that gender differences in labor market and family formation outcomes lead to differences in reference points. Since the earnings of men are higher on average, brothers will represent a higher reference point than sisters. Similarly, as women typically marry and have kids at an earlier age, they will have higher marriage and fertility rates at each particular age. Thus, the sex of the sibling matters in two different ways: by affecting the strength of the response to the reference point, and by affecting the value of the reference point as well. For labor market outcomes, the two effects go in the same direction for men, but in opposite directions for women. For family formation, the two effects go in the same direction for women, but in opposite directions for men. Hence, the predictions from the competition mechanism are clear only for men's income and for women's family formation: same-sex siblings should have a positive effect on these.

Another way in which siblings can have a direct impact is via the use of each other's labor market networks. The sex of the sibling is relevant in this case because men and women tend to work in different industries. For example, construction and manufacturing of machinery are male dominated industries, whereas health and social work and manufacturing of wearing apparel is mainly done by female workers. Indeed, Rao and Chatterjee (2018) analyzes answers to a job search questionnaire and finds that men find a job via their brother about twice as often as via their sister, whereas the reverse holds for women. Thus, same-sex siblings might improve each other's labor market outcomes via gendered job search networks.

We reviewed three different mechanisms in this subsection that might help explain our results.[36] Now we proceed with analyzing them empirically. We will look at labor market outcomes first, and then at family formation.

We start the labor market analysis by examining whether individuals with an opposite-sex sibling choose more gender-stereotypical occupations than individuals with a same-sex sibling, as predicted by the differential parental treatment channel. We use the Swedish version of ISCO occupational codes from the registers, observed between 2001-2011, and assign each person their mode occupation. To measure how gendered each occupation is, we calculate the share of women employed within that occupation. Column 1 of Table 8 shows that the gender of the sibling has no effect on this dimension of occupational choice.[37] As compared to those with an opposite-sex sibling, those with a same-sex sibling do not choose less stereotypical occupations. Thus, this prediction of the differential parental treatment channel is not supported by our data. Next, we turn to the competition mechanism. One way in which siblings can compete is by choosing more highly paid occupations. To examine this possibility, we look at the general earnings level within each person's occupation. More specifically, we calculate the natural logarithm of the population average income within their mode occupation. Column 2 of Table 8 shows that men with brothers indeed choose higher paying occupations than men with sisters, while there is no such effect for women. These results are in line with the competition mechanism (see Table 7). Finally, we turn to the network channel. In column 3 of Table 8, we show how a same-sex sibling affects individual's average unemployment rate (over ages 25-64 observed between 1986-2011). The results show that having a same-sex sibling decreases the probability of unemployment. The estimated effect is somewhat stronger for women. In column 4, we include a dummy for working in the same industry as the sibling, and find that the coefficient of the same-sex dummy loses significance for men and decreases for women. These results are in line with the explanation that same-sex siblings use each other's job search networks more efficiently than

[36]Other potential mechanisms relate to personality traits. Some psychology studies indicate that sibling spillovers may make children more feminine/masculine if they have a sister/brother (Rust et al., 2000). However, McHale et al. (1999) points out that deidentification (i.e. the tendency of children to differentiate themselves from their siblings) works in the opposite direction. In addition, results on same-sex education suggest that opposite-sex peers reinforce gender differences in traits such as risk-aversion (e.g. Booth et al. 2014). Brenøe (2017) and Cools and Patacchini (2017) both look at personality traits and find that these cannot explain results. Given these conflicting predictions and findings, we do not discuss these mechanisms any further.

[37]We get similar results when the outcome is a binary variable of gender stereotypical occupation (based on whether the share of women is below or above the median). There are also no effects on entrepreneurship (self-employment). The only specification in which we find a small but significant effect (only for women) is when we use a binary variable for STEM occupations: women are less likely to work in such occupations if they have a brother. This result confirms the finding of Brenøe (2017) that sibling sex-composition can help to understand the choice of STEM fields in particular. It seems however that the sex of the sibling is not so relevant for choosing gendered occupations more generally.

opposite-sex siblings.

In the last four columns, we conduct a "horse race" between the competition channel and the network channel. For men, the results suggest that the effect of a same-sex sibling on income comes mainly from the competition channel. For women, the coefficient of same-sex does not respond to the inclusion of occupational income, as expected from the results of column 2. However, the coefficient drops to half of its size when unemployment is included among the controls, which suggests that the job search channel can explain a substantial part of the effect.

After the labor market outcomes, we turn to family formation. Recall that differential parental treatment predicts that sisters will have a negative effect on the family formation of their siblings. However, the competition channel predicts that sisters have a positive effect on women. Thus, the two mechanisms go in opposite directions for women, which might explain why we found no effect on women's family formation in the twin sample and only a small positive effect in the singleton sample.

As we discussed at the beginning of this subsection, one way in which differential parental treatment manifests itself is that grandparents babysit for their daughters more than for their sons. This effect is more likely to be present if the grandparents live in the same municipality as their children. We make use of register data on locational information to explore this possibility. In particular, in Table 9 we interact the same-sex dummy with an indicator of whether the individual and his/her mother lives in the same municipality when the individual is thirty years old.[38] In line with the predictions, the interaction is positive and significant for having kids and the number of kids for men. For women, the coefficients are close to zero and insignificant. Interestingly, the interaction term is positive and significant for women's probability of ever getting married, as if competition between sisters was stronger if they lived close to their mother.

[38]Municipality of residence is observed between 1968-2011 so the number of missing observations is minimized at this age for our sample. See also Table A5 for detailed descriptions of the additional variables in this section.

Table 8: Exploring occupational mechanisms

	Share of women	Ln(occ. inc.)	Unemp. avg.	Unemp. avg.	Ln(permanent income)	Ln(permanent income)	Ln(permanent income)	Ln(permanent income)
Panel 0: Men								
SS sibling	-0.000	0.005***	-0.001**	-0.001	0.007***	0.006***	0.003***	0.002***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Same industry				-0.005***				
				(0.001)				
Unemp. avg,						-1.409***		-1.175***
						(0.005)		(0.005)
Ln(occ. inc.)							0.769***	0.678***
							(0.002)	(0.002)
Observations	456,561	456,561	456,561	456,561	456,561	456,561	456,561	456,561
Outcome mean	[0.28]	[12.72]	[0.09]	[0.09]	[12.54]	[12.54]	[12.54]	[12.54]
Panel 1: Women								
SS sibling	0.000	-0.000	-0.002***	-0.001***	0.004***	0.002*	0.004***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Same industry				-0.003***				
				(0.001)				
Unemp. avg,						-1.227***		-1.037***
						(0.005)		(0.005)
Ln(occ. inc.)							0.725***	0.614***
							(0.002)	(0.002)
Observations	444,383	444,383	444,383	444,383	444,383	444,383	444,383	444,383
Outcome mean	[0.69]	[12.51]	[0.12]	[0.12]	[12.19]	[12.19]	[12.19]	[12.19]

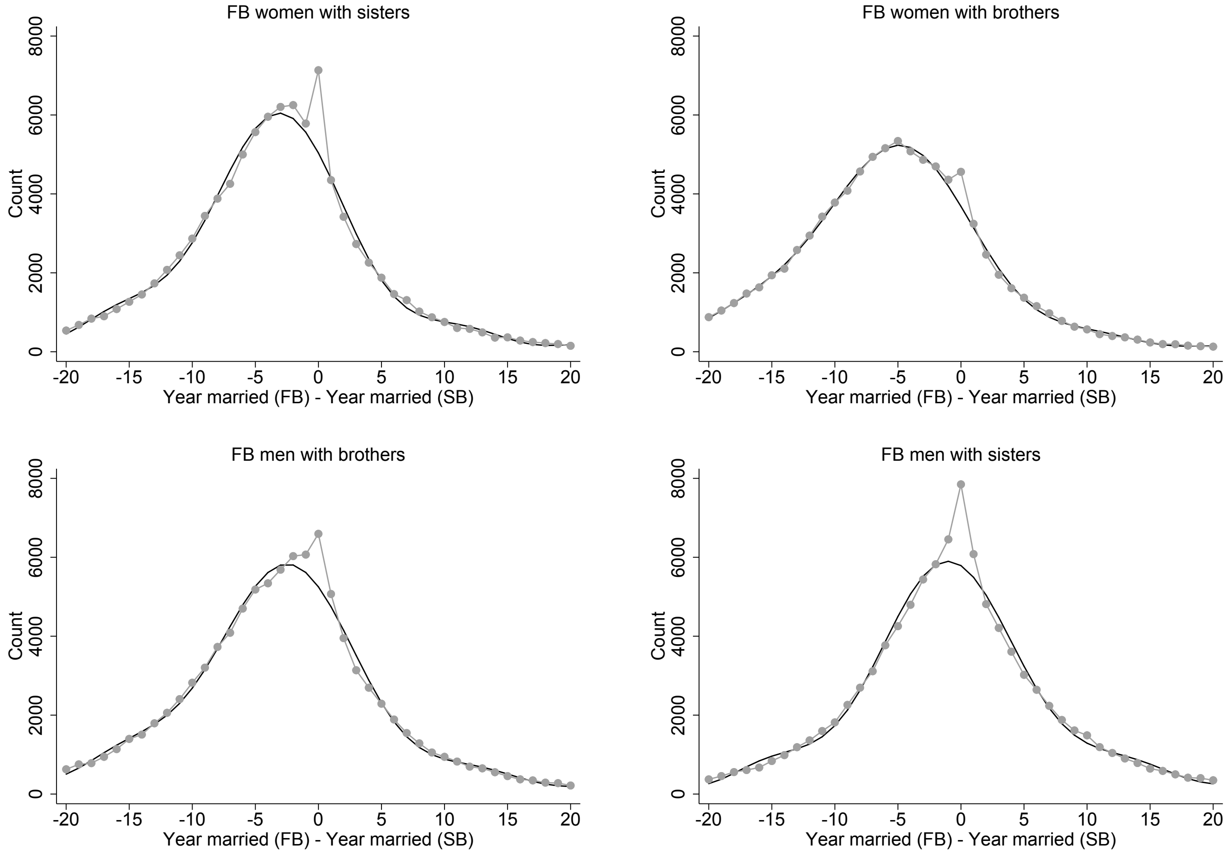
Note: The sample consists of first-born singletons born 1938-1977 whose second-born sibling was born at most 6 years later. "SS sibling" is a dummy which takes 1 if the second-born sibling is of the same sex as the first-born and 0 if the second-born sibling is of the opposite sex. "Ever married" is a dummy which takes 1 if the individual has ever been married, 0 otherwise. Permanent income is average income in years 1971-2011 (if subject is aged 25-64 in the given year). The occupational variables are described in Table A5. The sample is balanced across outcomes and control variables. All regressions control for cohort fixed effects. Robust standard are shown in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Exploring differential parental treatment in family formation

	Ever married	Any kids	Number of kids	Age at first birth
Panel 0: Men				
SS sibling	0.009***	0.004**	0.012**	-0.092***
	(0.002)	(0.002)	(0.005)	(0.024)
Same muni.*SS	0.003	0.004*	0.014**	-0.054
	(0.002)	(0.002)	(0.007)	(0.033)
Observations	499,427	499,427	499,427	399,676
Outcome mean	[0.68]	[0.80]	[1.80]	[28.85]
Panel 1: Women				
SS sibling	0.002	0.001	-0.001	-0.067***
	(0.002)	(0.001)	(0.005)	(0.021)
Same muni.*SS	0.005*	0.002	0.001	0.000
	(0.002)	(0.002)	(0.007)	(0.030)
Observations	470,214	470,214	470,214	413,492
Outcome mean	[0.75]	[0.88]	[2.00]	[26.41]

Note: The sample consists of first-born singletons born 1938-1977 whose second-born sibling was born at most 6 years later. "SS sibling" is a dummy which takes 1 if the second-born sibling is of the same sex as the first-born and 0 if the second-born sibling is of the opposite sex. The regressions include a dummy for the first-borns who, at age 30, live in the same municipality as their mother (Same muni.). The sample include those with non-missing information on municipality of residence for themselves and their mother. "Ever married" is a dummy which takes 1 if the individual has ever been married, 0 otherwise. All regressions include cohort fixed effects. Robust standard errors are shown in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Figure 1: Timing of marriage



Note: These figures plot the distributions of the following difference: calendar year the first-born (FB) got married - calendar year the second-born (SB) got married. The sample include sibling pairs where the first-born is born between 1938-1977 and where both siblings got married between 1969-2012. The gray dots are the actual counts. The black lines are the counterfactual distributions estimated by a 15-degree polynomial, taking out the effect from a dummy specified for the difference zero. The first figure is for first-born women with sisters (96,236 obs.), the second figure is for first-born women with brothers (92,798 obs.), the third figure is for first-born men with brothers (99,453 obs.) and the fourth figure is for first-born men with sisters (98,847 obs.).

With respect to the competition channel, recall that the underlying idea was that siblings represent reference points to each other. If this is true, we should see that individuals react to family formation events of their sibling by adjusting their own family formation. So far, we have only used information regarding the gender of the second-born sibling, but in order to examine if siblings react to each other's family formation events we now also consider the year the siblings get married. This analysis is of course highly explorative and cannot be given a causal interpretation. The aim of this exercise is simply to examine whether the patterns in the data are consistent with the notion of reference points and competition.

Figure 1 show distributions of the difference between the year the first-born got married and the year the second-born got married (conditional on that both got married). The gray connected dots show the actual number of sibling pairs. Observations to the left of zero are those sibling pairs where the first-born got married before the second-born while observations to the right represent the more atypical cases where the reverse is true. Although this is a selective sample, the irregularities at zero clearly stand out. Based on these graphs, siblings seem to have a tendency to marry the same year.

We illustrate the excess mass at zero by using an estimation strategy similar to those of Chetty et al. (2011) and Persson (2015). Specifically, we treat the counts as the dependent variable and the differences as the independent variable and fit a polynomial to the counts, including a dummy for the irregularity at zero. The predicted values from this model, when we exclude the contribution of the dummy, constitutes the counterfactual distributions shown in black in the figures. This exercise is mainly done for illustrative purposes but we can also obtain a crude measurement of the excess mass. We quantify the relative excess mass at zero as the difference between the actual and counterfactual count, relative to the counterfactual count. For women, this measure is 0.42 for those with sisters while it is 0.24 for those with brothers. For men, the measure is 0.26 for those with brothers and 0.36 for those with sisters. One interpretation of this pattern is that first-born siblings do not want to lag behind their younger sibling in getting married. This tendency seems to be strongest for women with sisters, which is consistent with the prediction that women react more strongly to family formation events of their sisters. Recall that the predictions were unclear for men, and indeed we do not see that men would react more strongly to their brother's marriage than to their sister's marriage.

5.2 Mechanisms related to constraints

Those who grow up with a same-sex sibling may be exposed to different kinds of constraints compared to those who grow up with an opposite-sex sibling. In this section, we focus on two potential mechanisms that are related to constraints: economies of scale and family size.

Rosenzweig and Wolpin (2000) argue that families economically benefit from having same-sex children because of hand-me-down savings, for example by reduction in child-expenditures on goods such as clothing. In other words, there might be economies of scale from having same-sex children. Therefore, a potential explanation of why same-sex siblings have better labor market outcomes is that they grew up in relatively better-off families than opposite-sex siblings.[39] The cost reductions implied by the economies of scale mechanism is expected to be particularly important for households of low socio-economic status. Thus, we explore this mechanism by examining heterogeneous effects with regard to parental education. Table 10 shows results for our main outcomes when we include interactions between same-sex and parental education categories, where having parents with low education is the reference category. Overall, the patterns do not seem to support the economies of scale explanation. The estimated effects of having a same-sex sibling do not seem to systematically differ by the socio-economic status of the parents.

[39]The effect might be more pronounced for men as recent research shows that they benefit more from additional resources such as higher school quality (Autor et al., 2016).

Table 10: Exploring economies of scale mechanism

	Years of schooling	Ln(permanent income)	Ever married	Any kids	Number of kids	Age at first birth
Panel 0: Men						
SS sibling	0.015	0.007***	0.009***	0.006***	0.022***	-0.123***
	(0.010)	(0.002)	(0.002)	(0.002)	(0.005)	(0.026)
Mixed par. educ.*SS	-0.000	0.001	0.003	0.001	-0.002	0.050
	(0.016)	(0.004)	(0.003)	(0.003)	(0.008)	(0.040)
High par. educ.*SS	-0.018	-0.001	-0.004	-0.004	-0.011	-0.038
	(0.018)	(0.004)	(0.003)	(0.003)	(0.009)	(0.042)
Missing par. educ.*SS	-0.016	0.010*	-0.000	-0.001	-0.006	-0.016
	(0.027)	(0.005)	(0.004)	(0.004)	(0.012)	(0.055)
Observations	528,606	528,606	528,606	528,606	528,606	419,902
Outcome mean	[11.93]	[12.46]	[0.70]	[0.80]	[1.78]	[28.39]
Panel 1: Women						
SS sibling	0.002	0.007***	0.003	-0.000	-0.004	-0.093***
	(0.010)	(0.002)	(0.002)	(0.001)	(0.005)	(0.023)
Mixed par. educ.*SS	-0.012	-0.004	0.002	0.004	0.005	0.066*
	(0.016)	(0.004)	(0.003)	(0.002)	(0.008)	(0.037)
High par. educ.*SS	0.005	-0.001	0.003	0.002	0.006	0.043
	(0.017)	(0.005)	(0.003)	(0.003)	(0.008)	(0.038)
Missing par. educ.*SS	-0.006	-0.001	0.004	0.007**	0.021*	-0.047
	(0.026)	(0.006)	(0.003)	(0.003)	(0.011)	(0.049)
Observations	498,741	498,741	498,741	498,741	498,741	435,395
Outcome mean	[12.27]	[12.09]	[0.76]	[0.88]	[1.91]	[26.23]

Note: The sample consists of first-born singletons born 1938-1977 whose second-born sibling was born at most 6 years later. “SS sibling” is a dummy which takes 1 if the second-born sibling is of the same sex as the first-born and 0 if the second-born sibling is of the opposite sex. “Ever married” is a dummy which takes 1 if the individual has ever been married, 0 otherwise. Permanent income is average income in years 1971-2011 (if subject is aged 25-64 in the given year). All regressions control for parental education dummies (Low par. edu. is the reference category) and cohort fixed effects. Robust standard errors are shown in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

The size of the family in which siblings grow up is also related to constraints. Theoretical models suggests that parents face a “quantity-quality trade-off” when deciding to have additional children. Because of both financial constraints (Becker and Lewis, 1973) and time constraints (Lundholm and Ohlsson, 2002), investment per child decreases with the number of children. If same-sex siblings grow up in larger families than opposite-sex siblings, the prediction from theory would be that they have worse labor market outcomes. On a more speculative note, family size could also affect the family formation of the children. On the one hand, they might be inclined towards having more children as they grew up in a bigger family themselves. On the other hand, they might see the constraints imposed by larger families better and decide to have smaller families themselves.

The family size channel is relevant if two conditions are met: 1) the sex composition of siblings affects family size and 2) family size affects the outcomes of siblings. We can empirically examine both conditions. The first condition is fulfilled for our sample of singletons: having a second-born sibling of the same sex significantly increases the number of siblings, by about 0.08 (untabulated result). This is in line with findings from previous literature (e.g. the estimate is very similar to the that of Angrist and Evans, 1998). For twins, we do not find a significant impact of sex composition on the number of siblings. This is also in line with our expectations: since twins already increase family size substantially, the sex composition of the pair is less likely to increase family size further. Recall from Section 4 that our main results are similar for male twins and singletons. This already gives and indication that family size does not drive the results on men.

A commonly used empirical strategy to examine the second condition is to use twin births as instruments for family size. Overall, these studies rarely find significant family size effects on child outcomes. For example, Black et al. (2005) find that education, earnings and teen births are all unaffected by family size in Norway. Åslund and Grönqvist (2010) reach similar conclusions about years of schooling and labor market outcomes in Sweden.

We can use the richness of the register data to employ the twin birth IV approach for the particular outcomes considered in our paper. To be consistent with our previous analysis, we focus on the impact of family size on first-born singletons in families with at least two births. Thus, we use twinning at second parity as an instrument for number of siblings. Table 11 shows the results. For men, we find no impact of family size on any of our outcomes. Family size also cannot drive our results for women. If anything, the impact of family size imply that our estimated effects of having a sister are biased towards zero for family formation outcomes. Recall that women with sisters have on average 0.08 more siblings than those with brothers. Thus, the estimated coefficients need to be multiplied with this to calculate the bias, which will therefore become very small.

Table 11: Twins at second parity as instrument for number of siblings

	Years of schooling	Ln(permanent income)	Ever married	Any kids	Number of kids	Age at first birth
Panel 0: Men						
Nr. of siblings	0.054	0.016	0.008	0.003	-0.009	0.132
	(0.051)	(0.010)	(0.008)	(0.008)	(0.023)	(0.115)
Observations	533,734	533,734	533,734	533,734	533,734	423,998
Outcome mean	[12.16]	[12.46]	[0.67]	[0.79]	[1.78]	[28.84]
Panel 1: Women						
Nr. of siblings	0.054	0.013	0.008	-0.015**	-0.069***	0.190*
	(0.048)	(0.012)	(0.008)	(0.007)	(0.022)	(0.105)
Observations	503,540	503,540	503,540	503,540	503,540	439,537
Outcome mean	[12.57]	[12.11]	[0.75]	[0.87]	[1.98]	[26.41]

Note: This table shows estimated effects of family size on our main outcomes. The sample consists of first-born singletons born 1938-1977 whose second-born sibling was born at most 6 years later. The specification is 2SLS, where number of siblings in the family is instrumented by an indicator for twins at second birth. Number of siblings include all full siblings in the first-born's family. "Ever married" is a dummy which takes 1 if the individual has ever been married, 0 otherwise. Permanent income is average income in years 1971-2011 (if subject is aged 25-64 in the given year). All regressions control for cohort fixed effects. Robust standard errors in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

6 Conclusion

In this paper we examine the impact of a sibling's gender on labor market outcomes and family formation. We study both twins and singletons by using two well-founded empirical strategies and two different data sets. In addition, we empirically explore several potential mechanisms that can explain why sibling gender matters. This comprehensive approach enables us to contribute to knowledge on the long-term effects of a sibling's gender.

We find that a same-sex sibling increases men's earnings, likelihood to form a family and number of children. These results are consistent between the twin and singleton samples. We find that a large part of the observed income premium can be attributed to occupational choice: men with brothers tend to choose higher paying occupations. We interpret this as a sign of competition between brothers. When it comes to the positive impact of brothers on men's family formation, we find that differential parental treatment may be an underlying explanation. In particular, parents have been found to support the family formation of daughters rather than family formation of sons. The existence of this type of substitution effect from sisters is consistent with our finding that the impact of a sibling's gender is more prominent for men who live close to their mother.

For women, the estimated effects of having a same-sex sibling are also positive in the singleton sample. However, we do not see these effects in the twin sample. Using the singleton data, we find that women benefit from having sisters because of shared labor market networks. If this is the main driver of the income results for singleton women, it is not a surprise that we do not find an income effect in the twin sample. Since the twins belong to an older generation, characterized by a lower female labor market participation, the twin sisters do not have the same opportunities to help each other with job search networks. For women's family formation, we suggest that competition between sisters implies a positive impact while differential parental treatment implies a negative impact. The presence of both mechanisms might explain why having a same-sex sibling has only a small effect on women's family formation outcomes in the singleton sample and no effect among twins.

The final question that can be asked is to which extent our empirical results generalize to other populations. The data that we have analyzed in this paper are tied to the Swedish context. Sweden is a country that has often been described in terms of high gender equality. Moreover, Sweden has a longstanding tradition of family friendly policies. This specific cultural and institutional context may be a determinant of the strength of the mechanisms at play. On the other hand, in our analysis we use data from different cohorts and data of samples of twins and singletons. We find that the main results are quite consistent across these cohorts and samples. This suggests that the findings might not be restricted to a specific context. Further research based on data from countries with a different cultural and institutional context is needed to shed light on this question.

Appendix

A.1 The selection bias problem in a formal framework

In Section 2 we explained verbally how a selection bias can arise when one tries to estimate the effect of an older sibling's gender on younger siblings. To show this formally, we need to consider a set of equations. We will analyze men and women separately so parameters can be different by gender. Nonetheless, for the ease of exposition we omit subscripts for own gender. We start with the equation for the latent variable Y_i^*:

$$Y_i^* = \alpha + \beta^o G_i^o + \gamma X_i + U_i \tag{2}$$

where Y_i^* is the outcome of individual i, G_i^o is the gender of the older sibling and X_i denotes observable exogenous covariates. U_i consists of other relevant variables with $E[U_i] = 0$. Parental preferences are denoted by P_i and they are included in U_i. They are relevant because parents with different preferences may raise their children differently. They are unobservable, so they have to be in U_i instead of X_i. Nonetheless, this does not lead to bias in the estimation of this equation because G_i^o and P_i are independent, due to the random assignment of gender.

The problem is that we cannot estimate the above equation because Y_i^* is a latent variable. Instead of observing Y_i^*, we observe

$$Y_i = \begin{cases} Y_i^* & if \quad S_i = 1 \\ missing & if \quad S_i = 0 \end{cases} \tag{3}$$

That is, we observe outcomes for a selected sample: only for those people who were actually born ($S_i = 1$). This selection depends on both parental preferences (P_i) and the gender of the older child:

$$S_i = f(P_i, G_i^o) \tag{4}$$

If we try to estimate the parameters using the selected sample, we need to consider the expected value of Y_i conditional on $S_i = 1$:

$$E[Y_i|S_i = 1, G_i^o, X_i] = \alpha + \beta^o G_i^o + \gamma X_i + E[U_i|S_i = 1, G_i^o, X_i] =$$

$$\alpha + \beta^o G_i^o + \gamma X_i + E[U_i|f(P_i, G_i^o) = 1, G_i^o, X_i] \tag{5}$$

The last term makes the selection bias visible. We condition on both $S_i = 1$ and G_i^o. Since selection depends on both P_i and G_i^o, this implicitly defines P_i: only certain parents will proceed to a next child after they have a girl (boy). Recall that U_i includes P_i because parental preferences may affect how children are raised. This implies that $E[U_i|f(P_i, G_i^o) = 1, G_i^o = 1, X_i] \neq E[U_i|f(P_i, G_i^o) = 1, G_i^o = 0, X_i]$, so the estimates suffer from selection bias.

A.2 What can researchers expect if they lack zygosity information

As we explained in Section 2, failing to distinguish twins by zygosity threatens identification. This is because such estimates would pick up not only the effect of the co-twin's gender, but also the impact of having a genetically identical sibling and other potential differences between MZ and DZ twins. Given the complex nature of this zygosity bias, it would be difficult to predict a priori how it would distort estimates.[40] However, using our data we can give an indication on what researchers can expect when they do not have access to zygosity information.

Table A1 reports estimates on all twins (including MZs) in SALT. In case of women, the results are very different from the unbiased estimates of Table 3. The coefficients are insignificant and close to zero for all outcomes, including education and age at first birth. For men, the results on income are similar to Table 3, while the estimates on family formation are somewhat smaller and the probability of having any kids is not significant any more.

One should be cautious when drawing conclusions from these results, as the distortion in a particular sample depends on multiple factors. For example, in our sample the share of DZ twins is almost three quarters. As MZ and DZ twins arise in different ways, their relative frequency can vary a lot from population to population or over time, which will affect the eventual magnitude of the distortion. To simulate this, we examined what happens when we weight observations by zygosity status. We found that as the relative weight of DZ twins decreases, a positive impact emerges on men's education while the results on their family formation gradually lose significance. The result on men's marriage loses significance when the share of DZ twins decreases to 55% and the result on their number of children becomes insignificant when the DZ share decreases to 35%. This is illustrated by Table A2, which shows estimation results from a weighted regression that simulates what would happen if only 35% of twins were dizygotics. As we can see, in such a case researchers would erroneously conclude that a co-twin's gender has no significant effect on any of the family formation variables.

In sum, zygosity bias distorts most of the results, but to different degrees and sometimes even in different directions. When researchers have no zygosity information, they should be most cautious when drawing conclusion about women's age at first birth and education, and men's probability of having kids.

[40] For example, how could we predict how a man or a woman would respond to having a genetically identical sibling?

Table A1: Main results on all twins (including monozygotics)

	Years of schooling	Ln(permanent income)	Ever married	Any kids	Number of kids	Age at first birth
Panel 0: Men						
SS co-twin	0.056	0.014**	0.010**	0.009	0.040**	-0.033
	(0.049)	(0.006)	(0.005)	(0.006)	(0.020)	(0.090)
Observations	17,978	17,978	17,978	17,978	17,978	14,680
Outcome mean	[10.83]	[12.44]	[0.88]	[0.82]	[1.89]	[27.87]
Panel 1: Women						
SS co-twin	-0.029	-0.004	-0.003	-0.006	0.005	-0.068
	(0.043)	(0.008)	(0.005)	(0.005)	(0.018)	(0.077)
Observations	20,099	20,099	20,099	20,099	20,099	17,453
Outcome mean	[10.82]	[11.93]	[0.89]	[0.87]	[1.98]	[25.02]

Note: The sample consists of all SALT twins (including monozygotics) born 1926-1958. "SS co-twin" is a dummy which takes 1 if the co-twin is of the same sex as the individual, and 0 if the co-twin is of the opposite sex. "Ever married" is a dummy which takes 1 if the individual has ever been married or cohabited and 0 otherwise. All regressions control for older sibling, missing sibling information and cohort fixed effects. Permanent income is average income in years 1971-2007 (if subject is aged 25-64 in the given year). The standard errors shown in parentheses are robust and clustered by the family of origin. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A2: How would results look like if only 35% of twins were dizygotic?

	Years of schooling	Ln(permanent income)	Ever married	Any kids	Number of kids	Age at first birth
Panel 0: Men						
SS co-twin	0.119**	0.016**	0.008	0.006	0.036	-0.016
	(0.057)	(0.007)	(0.006)	(0.007)	(0.022)	(0.101)
Observations	17,978	17,978	17,978	17,978	17,978	14,680
Panel 1: Women						
SS co-twin	0.053	0.006	-0.005	-0.008	0.000	0.046
	(0.049)	(0.009)	(0.005)	(0.006)	(0.020)	(0.087)
Observations	20,099	20,099	20,099	20,099	20,099	17,453

Note: The sample consists of all SALT twins (including monozygotics) born 1926-1958. In all regressions, weights are assigned to observations based on zygosity status, to simulate a situation when only 35% of twins are dizygotics. "SS co-twin" is a dummy which takes 1 if the co-twin is of the same sex as the individual, and 0 if the co-twin is of the opposite sex. "Ever married" is a dummy which takes 1 if the individual has ever been married or cohabited and 0 otherwise. All regressions control for older sibling, missing sibling information and cohort fixed effects. Permanent income is average income in years 1971-2007 (if subject is aged 25-64 in the given year). The standard errors shown in parentheses are robust and clustered by the family of origin. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

A.3 Additional tables and figures

Table A3: Robustness check: Main results for first-born singletons

	Years of schooling	Ln(permanent income	Ever married	Any kids	Number of children	Age at first birth
Panel 0: Men						
SS sibling	0.008	0.007***	0.009***	0.005***	0.018***	-0.119***
	(0.007)	(0.002)	(0.001)	(0.001)	(0.003)	(0.016)
Observations	528,606	528,606	528,606	528,606	528,606	419,902
Outcome mean	[11.93]	[12.46]	[0.70]	[0.80]	[1.78]	[28.39]
Panel 1: Women						
SS sibling	0.000	0.005***	0.005***	0.002**	0.001	-0.071***
	(0.006)	(0.002)	(0.001)	(0.001)	(0.003)	(0.014)
Observations	498,741	498,741	498,741	498,741	498,741	435,395
Outcome mean	[12.27]	[12.09]	[0.76]	[0.88]	[1.91]	[26.23]

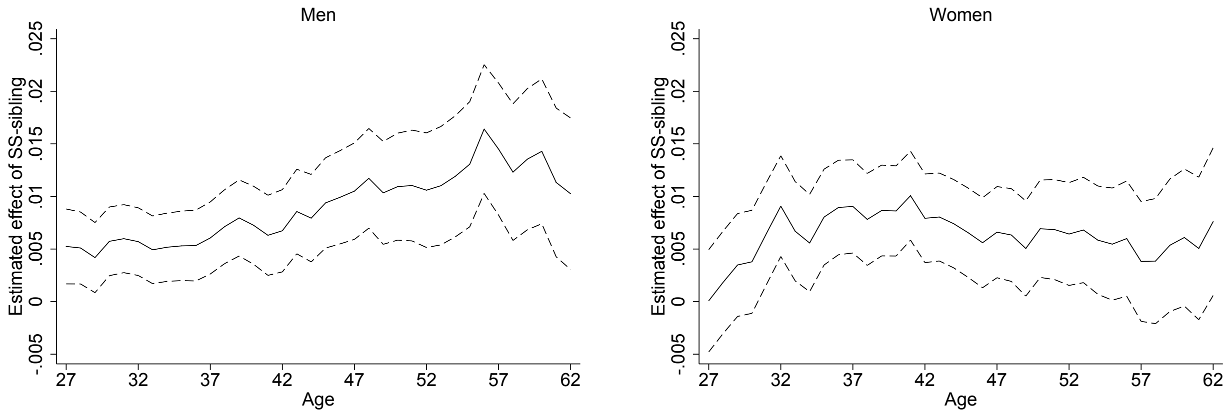
Note: The sample consists of first-born singletons born 1938-1977 whose second-born sibling was born at most 6 years later. "SS sibling" is a dummy which takes 1 if the second-born sibling is of the same sex as the first-born and 0 if the second-born sibling is of the opposite sex. "Ever married" is a dummy which takes 1 if the individual has ever been married, 0 otherwise. Permanent income is average income in years 1971-2011 (if subject is aged 25-64 in the given years). All regressions control for cohort fixed effects, age difference in months, age at first birth of mother and father fixed effects and parental education. Robust standard errors are shown in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Exploring selection bias: Results for second-born singletons

	Years of schooling	Ln(permanent income	Ever married	Any kids	Number of children	Age at first birth
Panel 0: Men						
SS sibling	0.001	0.000	0.000	-0.004***	-0.010***	-0.088***
	(0.007)	(0.002)	(0.001)	(0.001)	(0.003)	(0.017)
Observations	486,225	486,225	486,225	486,225	486,225	382,219
Outcome mean	[12.00]	[12.44]	[0.65]	[0.79]	[1.75]	[28.89]
Panel 1: Women						
SS sibling	0.008	0.005***	-0.005***	-0.001	-0.018***	0.015
	(0.007)	(0.002)	(0.001)	(0.001)	(0.003)	(0.016)
Observations	458,206	458,206	458,206	458,206	458,206	396,531
Outcome mean	[12.43]	[12.10]	[0.71]	[0.87]	[1.92]	[26.55]

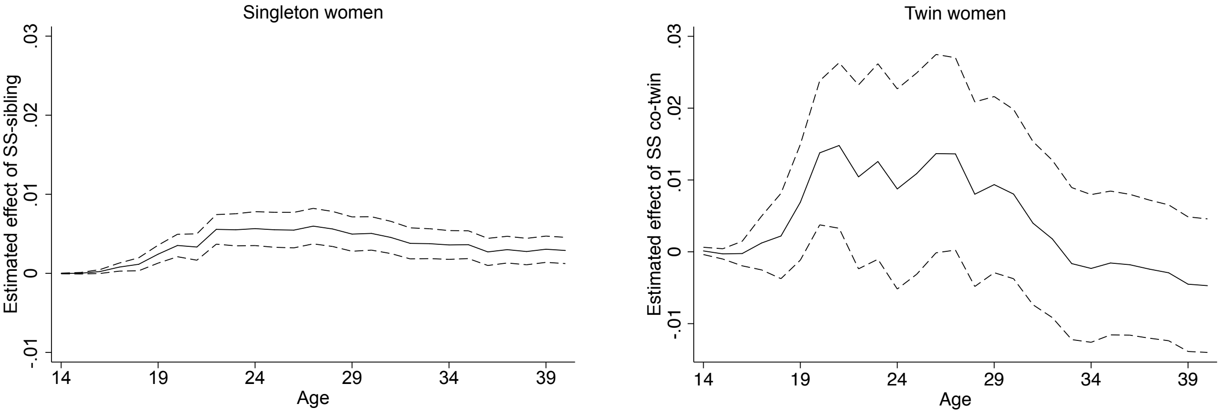
Note: The sample consists of second-born singletons born 1938-1977 whose first-born sibling was born at most 6 years earlier. "SS sibling" is a dummy which takes 1 if the first-born sibling is of the same sex as the second-born and 0 if the first-born sibling is of the opposite sex. "Ever married" is a dummy which takes 1 if the individual has ever been married, 0 otherwise. Permanent income is average income in years 1971-2011 (if subject is aged 25-64 in the given years). All regressions control for cohort fixed effects. Robust standard errors are shown in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Figure A1: Results on income at different ages, for first-born singletons



Note: The figures show results from separate regressions by age where the dependent variable is the natural logarithm of a five-year moving average of income. For example, at age 37 we use the average income over ages 35-39. The black lines show the estimated coefficients of the same-sex sibling dummy and the dotted lines show 95 percent confidence intervals. All regressions include cohort fixed effects. The first figure shows results for first-born singleton men and the second figure shows results for first-born singleton women.

Figure A2: Results on the probability of being a mother by a certain age



Note: The figures show results from separate regressions by age where the dependent variable is an indicator for having any children. The black lines show estimated coefficients of the same-sex sibling/same-sex co-twin dummy and the dotted lines show 90 percent confidence intervals. The first figure shows results for first-born singleton women and the second figure shows results for twin women. All regressions include cohort fixed effects and the regressions on the twin sample also control for older sibling and missing sibling information.

A.4 Variable definitions for analyzing potential mechanisms

Table A5: Variable definitions related to the analysis of the potential mechanisms

Variable	Data and definition
Occupation	Occupation is observed in terms of 4 digit occupational codes SSYK (Swedish version of ISCO) between years 2001-2011. Each individual's occupation in the singleton sample is the mode of observed occupations between 2001-2011.
Occupational income	We calculate the average income within each occupation using the entire Swedish population aged 25-64 between 2001-2011, then each individual in the singleton sample is assigned the natural log of the average income within their mode occupation.
Share of women in occupation	We calculate the share of women (between 0 and 1) employed in each occupation using the entire Swedish population observed between 2001-2011, then each individual in the singleton sample is assigned the share of women employed in their mode occupation.
Same industry	Industry is observed in terms of 2 digits industry codes (SNI) between year 1990-2010, there are 59 categories observed during this period (where the three largest industries, in terms of number of employed, are health and social work, education and other business activities). Each individual's industry in the singleton sample is the mode of observed industry between 1990-2010. Same industry is an indicator variable which equals 1 if the first-born and second-born singleton have the same mode industry, 0 otherwise.
Average unemployment	Unemployment status is observed between years 1986-2011, average unemployment is the fraction of years with unemployment status between ages 25-64. We observe unemployment at different age ranges for different cohorts, however we always include birth year fixed effects in all regressions to account for such differences.
Same municipality	Municipality of residence is observed between 1968-2011. We define an indicator variable which equals 1 if the individual and the mother lives in the same municipality when the individual is thirty years old, 0 otherwise.
Parental education	For each mother and father we determine whether their education is above the median years of schooling within the group of mothers/fathers who had their first child in the same year. Then we define four parental education categories: High = Both parents above median, Mixed = One parent above and the other parent equal/below median, Low = Both parents equal/below median, Missing = One or both parents' education are unobserved.

References

Abeliovich, D., Leiberman, J. R., Teuerstein, I., Levy, J., 1984. Prenatal sex diagnosis: Testosterone and fsh levels in mid-trimester amniotic fluids. Prenatal Diagnosis 4, 347–353.

Adams, B. N., 1999. Handbook of Marriage and the Family. Ch. Cross-cultural and US kinship, pp. 77–91.

Ahrenfeldt, L., Petersen, I., Johnson, W., Christensen, K., 2015. Academic performance of opposite-sex and same-sex twins in adolescence: A danish national cohort study. Hormones and behavior 69, 123–131.

Altonji, J. G., Cattan, S., Ware, I., May 2016. Identifying Sibling Influence on Teenage Substance Use. Journal of Human Resources, 0714–6474R1.

Andersson, G., Hank, K., Rønsen, M., Vikat, A., 2006. Gendering family composition: Sex preferences for children and childbearing behavior in the nordic countries. Demography 43 (2), 255–267.

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

Angrist, J. D., Pischke, J.-S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: Princeton University Press.

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

Autor, D., Figlio, D., Karbownik, K., Roth, J., Wasserman, M., May 2016. School Quality and the Gender Gap in Educational Achievement. American Economic Review 106 (5), 289–295.

Becker, G. S., Lewis, H. G., Mar. 1973. On the Interaction between the Quantity and Quality of Children. Journal of Political Economy 81 (2, Part 2), S279–S288.

Benirschke, K., Burton, G. J., Baergen, R. N., 2012. Multiple pregnancies. In: Pathology of the Human Placenta. Springer.

Björklund, A., Jäntti, M., Lindquist, M. J., 2009. Family background and income during the rise of the welfare state: Brother correlations in income for swedish men born 1932-1968. Journal of Public Economics 93 (5), 671–680.

Black, S. E., Breining, S., Figlio, D. N., Guryan, J., Karbownik, K., Nielsen, H. S., Roth, J., Simonsen, M., Jan. 2017. Sibling Spillovers. Working Paper 23062, National Bureau of Economic Research.

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

Booth, A., Cardona-Sosa, L., Nolen, P., 2014. Gender differences in risk aversion: do single-sex environments affect their development? Journal of economic behavior and organization 99, 126–154.

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

Breining, S., Doyle, J., 2015. Birth order and parental investment for boys vs. girls: Evidence from the genetic lottery. PhD thesis Chapter 1, Aarhus University.

Breining, S. N., Sep. 2014. The presence of ADHD: Spillovers between siblings. Economics Letters 124 (3), 469–473.

Breining, S. N., Joseph J. Doyle, J., Figlio, D. N., Karbownik, K., Roth, J., Jan. 2017. Birth Order and Delinquency: Evidence from Denmark and Florida. Working Paper 23038, National Bureau of Economic Research.

Brenøe, A. A., 2017. Sibling Gender Composition and Participation in STEM Education. Working Paper.

Butcher, K. F., Case, A., 1994. The effect of sibling sex composition on women's education and earnings. The Quarterly Journal of Economics 109 (3), 531–563.

Chen, S. H., Chen, Y.-C., Liu, J.-T., 2017. The impact of family composition on educational achievement. Journal of Human Resources forthcoming.

Chetty, R., Friedman, J. N., Olsen, T., Pistaferri, L., May 2011. Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. The Quarterly Journal of Economics 126 (2), 749–804.

Cohen-Bendahan, C. C., van de Beek, C., Berenbaum, S. A., 2005a. Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. Neuroscience and Biobehavioral Reviews 29, 353–384.

Cohen-Bendahan, C. C. C., van Goozen, S. H. M., Buitelaar, J. K., Cohen-Kettenis, P. T., 2005b. Maternal serum steroid levels are unrelated to fetal sex: A study in twin pregnancies. Twin Research and Human Genetics 8 (2), 173–177.

Conley, D., Sep. 2000. Sibship Sex Composition: Effects on Educational Attainment. Social Science Research 29 (3), 441–457.

Cools, A., Patacchini, E., 2017. Sibling Gender Composition and Women's Wages. Working Paper.

Dahl, G. B., Løken, K. V., Mogstad, M., 2014. Peer effects in program participation. American Economic Review 104, 2049–2074.

Danielsbacka, M., Tanskanen, A. O., Jokela, M., Rotkirch, A., 2011. Grandparental child care in europe: Evidence for preferential investment in more certain kin. Evolutionary Psychology 9 (1), 3–24.

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

Elston, R. C., Olson, J. M., Palmer, L., 2002. Biostatistical Genetics and Genetic Epidemiology. John Wiley and Sons.

Even, M. D., Dhar, M. G., vom Saal, F. S., 1992. Transport of steroids between fetuses via amniotic fluid in relation to the intrauterine position phenomenon in rats. Journal of Reproduction and Fertility 96, 709–716.

Fellman, J., Eriksson, A. W., 2006. Weinberg's differential rule reconsidered. Human Biology 78, 253–275.

Gielen, A. C., Holmes, J., Myers, C., 2016. Prenatal testosterone and the earnings of men and women. Journal of Human Resources 51 (1), 30–61.

Glass, A., Klein, T., 1981. Changes in maternal serum total and free androgen levels in early pregnancy: lack of correlation with fetal sex. American Journal of Obstetrics and Gynecology 140, 656–660.

Gray, R. H., Simpson, J. L., Bitto, A. C., Queenan, J. T., Li, C., Kambic, R. T., Perez, A., Mena, P., Barbato, M., Stevenson, W., Jennings, V., 1998. Sex ratio associated with timing of insemination and length of the follicular phase in planned and unplanned pregnancies during the use of natural family planning. Human Reproduction 13 (5), 1397–1400.

Hank, K., Kohler, H.-P., 2000. Gender preferences for children in europe: Empirical results from 17 ffs countries. Demographic Research 2.

Hansen, D., Møller, H., Olsen, J., 1999. Severe periconceptional life events and the sex ratio in offspring: follow up study based on five national registers. BMJ 319 (7209), 548–549.
URL `http://www.bmj.com/content/319/7209/548`

Hauser, R. M., Kuo, H.-H. D., 1998. Does the gender composition of sibships affect women's educational attainment? The Journal of Human Resources 33 (3), 644–657.

Heckman, J. J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.

Husby, H., Holm, N., Gernow, A., Thomsen, S., K, K. K., Gürtler, H., 1991. Zygosity, placental membranes and weinberg's rule in a danish consecutive twin series. Acta Geneticae Medicae et Gemellologiae 40, 147–152.

James, W. H., 1979. Is weinberg's differential rule valid? Acta Geneticae Medicae et Gemellologiae 28 (1), 69–71.

Jayachandran, S., Pande, R., 2017. Why are indian children so short? the role of birth order and son preference. American Economic Review 107 (9), 2600–2629.

Joensen, J. S., Nielsen, H. S., Jan. 2018. Spillovers in education choice. Journal of Public Economics 157, 158–183.

Kaestner, R., 1997. Are brothers really better? sibling sex composition and educational achievement revisited. The Journal of Human Resources 32 (2), 250–284.

Kahneman, D., Tversky, A., 1984. Choices, values and frames. American Psychologist 39 (4), 341–350.

Keim, S., Klärner, A., Bernardi, L., 2009. Who is relevant? exploring fertility relevant social networks. MPIDR Working paper WP 2009-001.

Kuegler, A., 2009. A curse of comparison? evidence on reference groups for relative income concerns. Policy Research Working Paper 4820.

Lichtenstein, P., Faire, U. D., Floderus, B., Svartengren, M., Svedberg, P., Pedersen, N. L., 2002. The swedish twin registry: a unique resource for clinical, epidemiological and genetic studies. Journal of Internal Medicine 252, 184–205.

Lundholm, M., Ohlsson, H., Aug. 2002. Who takes care of the children? The quantity-quality model revisited. Journal of Population Economics 15 (3), 455–461.

McHale, S. M., Crouter, A. C., Tucker, C. J., 1999. Family context and gender role socialization in middle childhood: Comparing girls to boys and sisters to brothers. Child development 70 (4), 990–1004.

McHale, S. M., Crouter, A. C., Whiteman, S. D., Feb. 2003. The Family Contexts of Gender Development in Childhood and Adolescence. Social Development 12 (1), 125–148.

Meulenberg, P. M. M., Hofman, J. A., 1991. Maternal testosterone and fetal sex. The Journal of Steroid Biochemistry and Molecular Biology 39 (1), 51–54.

Miller, E. M., 1994. Prenatal sex hormone transfer: A reason to study opposite-sex twins. Personality and Individual Differences 17, 511–529.

Mizuno, R., Aug. 2000. The male/female ratio of fetal deaths and births in Japan. The Lancet 356 (9231), 738–739.

Morduch, J., 2000. Sibling rivalry in africa. The American Economic Review 90 (2), 405–409, papers and Proceedings of the One Hundred Twelfth Annual Meeting of the American Economic Association.

Nicoletti, C., Rabe, B., 2014. Sibling spillover effects in school achievement. IZA Discussion Paper IZA DP No. 8615.

Oláh, L., E.M. Bernhardt, E., 2008. Sweden: Combining childbearing and gender equality. Demographic Research 19, 1105–1144.

Parish, W. L., Willis, R. J., 1993. Daughters, education, and family budgets taiwan experiences. The Journal of Human Resources 28 (4), 863–898, special Issue: Symposium on Investments in Women's Human Capital and Development.

Persson, P., 2015. Social insurance and the marriage market. Working Paper 2015:6, IFAU - Institute for Evaluation of Labour Market and Education Policy.

Pettersson-Lidbom, P., Thoursie, P. S., Vlachos, J., 2008. Does child gender affect sibling and family outcomes? Unpublished draft.

Phillips, D. I. W., 1993. Twin studies in medical research: can they tell us whether diseases are genetically determined? The Lancet 341, 1008–1009.

Pollet, T. V., Nelissen, M., Nettle, D., 2009. Lineage based differences in grandparental investment: evidence from a large british cohort study. Journal of Biosocial Science 41, 355–379.

Rao, N., Chatterjee, T., Mar. 2018. Sibling gender and wage differences. Applied Economics 50 (15), 1725–1745.

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., Jackson, R. B., 2010. Biology - Campbell, 9th Edition. Benjamin Cummings.

Rodeck, C. H., Gill, D., Rosenberg, D. A., Collins, W. P., 1985. Testosterone levels in midtrimester maternal and fetal plasma and amniotic fluid. Prenatal Diagnosis 5, 175–181.

Rosenzweig, M. R., Wolpin, K. I., Dec. 2000. Natural "Natural Experiments" in Economics. Journal of Economic Literature 38 (4), 827–874.

Rust, J., Golombok, S., Hines, M., Johnston, K., Golding, J., 2000. The role of brothers and sisters in the gender development of preschool children. Journal of Experimental Child Psychology 77, 292–303.

Ryan, B., Vandenbergh, J., 2002. Intrauterine position effects. Neuroscience and Biobehavioral Reviews 26, 665–678.

Singer, E., van Hoewyk, J., Maher, M. P., 2000. Experiments with incentives in telephone surveys. Public Opinion Quarterly 64, 171–188.

Swedish Institute, 1994. Family planning in sweden. fact sheets on sweden. Tech. Rep. ISSN-1101-6124.

Tapp, A. L., Maybery, M. T., Whitehouse, A. J., 2011. Evaluating the twin testosterone transfer hypothesis: A review of the empirical evidence. Hormones and Behavior 60, 713–722.

Toriola, A., Vääräsmäki, M., Lehtinen, M., A, A. Z.-J., Lundin, E., Rodgers, K., Lakso, H., Chen, T., Schock, H., Hallmans, G., Pukkala, E., Toniolo, P., Grankvist, K., Surcel, H., Lukanova, A., 2011. Determinants of maternal sex steroids during the first half of pregnancy. Obsterics and Gynecology 118 (5).

van de Beek, C., Thijssen, J. H., Cohen-Kettenis, P. T., van Goozen, S. H., Buitelaar, J. K., 2004. Relationships between sex hormones assessed in amniotic fluid, and maternal and umbilical cord serum: What is the best source of information to investigate the effects of fetal hormonal exposure? Hormones and Behavior 46, 663–669.

Vlietinck, R., C, D., R, D., den Berghe H, V., M, T., 1988. The validity of weinberg's rule in the east flanders prospective twin survey (efpts). Acta Geneticae Medicae et Gemellologiae 37, 137–141.

vom Saal, F. S., Bronson, F. H., 1980. Sexual characteristics of adult female mice are correlated with their blood testosterone levels during prenatal development. Science 208, 597–599.

Wilcox, A. J., Weinberg, C. R., Baird, D. D., 1995. Timing of sexual intercourse in relation to ovulation. The New England Journal of Medicine 333 (23), 1517–1521.