NBER WORKING PAPER SERIES

ESTIMATING CAUSAL EFFECTS OF FERTILITY ON LIFE COURSE OUTCOMES: EVIDENCE USING A DYADIC GENETIC INSTRUMENTAL VARIABLE APPROACH

Boyan Zheng
Qiongshi Lu
Jason Fletcher

Working Paper 30955
http://www.nber.org/papers/w30955

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2023

This research uses data from Health and Retirement Study (HRS). HRS is funded by NIA (grant number NIA U01AG009740) and is conducted by the University of Michigan. No direct support was received from grant U01AG009740 for this analysis. The authors gratefully acknowledge use of the facilities of the Center for Demography of Health and Aging at the University of Wisconsin-Madison, funded by NIA Center Grant P30 AG017266. We thank members of the Social Genomics Working Group at University of Wisconsin and the IGSS Conference for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Estimating Causal Effects of Fertility on Life Course Outcomes: Evidence Using A Dyadic Genetic Instrumental Variable Approach

Boyan Zheng, Qiongshi Lu, and Jason Fletcher

NBER Working Paper No. 30955
February 2023
JEL No. J10,J13,J22,J29

ABSTRACT

The causal effects of fertility are a central focus in the social sciences, but the analysis is challenged by the endogeneity of fertility choices. Earlier work has proposed several "natural experiments" from twin births or gender composition of earlier births to assess whether having more children affects adults' outcomes, though there are limitations to using rare (twins) and weak (gender composition) instrumental variables for fertility. This paper proposes a new "natural experiment" approach to assessing the causal effects of fertility by measuring the combination of couples' genetics in predicting fertility—a dyadic genetic instrumental variable, where the key idea (exclusion restriction) is that the interactions of the couple's genetics that shift the likelihood of fertility is unknown to the couples. We use a nationally representative sample of couples to examine the long-lasting effects of fertility on older adults' life outcomes, including labor market outcomes, personality traits, and subjective wellbeing. We find that fertility reduces females' extraversion and years of working and some evidence indicates that fertility reduces both males' and females' lifetime number of jobs worked.

Boyan Zheng
University of Wisconsin-Madison
bzheng9@wisc.edu

Qiongshi Lu
University of Wisconsin-Madison
425 Henry Mall
#2104, Biotechnology Center
Madison, WI 53706
qlu@biostat.wisc.edu

Jason Fletcher
University of Wisconsin-Madison
La Follette School of Public Affairs
1225 Observatory Drive
Madison, WI 53706
and NBER
jfletcher@lafollette.wisc.edu

Introduction

The causal effects of fertility are a central focus in the social sciences, but its identification is challenged by the endogeneity of fertility behavior. Fertility is negatively selected by socioeconomic background, and decisions on childbearing are co-determined with other life choices (Balbo, Billari, and Mills 2013, Lundberg and Rose 2000). To address these issues, previous research often uses instrumental variables to overcome the endogeneity problem. The two most popular instrumental variables for fertility are multiple births (Twin IV), and the gender composition of the first and second born (Sex-Mix IV). Researchers have used the two IVs to identify the causal effect of fertility on parents' labor force participation and earnings outcomes[1] (e.g., Angrist and Evans 1998; Cools, Markussen, and Strøm 2017; Jacobsen, Pearce, and Rosenbloom 1999), and generally reported that mothers, and not fathers, labor market outcomes are negatively affected by fertility.

The existing literature has several limitations. One limitation of both the Twin IV and Sex-Mix IV is that the approaches are restricted to higher-order births and cannot estimate the effects of having one child vs. no children. For the Twin-IV, additional limitations include the relative rarity of twin births and the increasing evidence that twin births are non-random and instead follow a social gradient (Bhalotra and Clarke 2019). For the Sex-Mix IV, the influence of having two same sex children on having a third child is relatively small (i.e. a weak instrument) and the estimates may elicit the effect only on those induced to have a third child due to strong gender preferences that may not reflect the average effect of fertility (i.e. a local average treatment effect). Alternative instruments may be helpful to advance the existing knowledge concerning the causal effects of fertility.

Motivated by these limitations and recent developments in the sociogenomic study of human fertility (Mills, Barban, and Tropf 2018), we propose a novel dyadic genetic instrumental variable (DGIV) based on newly available genetic data in the Health and Retirement Study. Emerging research has shown that fertility has a large and measurable genetic component. We leverage this new information by controlling for the genetic "main" effects for each spouse and propose that the *interaction* of the measures of the genetic dispositions toward fertility (i.e., polygenic scores, or PGS) can be used as an instrument for (couple level) fertility. Intuitively, we assume that spouses select one another based on traits related to levels of fertility but not on how these levels combine (interact) to predict fertility at the couple level.

In this paper, we first review the existing IV literature and discuss the limitations of the typically used Sex-mix and Twin IVs. Next, we formally explain our proposed IV and state the assumptions that support the validity of our instruments. Finally, we apply our proposed IV to a sample of 3,282 couples in the Health and Retirement Study (HRS). We analyze the causal effects of life course fertility on older adults' labor market outcomes, including work history, income, and wealth, to compare the consistency of our results with the previous IV literature. We also investigate the causal effects of fertility on parental personality traits and subjective wellbeing to extend the study of fertility and non-labor market outcomes using a causal framework. This paper contributes to the demographic literature on fertility by providing both innovative instrumental variable methodologies and new empirical findings.

Previous Literature

Theoretical Perspectives on the Effects of Fertility on Parents

Most studies on the economic consequences of fertility on parents are driven by Becker's (1993) household specialization theory. The theory posits that household members specialize in either housework (including childbearing) or labor market work in order to maximize overall production, which is the sum of housework production, labor market production, and utility from children. Due to socialization and biological differences, females are often assumed to be more efficient in housework, whereas males have comparative advantages in labor market work. Thus, the theory predicts that higher fertility causally reduces females' labor market participation. Meanwhile, as mothers specialize in housework after the birth of children, males or fathers are expected to allocate more time in the labor market to maintain or maximize the labor market production of the family. Thus, the household specialization theory also hypothesizes a positive effect of fertility on males' labor force participation.

Social science theories also indicate that fertility affects men's and women's broader outcomes, although the direction of the effects is ambiguous. One set of outcomes that receives focus is subjective wellbeing (SWB). Life course theory suggests the influences of children on parents as multifaceted and time-dependent. On the one hand, children are one of the most important social ties to parents. At younger ages, children provide a meaning of life and joy to parents, and they also serve as a source of social control that reduces parents' unhealthy and risky behavior (Umberson and Gove 1989; Umberson, Pudrovska, and Reczek 2010). When children grow up, they become a source of social support that benefits their older parents both emotionally and materially (Seltzer and Bianchi 2013; Umberson, et al. 2010; Umberson, Crosone, and Reczek 2013). However, rearing young children also demands significant mental and material resources from parents. These demands may cause significant parenting pressure

and reduce parental wellbeing (Nomaguchi and Milkie 2020; Umberson and Gove 1989).

Finally, psychological theories indicate that childbearing also changes parents' personality traits (Jokela et al. 2006). From the perspective of social investment theory (Roberts, Wood, and Smith 2005), engaging with social institutions and social roles, such as parenthood, leads to maturity in personality traits. Thus, this theory foresees a positive relationship between fertility and more stable, "positive" personalities. However, from a life course perspective, major life events such as childbirth have complex effects on personality traits. The potential disruptive effects of childbirth on living arrangements and the burden of parenting may lead to negative changes in personality as well (Hutteman et al. 2014, Fletcher and Padron 2016; Specht et al. 2011).

Causal Evidence on the Consequences of Fertility on Parents

Because reproductive behavior is endogenous to a number of individual and institutional factors, such as socioeconomic status, childhood family structure, personality, and social policies (Balbo et al. 2013; Guzzo and Hayford 2020), it is necessary to utilize statistical methods to improve the causal confidence of the research. Some studies used experimental designs to show that maternal identity negatively affects the evaluation of women's work performances (Benard and Correll 2010; Correll, et al., 2007), but most of the studies tend to use instrumental variables (IV) to estimate causal effects of fertility on the realized labor market outcomes, particularly labor income and employment status. The use of the IV method is motivated by the canonical studies by Rosenzweig and Wolfin (1980) and Angrist and Evans (1998). The two studies proposed and popularized two IVs for fertility: Sex-mix IV and Twin-IV. Sex-mix IV uses the gender composition of the first two children as the instrumental

variable for fertility. It assumes that parents would like to achieve a balanced gender ratio, and thus the biological randomness of the same-gender of the first two children leads to an increase in the likelihood of the third child. Twin IV uses multiple second births as the instrumental variable. It assumes that the birth of twins is random, and thus the birth of twins in the second parity leads to an exogenous increase in the total number of children. Other IVs include infertility shocks (Agüero and Marks 2008; Agüero and Marks 2011) and in vitro fertilization (IVF; Lundborg, Plug, and Rasmussen 2017), though these have been used less frequently[1].

With the proposed IVs, Angrist and Evans (1998) analyzed the 1980 and 1990 US census data and found that having an additional child causally reduces females' work time, employment rate, and labor income, but no effects on men were observed. Later studies using one or both of the IVs generally replicated Angrist and Evans's (1998) findings and extended their conclusions to other contexts such as Latin America, Europe, and East Asia (Bronars and Grogger 1994; Cáceres-Delpiano 2012; Chun and Oh 2002; Gruce and Galiani 2007; Daouli, Demousis, and Giannakopoulos 2009; Jacobsen, Pearce, and Rosebloom 1999). Whereas previous studies tend to rely on cross-sectional data, Cools, Markussen, and Strøm (2017) used Sex-mix IV and Norwegian administrative data to analyze the long-term effects of fertility on labor market outcomes. They found that the negative effects of fertility on females' work time and labor income are insignificant after the 30s for females without college education. Similar to previous studies, Cools et al. (2017) observed no effects of fertility on men. The findings based on Sex-mix IV and Twin IV are rich and consistent, but the validity and utility of the IVs have been questioned by more recent literature. We will discuss these issues in the next section.

[1] Another relevant literature uses policies as instruments for fertility (e.g. Geyer et al. 2015)

Compared with the rich literature on labor market outcomes, causal studies of fertility and a broader set of (non-labor market) outcomes are rare. Existing observational studies of the United States and European countries tend to conclude that number of children is negatively related to parental SWB during early adulthood, but the association disappears or becomes positive after middle adulthood (Deaton and Stone 2012; Stanca 2012; Glass, Simon, and Andersson 2016; Margolis and Myrskylä 2011; Umberson, Pudrovska, and Reczek 2010). Although studies suggest that mothers take the more burdensome tasks in childbearing and are subject to stronger parenting stress (Musick, Meier, and Flood 2016), evidence is mixed regarding whether there are gender differences in the association between fertility and SWB. Some research found no gender differences (Glass et al. 2016; Margolis and Myrskylä 2011), others report that men's happiness is positively related to the number of children (Nelson-Coffey et al. 2019).[2]

Finally, the association between parenthood and personality traits has shown mixed evidence, with little causal analysis available. Using a Finnish sample, Jekola et al. (2007) reported that having children predicted higher emotionality, a personality trait related to negative outcomes, and higher emotional stability only among men, lending partial support to the social investment theory. However, later studies report having children to be related to a decrease or no change in emotional stability, conscientiousness, and agreeableness (Hutteman

[2] In contrast to the previous observational research, Priebe (2020) applied Twin IV and Sex-mix IV to a large sample of parents in developing countries. The author reports that although OLS results show that an additional child is negatively related to happiness, IV-based results show that an additional child significantly increases parents' happiness. Although results in developed countries and developing countries may not be comparable, contrasts between Priebe's OLS and IV results still call existing observational studies into question and suggest further analysis using causal designs is needed.

et al. 2014; Scheppingen et al. 2016; Specht et al. 2011). South American Surveys asking for fertility intention also suggest the existence of mothers who prefer an unequal child sex ratio even when they have had one boy and one girl, suggesting the existence of defier groups (Clément 2017). The limited amount of literature and lack of causal focus indicates a need for more studies.

Limitations of the Sex-mix IV and Twin IV

Despite their significant contributions, the Sex-mix IV and Twin IV approaches have some limitations in terms of methodological design. First, Twin IV and Sibling IV are by design restricted to adults with at least two children. The identified treatment effects may not be generalized to childless or one-child parents. Alternatively, infertility (Agüero and Marks 2008; Agüero and Marks 2011) and IVF instruments (Lundborg, Plug, and Rasmussen 2017) can be applied to first-order births. However, although infertility is argued to be approximately random (Agüero and Marks 2008), its exogeneity is still challenged by some more recent findings that infertility is associated with marital status, age, health insurance, and graduate-level education (Louis et al. 2013; Thoma et al. 2013). Because IVF is defined as the success of the first IVF treatment, it is applicable only to mothers who have ever received any IVF treatment and since the use of IVF is selective, concerns remain about the generalizability of the results.

In addition, when interpreted using the Local Average Treatment Effect (LATE) framework, Sex-mix IV measures the causal effects of fertility on the complier groups who wish to balance children's sex composition by having a higher-order birth. It also assumes the absence of the defiers group who have strong sex preferences so that they want to have more

children when the sex composition is already balanced. Although these assumptions may hold in Western contexts, they may not be applicable to cultural contexts with gender preferences for births (Chun and Oh 2002). Rosenzweig and Wolpin (2000) also suggested that child sex composition may have economics of scale effects. For example, same-sex children may share their clothes to reduce expenditures, which allows more investment in education. Indeed, recent studies have provided support for the validity of sex-mix IV. For example, Huber (2010) did not find economics of scale effects using US Census data, and Clément (2017) theorized that LATE can be identified when the defier group is small or the LATE on defiers has the same sign as LATE on compliers (i.e., although IV does not work as the same on defiers, treatments still affect them). But these concerns still suggest a need for more IVs to solidify our knowledge (Black et al. 2022).

Twin IV, on the other hand, seems to clearer claim to random variation. The birth of twins was thought to be random, and its positive effects on the number of children unrelated to other family and individual characteristics, although the parents need to desire a second parity in order to be eligible for the analysis. However, recent studies have challenged the exogeneity of the Twin IV. For example, Bhalotra and Clarke (2019) found that in the US, better maternal health and higher maternal educational attainments are significant predictors of twin births. The chances of twin births can also be significantly increased by the use of fertility treatments, such as IVF.[3]

[3] Braakmann and Wildman (2015) also reported that the use of fertility treatments is associated with a 10% increase in the likelihood of twin births. Although other studies proposed that monozygotic twins can be more strictly exogenous (Farbmacher, Guber, and Vikström 2016), such information is usually unavailable in large demographic datasets.

The Present Study

As reviewed in the previous section, the existing literature concerning the causal effects of fertility has some limitations. Empirically, existing studies tend to focus on labor market outcomes, but the effects of childbearing on psychological outcomes are also useful knowledge that demands causal analysis. Although a few studies have applied causal methods to study the impacts of fertility on non-labor market outcomes (Priebe 2020), the literature is very limited compared to the study of labor market outcomes. Finally, it is worth noting that existing studies tend to rely on working-age adults, whereas few studies consider the causal effect of fertility on older adults and therefore do not examine full life course effects. Methodologically, the issues with the generalizability and validity of Sex-mix IV and Twin-IV imply the need to seek new IVs to improve the robustness of our knowledge of fertility's consequences.

To address these limitations, this paper contributes to the current literature in three respects. First, we propose a new dyadic genetic IV (DGIV). Specifically, we utilize the interaction of spousal genetic predispositions for fertility as an instrument for realized fertility. The proposed DGIV does not impose sample restrictions on parents with multiple births. We provide formal statements of identification assumptions and interpretations based on LATE in the appendix. Second, we focus on the effects of fertility on older adults, which allows us to measure the life-course effects of fertility on labor market participation. Third, we extend the use of the proposed IV to investigate the impacts of fertility on personality and mental health outcomes. These analyses will contribute to the causal knowledge of the impacts of fertility on a broader set of outcomes.

Data and Empirical Method

Sample Construction

We use data from the Health and Retirement Study (HRS). HRS is a biannual and nationally representative longitudinal survey of older adults in the United States. HRS surveyed about 20,000 respondents each wave and has so far included about 45,000 unique older adults across all waves. Beginning in 1992, HRS has documented rich data for older adults' financial conditions, labor force participation, and fertility history. Since 2006, HRS started collecting genetic data of HRS respondents and their co-residential spouses. A Psychosocial and Lifestyle Questionnaire (PLQ) was introduced in the same year, which collects information about respondents' psychological traits and non-cognitive skills. The recent advancements in HRS's data collections provide unique opportunities to use genetics to identify the effects of fertility on non-labor market outcomes. We restrict the analytical sample to HRS couples. A respondent who is identified with two or more different spouses in HRS is not included in the sample. Because polygenic scores (PGS) predicted from the European-ancestry-based genome-wide association study (GWAS) are less reliable among the non-European population (Martin et al. 2017), we further restrict the HRS couples to be both White or European Ancestry (defined by HRS's genetic data; for detail, see Ware et al. 2018). After excluding cases with missing values, we obtained an analytical sample of 6,564 White older adults or 3,282 unique couples. Using these cases, we construct our measures (see the next section) using all person-wave observations over 50 years old so that our analysis focuses on older adults. Specific sample sizes vary by the outcome. In Appendix A, we outline our procedure for sample construction.

Measures

Our DGIV is the interaction term of each spouse's genetic propensity for fertility (details

will be given in the next section). To construct our GIV, we use a PGS for the number of children ever born (NEB) to operationalize "genetic predisposition". We use the officially released NEB PGS from the HRS team (Ware et al. 2018), which is constructed based on Barban et al. (2016)[3]. The treatment variable of our study is fertility, defined as the number of biological children for the couple. We use the wife's (or female partner if not married) reported fertility as the measure of the couple's fertility whenever the data is available. If the wife's fertility information is unavailable, the husband's data are used.

We considered two sets of life course outcomes. The first set is labor market outcomes, which include work history and family finances. Work history measures include: (1) ever worked: a dummy variable indicating whether a respondent has ever worked on at least one job, (2) total years worked: a continuous measure of the total number of years that a respondent worked for, and (3) the number of jobs worked: a continuous measure of the total number of jobs that a respondent worked on. In the HRS, work history is collected in each wave and is thus a time-varying measure. We use the last available report of work history to maximize the coverage over the life course. Family finance measures include respondents' long-term income, long-term wealth, and long-term earnings. They are constructed as the mean of the available waves of household income, household wealth, and individual earning data adjusted by the 2016 consumer price index (CPI). We use the imputation values provided by Rand HRS so that the proportion of missingness is zero. Because of the negative values of household wealth, we used an inverse hyperbolic sine transformation on all three variables. The inverse hyperbolic sine has a similar property to the natural log transformation so that coefficients in the model can be interpreted as a semi-elasticity or elasticity (Bellemare and Wichman 2020). Note that

work history measures are constructed using all person-wave observations over 50, but financial outcomes are constructed using all available observations.

The second set of outcomes includes personality traits and subjective wellbeing outcomes. The personality traits are operationalized as the Big-Five personality traits, defined as five composite scales that capture dimensions of human personality in a continuum (Goldberg 1990). The five scales are openness, conscientiousness, extraversion, agreeableness, and neuroticism. We created HRS Big-Five scales using items in PLQ. The yielded scales range from 1 to 4. Following the previous literature about childlessness and older adults' subjective wellbeing (Umberson et al. 2010; Nomaguchi and Milkie 2020), we consider three measures of subjective wellbeing, including depression, life satisfaction, and loneliness. For depression, we used the CESD scale provided by Rand HRS, which ranges from 0 to 8. For life satisfaction, we used the 7-point scale of Diener's measure of life satisfaction (Diener et al. 1985), which ranges from 1 to 7. The scale is available in PLQ since 2008. The loneliness variable is constructed using the 3-item adapted UCLA-R loneliness scale (Hughes et al. 2004) and ranges from 0 to 6. The scale is available since 2006. Except for the depression scale, all the scales are constructed based on the official manual (Smith, Ryan, and Sonnega 2019). When more than half of the items for PLQ-based scales are missing, the respective scale will be set to missing. All of the original measures are time-varying, and we take the mean of the measures across waves to create time-invariant outcome variables. Similar to work history outcomes, only person-wave observations over 50 years old are used.

We include spouses' years of birth and gender as demographic controls. In addition, we recognize that the current version of PGS may be subject to three sources of confounding:

population stratification, pleiotropy, and parental genetics effects (Young et al. 2019). Population stratification refers to the differential frequencies of alleles in ancestry groups, making PGS capture non-genetic effects. We minimize population stratification by controlling for the spouses' first principal components (PC; for detail, see Price et al. 2006) and 100 cross-spousal interaction terms of the spouses' PCs. Pleiotropy generally implies that the same genetic invariant may affect multiple life outcomes, which could invalidate the exclusion restriction of NEB PGS as instruments. We account for this problem by controlling for spouses' education (EA) PGS as covariates in the model and note that we also control for NEB-PGS for each spouse as the main effects. The principal components and EA PGS data are also drawn from the HRS team's public release (Ware et al. 2018). Finally, parental or sibling genetics are necessarily related to individuals' genetics and may directly affect individuals' life outcomes. Thus, PGS may also capture the indirect genetic effects from parents and siblings. As reviewed earlier, Mills et al. (2016) showed that the NEB PGS we use captures parental effects. When we take PGS as a measure of the causal effects of individuals' own genetics, such indirect genetic effects are confounding effects. In the next section, we will explicate how our IV method can account for this source of confounding.

Analytical Method: The Genetic IV

We use 2SLS to identify the causal effects of fertility on older adults' life course outcomes. The first stage is specified as:

$$F_j = b_0 + b_1 G_{0j} + b_2 G_{1j} + b_3 G_{0j} \times G_{1j} + \boldsymbol{X_{ij}\gamma} + \epsilon_{ij} \qquad (1)$$

Where j denotes a couple, $0j$ denotes the husband or male spouse, $1j$ denotes the wife or female spouse, F_j denotes fertility or the number of children born to the couple, and G

denotes genetic disposition toward fertility, which is operationalized as NEB PGS in this paper. In addition, $\boldsymbol{X_{ij}}$ denotes covariates, $\boldsymbol{\gamma}$ is a vector of coefficients for the covariates, and ϵ_{ij} denotes the error term. Covariates include the spouses' birth year, gender, principal components, and EA PGS. In addition, we include interaction terms of all principal components (100 interaction terms in total) and spousal interaction of EA PGS for additional controls. Finally, the spousal interaction of NEB PGS, $G_{0j} \times G_{1j}$, is the excluded instrument.

The second stage is specified as:

$$Y_{ij} = b_0 + b_1 G_{0j} + b_2 G_{1j} + b_3 \hat{F}_j + \boldsymbol{X_{ij}}\boldsymbol{\delta} + \eta_{ij} \quad (2)$$

Where Y_{ij} refers to the outcome of interest, $\hat{F}_j$ refers to the predicted value of fertility by Eq. (1), $\boldsymbol{X_{ij}}$ denotes the same set of covariates as in Eq. (1), $\boldsymbol{\delta}$ is a vector of coefficients for covariates, and η_{ij} denotes the error term.

Mirroring the conventional IV method (Angrist, Imbens, and Rubin, 1996), we make (A1) relevance assumption, (A2) monotonicity assumption, (A3) exclusion restriction, and (A4) exchangeability assumptions. The assumptions are also presented in the counterfactual framework in Appendix B. The relevance assumption suggests that spouses' genetic propensity has a multiplicative or interaction effect on realized fertility. Despite the fact that the relevance assumption is empirically testable, we also argue that this assumption has theoretical foundations. Although genetics is often regarded as an individual characteristic, fertility behavior is essentially a couple-level dyadic outcome that relies on the characteristics of both spouses. In this sense, genetics for fertility is a "couple-level" characteristic, and a spouse's genetic propensity for higher fertility is likely to have multiplier effects in increasing the number of children ever born.

The exclusion restriction requires that multiplicative effects of spouses' genetic propensity for fertility on life outcomes are only transmitted via realized fertility. Our rationale is that the spousal interaction term of NEB PGS is associated with an outcome variable either due to its association with the respective main terms (G_{0j}, G_{1j}) or through pleiotropy. Controlling for the main effect of NEB PGS will effectively rule out the former pathway. For the latter pathway, we can understand the pleiotropy problem as the association between NEB PGS and other PGS variables due to shared underlying genetic variants between PGS variables. If NEB is associated with another PGS variable that also impacts fertility through certain behavioral mechanisms, the exclusion restriction would be violated. We account for this problem by controlling for EA PGS, because EA PGS predicts educational attainments, income, cognitive ability, and other social status outcomes that significantly select fertility behavior. We argue that by controlling for EA PGS, including the spousal interaction of EA PGS, we can reasonably assume the exclusion restriction. Nevertheless, we acknowledge a more direct and ideal control for pleiotropy would be the PGS for the respective phenotype, which may capture and block the effects of the shared genetic variants causing pleiotropy biases. Yet such PGS has limited availability. For example, the HRS has no PGS for labor market outcomes such as the number of jobs. To examine the potential influences of the uncaptured pleiotropy, we add robustness analysis by controlling for the neuroticism, extraversion, depression, and life-satisfaction PGS in the analysis of respective phenotypes, as these PGS have been officially provided (Ware et al. 2016). We expect that if EA PGS is a good control or pleiotropy has minimal influences, adding these controls would not change our estimates.

(A3) Exchangeability Assumption: conditional on covariates, the causal effects of the

spousal interaction of genetic propensity on life outcomes can be identified. Key to our assumptions is that we do not assume that the main effects of genetic propensities on life outcomes are identifiable. We acknowledge that, unless particular assumptions of the causal structure are made, the identifiability of main effects is usually the pre-condition for the identifiability of the interaction effect (Vanderweele 2009). That the main effects of PGS variables are confounded (Young et al. 2019, Mills et al. 2016) does challenge the validity of our Assumption A3. In order to justify our Assumption (A3), we assume the independence of confounding structure between spouses. Specifically, we made the following sub-assumptions: Sub-Assumption (1) the spouses' genetic measures do not share any confounders; Sub-Assumption (2) the confounding effect of one spouse's genetic measure (for example, parental effects) is not moderated by the genetic propensity of the other spouse's genetic measure. Regarding Sub-Assumption (1), we believe that the genetic unrelatedness of spouses ensures that the spouses do not share the source of confounding for their genetic measures. For example, genetically unrelated individuals are unlikely to share the same source of parental effects in their PGS measures. Regarding Sub-Assumption (2), we think there are few substantive mechanisms that can explain such cross-spousal interaction effects. In Appendix C, we provide an informal proof regarding how the two sub-assumptions support the validity of Assumption (A3).

(A4) Monotonicity Assumption: this assumption indicates that given a higher level of multiplicative interaction of the couple's NEB PGS, all couples are expected to either have more children (compliers) or have the same number of children (always-takers or never-takers), and no couples would have the lower expected number of children (i.e., no defiers). This

assumption also specifies our compliers or our local average treatment effects (LATE), as the compliers to our IV are those who can be affected positively by stronger genetic dispositions toward fertility. The compliers to our IV thus are distinctive from the compliers to classical sex-mix IV, which refers to parents who prefer an equal sex ratio of their children. We also assume the absence of defiers whose tendency to have children is negatively affected by stronger genetic propensity.

Given the specified assumptions, we proceed to the summary statistics, the first stage results, the falsification test, and the 2SLS results in the next section. In the summary statistics, we will compare the distribution of key variables and covariates between spouses. In the first stage results, we show OLS estimates of the effects of spousal interaction of NEB PGS on fertility and the respective F-statistics for the interaction term. In the falsification test, we try to obtain the OLS, 2SLS, and Intention-to-Treatment (ITT) estimates of the effect of fertility on educational attainments. Finally, we show the OLS and 2SLS estimates of the causal effect of fertility on our selected outcome variables.

Results

Summary Statistics

[TABLE 1 ABOUT HERE]

Summary statistics by gender are presented in Table 1. The statistical significance of gender differences is also presented for most variables. Although almost all respondents have worked at least one job, the male respondents have a much higher number of years worked. The labor income for the male respondents is higher as well, corresponding to the well-documented gender pay gap. Although our analytical sample is composed of couples, family

income and wealth still slightly differ by gender. This is because HRS co-residential couples tend to report slightly different household incomes and wealth. Regarding gender differences in the non-labor market outcomes, we can observe a mixture in terms of the direction of gender differences. For personality traits, the female respondents are rated higher in extraversion, openness, and agreeableness but also higher in neuroticism, a trait related to negative health and social consequences. This pattern corresponds to the existing literature (Weisberg, Young, and Hirsh 2011). The female respondents also exhibit a higher level of loneliness and depression, which is also consistent with the population pattern recorded in the psychology literature (Luhmann and Hawkley 2016).[4]

First Stage Results: The Relevance Assumption

[TABLE 2 ABOUT HERE]

Relevance is the pre-condition for the use of the IV method and is the only assumption that can be empirically tested in a sufficient way. Table 2 presents the first stage results that assessed the relevance assumption. Model 3 is used as the final first stage model in the subsequent 2SLS regressions. In Model 1, we can see that the spousal genetic interaction term has a significantly positive coefficient on the spouses' number of children. This result supports our hypothesis of the multiplicative effect of spousal genetics. Controlling for EA PGS and PCs, including their interactions (i.e., interactions between spouses' EA PGS, and 100 cross-

[4] In Table D1, Appendix D, we compared our analytical sample with all White/European Ancestry HRS respondents with genetic data (N = 10,290). This larger White genetic subsample can be regarded as the pool from which our analytical sample is drawn. Note that the subjects in the White genetic subsample are not necessarily couples. The comparison shows that our analytical sample is very similar to the overall White genetic subsample, although our analytical sample is slightly advantaged in household income, household wealth, and subjective wellbeing (life satisfaction). Respondents in our analysis are living with their spouses in at least one wave, but the White genetic sample includes living-alone respondents. Co-residence is expected to lead to advantages in household finances and mental health.

spousal interaction terms of PCs), has minimal impacts on the genetic interaction effect, which we interpret as suggestive evidence that the inclusion of additional genetic measures would not affect our results and indirect evidence against large pleiotropic effects. This pattern indicates that the spousal interaction effect of NEB PGS is robust. At the bottom of the table, we also report F-statistics for the spousal genetic interaction term. Model 3 shows that F-statistics (15.87) (Stock and Yogo 2005). In the subsequent analyses, the sample sizes may vary by outcome variable, and F-statistics vary as well. We note that for all the 2SLS results reported in the subsequent sections, F-statistics is consistently larger than 10.

Falsification Test

[TABLE 3 ABOUT HERE]

Table 3 presents the results from the falsification test, which estimates the "effect" of the number of children born on educational attainment. Since most (but not all) fertility occurs after schooling (see Rosenbaum 2020 and others), we ask whether the genetic IV can "correct" for any estimated effect of completed fertility on schooling, which is likely to reflect reverse causality and confounding effects. Indeed, since education tends to delay and reduce fertility behavior (Guzzo and Hayford 2020), it is not surprising that OLS yields a negative coefficient (Row 1, Table 3). Consistent with our expectation, 2SLS results imply that fertility has no causal effects on education. We acknowledge that despite statistical insignificance, the coefficients of 2SLS are positive and large in magnitude. To account for the limitation, we also implement ITT analysis, which is an OLS regression of the educational outcomes directly on the proposed IV or spousal genetic interactions. We can see that the ITT coefficient sizes are very small and close to zero (< 0.1), so their statistical insignificance cannot be simply

explained by large standard errors. These results lend support to the validity of our proposed IV.

Main Results

[TABLE 4 ABOUT HERE]

Table 4 presents the estimated effects of completed fertility on life course labor market outcomes. For work history outcomes, OLS results indicate that fertility is associated with all of the outcomes for females, but 2SLS results show that fertility only reduces older females' total years worked ($p < 0.05$). The 2SLS estimates are also substantially larger than OLS in absolute magnitudes, indicating that an additional child reduces women's length of working by five years. The differences in 2SLS and OLS estimates correspond to previous IV-based studies (e.g., Angrist and Evans 1997). On the other hand, males' total years worked is unaffected, as was observed by previous studies (Angrist and Evans 1997; Cools 2017). A more novel finding regarding the 2SLS estimates is that an additional child similarly reduced the number of jobs worked for males by about 0.5 ($p < 0.1$). Different from previous studies, this result suggests that males' labor force participation is not free from the effect of fertility. Given that males' years of working were not influenced, a reduced number of jobs may imply that males are less likely to change their jobs when they have children perhaps because they desire stable jobs. However, we acknowledge that these findings are suggestive and not conclusive due to the lack of precision in the estimates.

Meanwhile, although OLS results show that an additional child decreases long-term household income and wealth by 3% and 20% and decreases females' labor income by 12.4%, the 2SLS results do not support the conclusions. 2SLS results provide no evidence regarding

the causal effects of fertility on older adults' long-term income. Nonetheless, we note HRS records older adults' income at later ages when childbearing is finished. The conclusion of null effects of fertility may simply imply that within each gender, the differences in income at older ages are not influenced by the total number of children. This result does not indicate that working-age females' income at younger ages is not affected by fertility. In addition, we recognize that the 2SLS coefficients are negative and larger in terms of magnitudes than OLS results. It is likely that null findings result from larger standard errors (2SLS standard errors are ten times larger than OLS). Larger sample sizes would be necessary to make clearer conclusions concerning these outcomes.

[TABLE 5 ABOUT HERE]

Table 5 presents OLS and 2SLS results for non-labor market outcomes. For personality traits, OLS results suggest that fertility is positively related to agreeableness for both males and females. However, 2SLS results show that an additional child significantly reduces females' extraversion by 0.194 ($p < 0.05$). Marginally significant evidence suggests that the birth of an additional child reduces, instead of increasing, females' agreeableness ($p < 0.1$). For males, no effects of fertility on personality are observed by 2SLS regression. This pattern may be explained by the fact that the increased number of children primarily restrains mothers' capability and frequency of social interactions in early and middle adulthood, which leads to differences in personality traits in the later life. Furthermore, 2SLS results did not show any significant effects of fertility on depression, life satisfaction, or loneliness on either males or females. These results show the need for much larger samples when using IV methods on these outcomes.

Finally, we examine the potential bias due to pleiotropy. As noted earlier, we do so by adding spousal PGS variables for neuroticism, extraversion, depression, and life satisfaction, and the spousal interaction of the PGS variables to the respective 2SLS model of each outcome. The results, shown in Table E1 in Appendix E, suggest that adding these more direct controls for pleiotropy leads to minor changes to the 2SLS estimates. It implies that our results are robust to the potential violation of exclusion restriction due to pleiotropy.

Conclusion and Discussion

In this paper, we made methodological and empirical contributions to the literature on the causal consequences of fertility on older adults. Methodologically, we propose a novel genetic instrumental variable that utilizes the recent advancements of genotyping technology and large-scale genetic survey data. The proposed IV is the interaction of the spousal genetic propensity for higher fertility (Dyadic Genetic Instrumental Variable, DGIV). The validity of the proposed DGIV relies on two key assumptions: (1) spouses do not share any unobserved confounders for the measured genetic dispositions or PGS measures, and (2) the effect of one spouse's genetic propensity for fertility on realized fertility does not interact with any unobserved confounders to the measures of the genetic propensity of the other spouse. Essentially, these two assumptions imply the independence of the confounding structure of the PGS measurements for spouses. We believe that the fact that spouses are unrelated genetically is consistent with the validity of these two assumptions as well as the validity of our proposed DGIV. We argue that our proposed DGIV overcomes the difficulty that the measurements of genetic propensities tend to be confounded by family and environmental factors and can serve as a new tool for analyzing human fertility behavior.

Using a sample of 3,282 couples from HRS, we apply the proposed GIV to analyze the impacts of fertility on labor market outcomes and non-labor market outcomes of personality traits and subjective wellbeing. Since we utilized an older adult sample whose career path is finished, our estimates for labor market outcomes provide a summary measure of the life-course labor force participation, a novel perspective that is rare in the previous reach due to the limitation of data. Similar to previous works using census data and Twin and Sibling IV (Angrist and Evans 1998; Jacobsen et al. 1999), we find that an additional child leads to reduced lifetime work length for females by five years. Utilizing the advantage that HRS is a survey of older adults, we also analyzed how the number of jobs throughout the life course is affected by fertility. We find suggestive evidence that males' work for 0.5 fewer jobs if they have one more child. For males, this effect may be explained by fewer job changes or desire for stable work schedules after significant family change. This finding also to some degree challenges the previous finding that males are unaffected by fertility behavior. We also believe that the consistency of our findings on labor force participation with the previous IV literature provides empirical support for the validity of our DGIV.

Regarding personality traits, our results show that fertility causally decreases older females' extraversion. This may be the result of the fact that mothers have reduced social interactions due to childcare workloads and less labor force experience. Suggestive evidence also shows that females have reduced agreeableness if they have more children. This may be related to the increased life conflicts, such as work-life conflicts, that women face after childbearing (Nomaguchi and Milkie 2020). We also analyzed links between completed fertility and mental health in older age but the results were too imprecise to be conclusive.

This study has two more general implications. First, we believe that our paper adds more evidence that fertility has unequal impacts for males and females, particularly in underexamined non-labor market outcomes. We suggest that more causality-driven studies would be necessary to enrich our knowledge in subjective wellbeing outcomes, a topic that is still understudied compared to labor market outcomes. Second, this study demonstrates a wider utility of the use of genetics in social science research. Previous sociogenomic research tends to focus on a relatively narrower range of topics, including gene-by-environment interaction, nature versus nurture, or heritability and intergenerational mobility (Freese and Shostak 2009; Conley 2016). This study shows that genes can be used as a statistical tool that benefits a much wider range of social science inquiries. While many studies have used genetic measures as instrumental variables, previous work has uniformly focused on individual level genetic instruments, which have clear concerns over being invalid (i.e. not excludable from the second stage regression) (Conley and Fletcher 2017). We extend the use of genetic instruments to a more defensible scenario by constructing couple-level instruments (DGIV) to predict couple-level endogenous outcomes. Future methodological and empirical literature could continue to explore the potential of genes as a tool that is beyond current uses.

Nevertheless, we acknowledge that this study has several limitations. First, the biological or molecular mechanisms behind the significant spousal genetic interaction effects on realized fertility are unclear. A mechanism-based explanation for this interaction effect would be essential for consolidating the validity of our proposed DGIV. Second, a sample size of 3,282 coupes limited our ability to estimate precise results in several cases. Empirically, the previous IV studies tend to rely on census (e.g. Angrist and Evans 1998; Jacobsen et al. 1999) or register

data (e.g. Cools et al. 2017). The sample sizes of these more powerful IV studies are at the level of hundreds of thousands of respondents. The empirical power of our analysis is unavoidably limited by the more modest sample sizes compared to the older studies. The empirical application of our approach using large samples would be important for future research. Finally, our study restricts the analysis to European ancestry respondents due to the current lack of useful polygenic scores for other population groups. However, we do expect that this inequality in the availability of genetic data and the production of genetic knowledge will be reduced in the future so that work will not be forced to limit analysis to respondents of European ancestry.

Note

1. Although a large literature has examined the effects of early (i.e. teenage) fertility on longer term outcomes (Furstenberg 2013; Sweeny and Raley 2014, Kane et al. 2013, Fletcher and Wolfe 2009, Fletcher 2012; Fletcher and Polos 2017 reviews this literature), our paper focuses on total fertility and adults.
2. Fletcher and Kim (2018) applied the twin-IV to an American adolescent sample and found that sibship sizes causally affect adolescents' personality traits, but similar methods have not been applied to the study of parents.
3. Later analyses reveal that the NEB PGS is confounded by childhood family environments, but the magnitude of confounding is limited and the coefficient of the NGB PGS remains large after family fixed effects ruled out family environmental confounders (Mills, Barban, and Tropf 2016).

References

Agüero, J. M., & Marks, M. S. (2011). Motherhood and female labor supply in the developing world: Evidence from infertility shocks. Journal of Human Resources, 46(4), 800–826. https://doi.org/10.3368/jhr.46.4.800

Agüero, J. M., & Marks, M. S. (2008). Motherhood and Female Labor Force Participation: Evidence from Infertility Shocks. American Economic Review, 98(2), 500–504. https://doi.org/10.1257/aer.98.2.500

Angrist, J. D., & Evans, W. N. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. American Economic Review, 88(3), 450–477. https://doi.org/10.2307/116844

Angrist, J., Lavy, V., & Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. Journal of Labor Economics, 28(4), 773–823. https://doi.org/10.1086/653830

Balbo, N., Billari, F. C., & Mills, M. (2013). Fertility in Advanced Societies: A Review of Research. European Journal of Population, 29(1), 1–38. https://doi.org/10.1007/s10680-012-9277-y

Barban, N., Jansen, R., De Vlaming, R., Vaez, A., Mandemakers, J. J., Tropf, F. C., Shen, X., Wilson, J. F., Chasman, D. I., Nolte, I. M., Tragante, V., Van Der Laan, S. W., Perry, J. R. B., Kong, A., Ahluwalia, T. S., Albrecht, E., Yerges-Armstrong, L., Atzmon, G., Auro, K., … Mills, M. C. (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behavior. Nature Genetics, 48(12), 1–7. https://doi.org/10.1038/ng.3698

Becker, G. (1993). A Treatise On The Family (Vol. 22, Issue 1). Harvard University Press. https://doi.org/10.1111/j.1545-5300.1983.119_8.x

Bellemare, M. F., & Wichman, C. J. (2020). Elasticities and the Inverse Hyperbolic Sine Transformation. Oxford Bulletin of Economics and Statistics, 82(1), 50–61. https://doi.org/10.1111/obes.12325

Benard, S., & Correll, S. J. (2010). Normative discrimination and the motherhood penalty. In Gender and Society (Vol. 24, Issue 5). https://doi.org/10.1177/0891243210383142

Bhalotra, S., & Clarke, D. (2020). The Twin Instrument: Fertility and Human Capital Investment. Journal of the European Economic Association, 18(6), 3090–3139. https://doi.org/10.1093/jeea/jvz058

Black, D. A., Joo, J., LaLonde, R., Smith, J. A., & Taylor, E. J. (2022). Simple tests for selection: Learning more from instrumental variables. *Labour Economics*, *79*, 102237

Braakmann, N., & Wildman, J. (2016). Reconsidering the effect of family size on labour supply: the twin problems of the twin birth instrument. Journal of the Royal Statistical Society: Series A (Statistics in Society), 179(4), 1093–1115. https://doi.org/10.1111/rssa.12160

Bronars, S. G., & Grogger, J. (1994). Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment. The American Economic Review, 84(5), 1141–1156.

Cáceres-Delpiano, J. (2012). Can We Still Learn Something From the Relationship Between Fertility and Mother's Employment? Evidence From Developing Countries. Demography, 49(1), 151–174. https://doi.org/10.1007/s13524-011-0076-6

Chun, H., & Oh, J. (2002). An instrumental variable estimate of the effect of fertility on the labour force participation of married women. Applied Economics Letters, 9(10), 631–634. https://doi.org/10.1080/13504850110117850

Cools, S., Markussen, S., & Strøm, M. (2017). Children and Careers: How Family Size Affects Parents' Labor Market Outcomes in the Long Run. Demography, 54(5), 1773–1793. https://doi.org/10.1007/s13524-017-0612-0

Conley, D. (2016). Socio-genomic research using genome-wide molecular data. Annual Review of Sociology, 42, 275-299.

Correll, S. J., Benard, S., & Paik, I. (2007). Getting a job: Is there a motherhood penalty? American Journal of Sociology, 112(5), 1297–1338. https://doi.org/10.1086/511799

Daouli, J., Demoussis, M., & Giannakopoulos, N. (2009). Sibling-sex composition and its effects on fertility and labor supply of Greek mothers. Economics Letters, 102(3), 189–191. https://doi.org/10.1016/j.econlet.2009.01.002

Deaton, A., & Stone, A. A. (2014). Evaluative and hedonic wellbeing among those with and without children at home. Proceedings of the National Academy of Sciences of the United States of America, 111(4), 1328–1333. https://doi.org/10.1073/pnas.1311600111

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. Journal of Personality Assessment, 49(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Dima, B., Dima, S., Nachescu, M., Saccon, C., & Nachescu, M. (2015). Professional Autonomy and IFRSS Adoption. Journal of Accounting and Management Information Systems, 14(4), 627–654.

Farbmacher, H., Guber, R., & Vikström, J. (2018). Increasing the credibility of the twin birth instrument. Journal of Applied Econometrics, 33(3), 457–472. https://doi.org/10.1002/jae.2616

Fletcher, J. M. (2012). The effects of teenage childbearing on the short-and long-term health behaviors of mothers. *Journal of Population Economics*, *25*(1), 201-218.

Fletcher, J., & Padrón, N. (2016). The effects of teenage childbearing on adult soft skills development. *Journal of population economics*, *29*(3), 883-910.

Fletcher, J. M., & Polos, J. (2017). Nonmarital and teen fertility. In *The Oxford Handbook of Women and the Economy*.

Fletcher, J. M., & Wolfe, B. L. (2009). Education and labor market consequences of teenage childbearing evidence using the timing of pregnancy outcomes and community fixed effects. *Journal of Human Resources*, *44*(2), 303-325.

Fletcher, J. M., & Kim, J. (2019). The effect of sibship size on non-cognitive Skills: Evidence from natural experiments. Labour Economics, 56(October 2017), 36–43. https://doi.org/10.1016/j.labeco.2018.11.004

Freese, J., & Shostak, S. (2009). Genetics and social inquiry. Annual Review of Sociology, 35, 107-128.

Furstenberg, F. F. (2003). Teenage Childbearing as a Public Issue and Private Concern. Annual Review of Sociology, 29, 23–39. https://doi.org/10.1146/annurev.soc.29.010202.100205

Geyer, J., Haan, P., & Wrohlich, K. (2015). The effects of family policy on maternal labor supply: Combining evidence from a structural model and a quasi-experimental approach. *Labour Economics*, *36*, 84-98

Glass, J., Simon, R. W., & Andersson, M. A. (2016). Parenthood and happiness: Effects of work-family reconciliation policies in 22 OECD countries. American Journal of Sociology, 122(3), 886–929. https://doi.org/10.1086/688892

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. Journal of Personality and Social Psychology, 59(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Guzzo, K. B., & Hayford, S. R. (2020). Pathways to Parenthood in Social and Family Contexts: Decade in Review, 2020. Journal of Marriage and Family, 82(1), 117–144. https://doi.org/10.1111/jomf.12618

Huber, M. (2015). Testing the validity of the sibling sex ratio instrument. Labour, 29(1), 1–14. https://doi.org/10.1111/labr.12045

Hughes, M. E., Waite, L. J., Hawkley, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys: Results from two population-based studies. Research on Aging, 26(6), 655–672. https://doi.org/10.1177/0164027504268574

Hutteman, R., Bleidorn, W., Keresteš, G., Brković, I., Butković, A., & Denissen, J. J. A. (2014). Reciprocal associations between parenting challenges and parents' personality development in young and middle adulthood. European Journal of Personality, 28(2), 168–179. https://doi.org/10.1002/per.1932

Jacobsen, J. P., Pearce, J. W., & Rosenbloom, J. L. (1999). The effects of childbearing on married women's labor supply and earnings: Using twin births as a natural experiment. Journal of Human Resources, 34(3), 449–474. https://doi.org/10.2307/146376

Jokela, M., Kivimäki, M., Elovainio, M., & Keltikangas-Järvinen, L. (2009). Personality and Having Children: A Two-Way Relationship. Journal of Personality and Social Psychology, 96(1), 218–230. https://doi.org/10.1037/a0014058

Kane, J. B., Morgan, S. P., Harris, K. M., & Guilkey, D. K. (2013). The educational consequences of teen childbearing. *Demography*, *50*(6), 2129-2150.

Louis, J. F., Thoma, M. E., Sørensen, D. N., Mclain, A. C., King, R. B., Sundaram, R., Keiding, N., & Buck Louis, G. M. (2013). The prevalence of couple infertility in the United States from a male perspective: Evidence from a nationally representative sample. Andrology, 1(5), 741–748. https://doi.org/10.1111/j.2047-2927.2013.00110.x

Lundberg, S., & Rose, E. (2000). Parenthood and the earnings of married men and women. *Labour Economics*, *7*(6), 689-710.

Luhmann, M., & Hawkley, L. C. (2016). Supplemental Material for Age Differences in Loneliness From Late Adolescence to Oldest Old Age. Developmental Psychology, 52(6), 943–959. https://doi.org/10.1037/dev0000117.supp

Lundborg, P., Plug, E., & Rasmussen, A. W. (2017). Can Women Have Children and a Career? IV Evidence from IVF Treatments. American Economic Review, 107(6), 1611–1637. https://doi.org/10.1257/aer.20141467

Margolis, R., & Myrskylä, M. (2011). A Global Perspective on Happiness and Fertility. Population and Development Review, 37(1), 29–56. https://doi.org/10.1111/j.1728-4457.2011.00389.x

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., Daly, M. J., Bustamante, C. D., & Kenny, E. E. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. American Journal of Human Genetics, 100(4), 635–649. https://doi.org/10.1016/j.ajhg.2017.03.004

Mills, M. C., Barban, N., & Tropf, F. C. (2018). The sociogenomics of polygenic scores of reproductive behavior and their relationship to other fertility traits. RSF: The Russell Sage Foundation Journal of the Social Sciences, 4(4), 122–136. https://doi.org/10.7758/rsf.2018.4.4.07

Musick, K., Meier, A., & Flood, S. (2016). How Parents Fare: Mothers' and Fathers' Subjective Wellbeing in Time with Children. American Sociological Review, 81(5), 1069–1095. https://doi.org/10.1177/0003122416663917

Nelson-Coffey, S. K., Killingsworth, M., Layous, K., Cole, S. W., & Lyubomirsky, S. (2019). Parenthood Is Associated With Greater Well-Being for Fathers Than Mothers. Personality and Social Psychology Bulletin, 45(9), 1378–1390. https://doi.org/10.1177/0146167219829174

Nomaguchi, K., & Milkie, M. A. (2020). Parenthood and Wellbeing: A Decade in Review. Journal of Marriage and Family, 82(1), 198–223. https://doi.org/10.1111/jomf.12646

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8), 904–909. https://doi.org/10.1038/ng1847

Priebe, J. (2020). Quasi-experimental evidence for the causal link between fertility and subjective wellbeing. Journal of Population Economics, 33(3), 839–882. https://doi.org/10.1007/s00148-020-00769-3

Roberts, B. W., Wood, D., & Smith, J. L. (2005). Evaluating Five Factor Theory and social investment perspectives on personality trait development. Journal of Research in Personality, 39(1 SPEC. ISS.), 166–184. https://doi.org/10.1016/j.jrp.2004.08.002

Rosenbaum, P. (2020). Does early childbearing matter? New approach using Danish register data. *Labour Economics*, *65*, 101852

Rosenzweig, M. R., & Wolpin, K. I. (2000). Natural "natural experiments" in economics. Journal of Economic Literature, 38(4), 827–874. https://doi.org/10.1257/jel.38.4.827

Rosenzweig, Mark R., and Kenneth I. Wolpin. (1980) "Testing the quantity-quality fertility model: The use of twins as a natural experiment." Econometrica: journal of the Econometric Society: 227-240.

Seltzer, J. A., & Bianchi, S. M. (2013). Demographic change and parent-child relationships in adulthood. Annual Review of Sociology, 39, 275–290. https://doi.org/10.1146/annurev-soc-071312-145602

Smith, J., Fisher, G., Ryan, L., Clarke, P., House, J., & Weir, D. (2013). Psychosocial and lifestyle questionnaire. *Survey Research Center, Institute for Social Research*.

Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and Change of Personality Across the Life Course: The Impact of Age and Major Life Events on Mean-Level and Rank-Order Stability of the Big Five. Journal of Personality and Social Psychology, 101(4), 862–882. https://doi.org/10.1037/a0024950

Stanca, L. (2012). Suffer the little children: Measuring the effects of parenthood on wellbeing worldwide. Journal of Economic Behavior and Organization, 81(3), 742–750. https://doi.org/10.1016/j.jebo.2010.12.019

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, *80*(4.2), 1.

Sweeney, M. M., & Raley, R. K. (2014). Race, ethnicity, and the changing context of childbearing in the United States. Annual Review of Sociology, 40, 539–558. https://doi.org/10.1146/annurev-soc-071913-043342

Thoma, M. E., McLain, A. C., Louis, J. F., King, R. B., Trumble, A. C., Sundaram, R., & Buck Louis, G. M. (2013). Prevalence of infertility in the United States as estimated by the current duration approach and a traditional constructed approach. Fertility and Sterility, 99(5), 1324-1331.e1. https://doi.org/10.1016/j.fertnstert.2012.11.037

Thomas, P. A., Liu, H., & Umberson, D. (2017). Family Relationships and Well-Being. Innovation in Aging, 1(3), 1–11. https://doi.org/10.1093/geroni/igx025

Umberson, D. (1992). Gender, marital status and the social control of health behavior. Social Science and Medicine, 34(8), 907–917. https://doi.org/10.1016/0277-9536(92)90259-S

Umberson, D., Crosnoe, R., & Reczek, C. (2010). Social relationships and health behavior across the life course. Annual Review of Sociology, 36, 139–157. https://doi.org/10.1146/annurev-soc-070308-120011

Umberson, D., & Gove, W. R. (1989). Parenthood and Psychological Well-Being. Journal of Family Issues, 10(4), 440–462. https://doi.org/10.1177/019251389010004002

Umberson, D., Pudrovska, T., & Reczek, C. (2010). Parenthood, childlessness, and wellbeing: A life course perspective. Journal of Marriage and Family, 72(3), 612–629. https://doi.org/10.1111/j.1741-3737.2010.00721.x

Vanderweele, T. J. (2009). On the distinction between interaction and effect modification. Epidemiology, 20(6), 863–871. https://doi.org/10.1097/EDE.0b013e3181ba333c

Ware, E., Schmitz, L., Gard, A., & Faul, J. (2018). HRS polygenic scores—Release 3: 2006–2012 genetic data. Ann Arbor: Survey Research Center, University of Michigan.

Weisberg, Y. J., De Young, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. Frontiers in Psychology, 2(AUG), 1–11. https://doi.org/10.3389/fpsyg.2011.00178

Young, A. I., Benonisdottir, S., Przeworski, M., & Kong, A. (2019). Deconstructing the sources of genotype-phenotype associations in humans. Science, 365(6460), 1396–1400. https://doi.org/10.1126/science.aax3710

Table 1. Summary statistics (mean and standard deviation) of the analytical sample.

Variable	Female	Male	T-Tests of Differences
Labor Market Outcome			
Ever worked	0.96 (0.19)	0.996 (0.07)	***
Total years of working	22.74(11.71)	35.39 (9.47)	***
Number of jobs	1.74 (1.08)	2.12 (1.21)	***
Long-term household income	10.89 (12.83)	11.13 (12.83)	Insig.
Long-term net household wealth	70.23 (117.82)	71.60 (123.06)	Insig.
Long-term labor income	1.90 (2.91)	3.45 (5.58)	***
Non-Labor Market Outcome			
Neuroticism	2.10 (0.56)	1.95 (0.53)	***
Extraversion	3.23 (0.52)	3.13 (0.51)	***
Openness	2.94 (0.51)	2.94 (0.50)	Insig.
Agreeableness	3.65 (0.36)	3.37 (0.45)	***
Conscientiousness	3.40 (0.37)	3.27 (0.40)	***
Depression	1.21 (1.34)	0.91 (1.10)	***
Life Satisfaction	5.04 (1.29)	4.99 (1.22)	†
Loneliness	1.31 (1.37)	1.14 (1.31)	***
Falsification Test			
Years of schooling	13.34 (2.27)	13.53 (2.69)	Insig.
Treatment and Instruments			
Fertility (Number of children)	2.59 (1.49)	2.59 (1.49)	--
NEB PGS	0.04 (1.01)	0.002 (1.08)	--
Covariates			
Year of Birth	1944 (10.92)	1940 (10.89)	***
EA PGS	0.001 (0.99)	0.03 (0.99)	--

Data source: The Health and Retirement Study, 2004–2016.

Note: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. "Insig." stands for insigificnat. Standard deviations are presented in parentheses. Population refers to all HRS respondents with non-missing values for any of the variables. To save some digits, long-term household income, long-term household wealth, and long-term labor income are reported in thousands. Year of birth is rounded to integers. PGS measures have been standardized. Fertility variable is constructed to be the same for males and females. Equal variance is assumed for T-tests. T-tests are not implemented on EA PGS, NEB PGS, and fertility, because they are constructed to have no group differences in means. Summary statistics of PCs are not reported.

Table 2. Regression of Couples' Fertility on Proposed IV and Control Variables.

	Model 1	Model 2	Model 3
	Outcome: Number of Children Ever Born		
Husband's NEB PGS	0.322***	0.326***	0.316***
	(0.025)	(0.025)	(0.025)
Wife's NEB PGS	0.428***	0.429***	0.434***
	(0.027)	(0.027)	(0.028)
Husband's NEB PGS × Wife's NEB PGS	0.103***	0.104***	0.102***
	(0.025)	(0.025)	(0.026)
Husband's Age	-0.011*	-0.011*	-0.011*
	(0.005)	(0.005)	(0.005)
Wife's Age	0.039***	0.039***	0.039***
	(0.005)	(0.005)	(0.005)
Husband's Education PGS		0.039	0.022
		(0.024)	(0.024)
Wife's Education PGS		-0.002	-0.008
		(0.024)	(0.024)
Husband's Education PGS × Wife's Education PGS		0.018	0.006
		(0.022)	(0.023)
Couple's PCs, Main Effects	Controlled	Controlled	Controlled
Couple's PCs, Interaction Effects			Controlled
F-statistics for the proposed GIV (Husband's PGS × Wife's PGS)	16.55	16.86	15.87
R-Squared	0.20	0.20	0.23
N	6,564	6,564	6,564

Data source: Health and Retirement Study, 2004–2016.

Note: † p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Cluster standard errors at the spousal level are presented in paretheses.

Table 3. OLS, 2SLS, and ITT Estimation of the Effects of Fertility on Educational Attainments

Outcome Variable	Spouses' Mean Years of Schooling	Husband's Years of schooling	Wife's Years of schooling
Ordinary Least Square (OLS)	-0.111***	-0.074*	-0.147***
	(0.028)	(0.037)	(0.029)
Two-State Least Square (2SLS)	0.218	0.367	0.125
	(0.314)	(0.437)	(0.324)
Intention to Treatment (ITT)	0.026	0.040	0.015
	(0.031)	(0.042)	(0.032)
Control Variables	Yes	Yes	Yes

Data source: Health and Retirement Study, 2004–2016.

Note: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. In ITT model, the coefficients reflect the direct regression of years of schooling on the spousal genetic interaction term. Cluster standard errors at the spousal level are presented in parentheses.

Table 4. Selected coefficients of OLS and 2SLS regressions of selected labor market outcomes on realized fertility.

Gender	Ever Worked	Total years of working	Number of Jobs	Long-term Household Income	Long-term Net Household Wealth	Long-term Labor Income
			OLS			
Male	-0.001	0.042	-0.015†	-0.030***	-0.215***	0.044
	(0.001)	(0.140)	(0.008)	(0.008)	(0.044)	(0.050)
Female	-0.006**	-1.648***	0.006	-0.029**	-0.210***	-0.124*
	(0.002)	(0.196)	(0.02)	(0.009)	(0.045)	(0.054)
			2SLS			
Male	-0.003	-0.758	-0.504†	-0.054	-0.568	0.217
	(0.010)	(1.663)	(0.295)	(0.100)	(0.522)	(0.594)
Female	-0.043	-4.826*	-0.391	-0.009	-0.504	-0.009
	(0.027)	(2.309)	(0.250)	(0.101)	(0.53)	(0.643)

Data source: Health and Retirement Study, 2004–2016.

Note: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Cluster standard errors at the spousal level are presented in parentheses. Control variables are adjusted in all models. Long-term income, wealth, and labor income are rescaled by hyperbolic sine function, and the respective coefficients can be interpreted as semi-elasticity. Detailed results are available in Table D2.1 to D2.4 in Appendix D.

Table 5. Selected coefficients of OLS and 2SLS regressions of selected non-labor market outcomes on realized fertility.

Gender	Neuroticism	Extraversion	Openness	Agreeable-ness	Conscientio-usness	Depression	Life Satisfaction	Loneliness
				OLS				
Male	-0.002	0.010	-0.011	0.015*	-0.004	0.027†	0.013	0.009
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.015)	(0.017)	(0.018)
Female	-0.010	-0.005	-0.015*	0.013**	-0.006	0.039*	-0.013	0.034†
	(0.008)	(0.007)	(0.007)	(0.005)	(0.005)	(0.018)	(0.018)	(0.019)
				2SLS				
Male	0.033	-0.087	0.075	-0.045	-0.013	0.152	0.216	0.046
	(0.081)	(0.081)	(0.076)	(0.071)	(0.062)	(0.174)	(0.19)	(0.201)
Female	-0.078	-0.194*	-0.022	-0.105†	-0.052	-0.221	0.091	0.184
	(0.088)	(0.091)	(0.079)	(0.063)	(0.059)	(0.209)	(0.208)	(0.217)

Data source: Health and Retirement Study, 2004–2016.

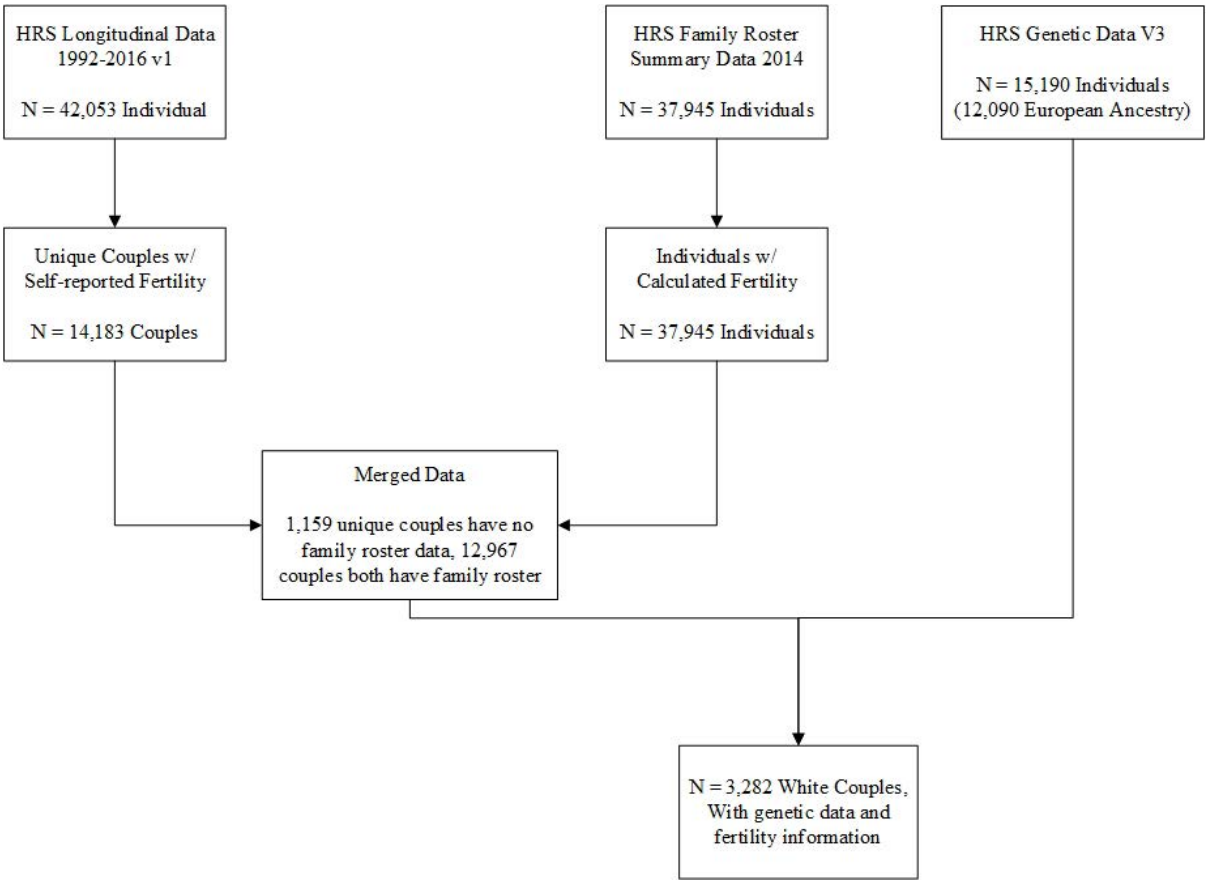
Note: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Cluster standard errors at the spousal level are presented in parentheses. Control variables are adjusted in all models. Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness are measured by Big-5 personality scale and ranges from 1 to 4. Depression is measured by the short CESD scale provided and ranges from 0 to 8. Life-satisfaction is measured by Diener's measure of life satisfaction and ranges from 1 to 7. Loneliness is measured by the 3-item adapted UCLA-R loneliness scale and ranges from 0 to 6. Cognitive ability is measured by the Rand HRS sum scores that range from 0 to 35. Detailed results are available in Table D3.1 to D3.4 in Appendix D.

Supplementary Materials for

Estimating Causal Effects of Fertility Outcomes: Evidence Using A Dyadic Genetic Instrumental Variable Approach

Appendix A: Sample Construction

Figure A1. Construction of Fertility IV Data using RAND HRS Data Files



We construct our analytical sample using two HRS datasets: (1) Rand HRS Longitudinal Data 1992-2006 v1 (2) Rand HRS Family Roster Data 2014.

Rand HRS Longitudinal Data is used to identify unique couples and gather basic demographic information. By unique couple, we refer to the females and males who identify no more than one spouse in Rand HRS longitudinal data file 1992-2016 data file. Using this standard, we identified a sample of 14,183 unique couples.

Although Rand HRS Longitudinal Data provides a fertility variable, "Number of Children Ever Born", we are concerned that this self-report variable has limited accuracy. We reach the Rand HRS

Family Roster Data 1992-2014 to obtain fertility data calculated from the family roster. Rand HRS Family Roster Data has two files: a summary file, where each observation represents a HRS respondent regardless of whether they have kids, and a kid-roster file, where each observation represents a child of the HRS respondent.

The summary data provides a calculated variable of "respondent's own kid" in the summary file. Rand used "kid" to refer to the children of HRS respondents. Rand HRS constructed this variable using the kid-roster data. According to Rand HRS document, this variable summarizes the number of children in the family roster file whose relationship with HRS respondent can be confidently considered as biological child:

> "Respondent's own kid (RwOWNKIDKN) is a count of Respondent's own kids … the variable is the sum up of reported kids who have good linkage".

The summary data provides data for N = 37,945 cases. We merged the cleaned family roster data with the unique couple data. Among all the unique couples, N = 1,159 unique couples have no family roster data. Using the merged dataset, we create a realized fertility variable based on the following rules:

1. When the female (wife) has "respondent's own kid" information, the female's report is used as the self-reported couple fertility
2. When the female's report information is missing, the male's report is used as the self-reported couple fertility
3. When neither of the spouses has non-missing respondent's own kid variable, the realized fertility variable is set to missing.
4. If both spouses self-report to give birth to zero children (e.g. both spouses have non-

missing "number of children ever born" variables that have values of zero) and have missing respondent's own kid variable, then the Rand-calculated couple fertility variable is imputed with zero.

Finally, we merged the unique couple data with the HRS Genetic Data V3, which includes N = 15,190 cases. Among them, 12,090 are identified as European Ancestry by the HRS team. The distribution of fertility of the Rand-calculated couple fertility among the 3,289 white couples with genetic data are given in Table A. After omitting 8 couples that have no fertility information, the final analytical sample includes 3,282 unique couples. The fertility information for the analytical sample is presented in Table A1.

Table A1. Summary of the Distribution of Fertility Information in the Analytical Sample

Reported Couple Fertility (# of children)	Frequency (# of couples)	Percent (%)
0	216	6.58
1	378	11.52
2	1,173	35.74
3	813	24.77
4	406	12.37
5	168	5.12
6	76	2.32
7	27	0.82
8	15	0.46
9	6	0.18
10	3	0.09
19	1	0.03
Total Number of Couples	3,282	100

Data source: The Health and Retirement Study, 2004–2016.
Note: Mean = 2.59. SD = 1.49.

Appendix B. Genetic IV Identification Assumptions.

In this Section, we present our IV Assumptions via counterfactual notations.

$$E\left[F_j\left(G_{0j}, G_{1j}\right) - F_j\left(G'_{0j}, G_{1j}\right)\right] \neq E\left[F_j\left(G_{0j}, G'_{1j}\right) - F_j\left(G'_{0j}, G'_{1j}\right)\right] \qquad \text{(B1)}$$

Where G'_{0j} and G'_{1j} denotes counterfactual values of G_{0j} and G_{1j} so that $G_{0j} \neq G'_{0j}$ and $G_{1j} \neq G'_{1j}$, and $F_j(\cdot)$ stands for the potential outcomes of fertility given certain values of genetic dispositions.

Formula (B1) explicates Assumption (A1), which specified a causal interaction effect of the spouses' genetic dispositions on fertility (Vanderweele 2009). This implies that the causal effect of one spouse' genetic dispositions is contingent on the other's genetic dispositions. Thus, Assumption (A1) serves as our relevance assumption, which states that the couple's genetic dispositions have a multiplicative effect on their realized fertility.

$$Y_{ij}\left(G_{0j}, G_{1j}, F_j\right) - Y_{ij}(G'_{0j}, G_{1j}, F_j) = Y_{ij}\left(G_{0j}, G'_{1j}, F_j\right) - Y_{ij}(G'_{0j}, G'_{1j}, F_j) \qquad \text{(B2)}$$

Where $Y_{ij}(\cdot)$ stands for the potential outcomes of life outcomes given certain values of spousal genetic dispositions.

Formula (B2) explicates our Assumption (A2). Assumption (A2), or the exclusion restriction, states that any interaction effects of the spouses' genetics on life outcomes are transmitted via realized fertility.

$$\begin{aligned} E\{Y_{Ij}\left(G_{0j}, G_{1j}\right) - Y_{Ij}\left(G'_{0j}, G_{1j}\right) - \left[Y_{Ij}\left(G_{0j}, G'_{1j}\right) - Y_{Ij}\left(G'_{0j}, G'_{1j}\right)\right] \big| \boldsymbol{X_{ij}}\} \\ = E\left(Y_{Ij} \middle| G_{0j}, G_{1j}, \boldsymbol{X_{ij}}\right) - E\left(Y_{Ij} \middle| G'_{0j}, G_{1j}, \boldsymbol{X_{ij}}\right) \\ - \left[E\left(Y_{Ij} \middle| G_{0j}, G'_{1j}, \boldsymbol{X_{ij}}\right) - E\left(Y_{Ij} \middle| G'_{0j}, G'_{1j}, \boldsymbol{X_{ij}}\right)\right] \end{aligned} \qquad \text{(B3)}$$

Formula (B3) explicates Assumption (A3), or exchangeability assumption. Assumption (A3) states that conditional on all control variables, the causal effects of the spousal interaction

of genetic dispositions on life outcomes can be identified.

In our parametric model, Assumption (A) implies that the spousal interaction of NEB PGS, $G_{0j} \times G_{1j}$, does not share the same cause with the outcome variable, holding constant all covariates. In other words, we argue that even if $E(G_{0j}\varepsilon_{0j}|\boldsymbol{X_{ij}}) \neq 0$ and $E(G_{1j}\varepsilon_{1j}|\boldsymbol{X_{ij}}) \neq 0$, controlling for the "contaminated" main effects will keep the spousal interaction term free from each spouses' family-level confounders. In addition, let u_{0j} and u_{1j} stand for the unobserved confounders for the husband's and wife's genetics effects, the sub-assumptions can be parametrically presented as (1) $E\left(G_{0j}^{NEB}\mathrm{u}_{1j}\right) = 0$ and $E\left(G_{1j}^{NEB}\mathrm{u}_{0j}\right) = 0$; (2) the interaction terms $G_{0j}^{NEB}\mathrm{u}_{1j}$ and $G_{1j}^{NEB}\mathrm{u}_{0j}$, or any other higher-order interactions that involve any of these two terms, have zero coefficients on the outcome variables.

$$F_j\left(G_{0j}, G_{1j}\right) - F_j\left(G'_{0j}, G_{1j}\right) \geq F_j\left(G_{0j}, G'_{1j}\right) - F_j\left(G'_{0j}, G'_{1j}\right) \text{ if } G_{1j} > G'_{1j} \quad \text{(A4)}$$

Where $F_j(\cdot)$ stands for the potential outcomes of the number of children born to the couple given certain values of spousal genetic dispositions.

Assumption (A4), or monotonicity assumption, states that there is no couple who acts against the interaction between genetic dispositions. In other words, the interaction effect is either positive (compliers) or zero (always-takers or never-takers). There are no defiers who display a negative interaction effect.

Note that for the purpose of simplicity, we present counterfactual notations to our proposed IV by fixing the change in the male's genetic disposition and use the reference level of the female' s genetic dispositions as the moderator. We note that our assumptions also apply to the case when the female's genetic dispositions are presented as the moderator in counterfactual notations.

Appendix C. The Assumption of Independence of Confounding Structure

In this Section, we would prove that our stated assumptions support the identification of causal interaction effects without assuming absence of unobserved confounders. We will use general notations such as X_1, X_2, instead of notations specific to this paper (G_{0j}, G_{1j}), to facilitate reading and show that our assumptions are applicable to general setting not limited to genetics. We use X_1, X_2 to refer to the treatment variables whose causal interaction effects are of analytical interests. We use U_1 and U_2 to denote two unobserved variables that may cause X_1, X_2, and Y, the outcome of interests, at the same time, and therefore U_1 and U_2 are confounders to the causal effects of X_1, X_2 on Y. We also use $\boldsymbol{C}$ to denote a set of unobserved confounders that may cause X_1, X_2, and Y.

Given the specifications above, an estimand for the causal interaction effects can be specified as (Vanderweele 2009)

$$E\{Y(x_1, x_2) - Y(x_1', x_2) - [Y(x_1, x_2') - Y(x_1', x_2')] | \boldsymbol{C}, U_1, U_2\} \quad (1)$$

Where $Y(\cdot)$ denotes the potential outcomes given certain values of X_1 and X_2, and the lower case x_1, x_2, x_1', and x_2' denotes a particular value assigned to X_1 and X_2. $Y(x_1, x_2)$ is an abbreviation of $Y(X_1 = x_1, X_2 = x_2)$.

The first part of Formula (1), $Y(x_1, x_2) - Y(x_1', x_2)$, denotes the effect of change in X_1 from x_1' to x_1 on Y, given certain levels of $X_2 = x_2$, $\boldsymbol{C}$, U_1 and U_2. $Y(x_1, x_2') - Y(x_1', x_2')$ denotes the effect of the same magnitude of change in X_1 on Y, when the level of X_2 has been changed to a different value, x_2', and $\boldsymbol{C}$, U_1 and U_2 are hold constant. Therefore, this estimand describes the causal interaction effect where X_2 causally moderates the effect of X_1 on Y. Note that with a simple rearrange of these counterfactual terms, X_1

can be presented as the moderator as well. We just arbitrarily choose to present X_2 as the moderator.

$$E(Y|x_1, x_2, \boldsymbol{C}, U_1, U_2) - E(Y|x_1', x_2, \boldsymbol{C}, U_1, U_2) - [E(Y|x_1, x_2', \boldsymbol{C}, U_1, U_2) - E(Y|x_1', x_2', \boldsymbol{C}, U_1, U_2)] \quad (2)$$

The formula (Estimator (2)) above presents an unbiased estimator of the causal interaction effect. The estimator is unbiased because $E(Y|X_1, X_2, \boldsymbol{C}, U_1, \boldsymbol{U_2}) = E[\, Y(x_1, x_2)|\boldsymbol{C}, U_1, \boldsymbol{U_2}]$ by backdoor criterion. However, because U_1 and U_2 are in fact unobserved, only the following estimator can be estimated via observed data,

$$E(Y|x_1, x_2, \boldsymbol{C}) - E(Y|x_1', x_2, \boldsymbol{C}) - [E(Y|x_1, x_2', \boldsymbol{C}) - E(Y|x_1', x_2', \boldsymbol{C})] \quad (3)$$

Because confounders U_1 and U_2 are not properly controlled, the estimator above (Estimator (3)) cannot be generally used to identify the causal interaction effects specified by Formula (1). Nonetheless, we can make two assumptions to make Estimator (3) an unbiased estimator:

Sub-Assumption (1): $X_i \perp U_j|\boldsymbol{C}$, where $(i, j) \in \{(1,2), (2,1)\}$

Sub-Assumption (1) states that conditional on observed covariates, the X_1 is independent from U_2 and X_2 is independent from U_1. In other words, X_1 and X_2 (in the case of this paper, the spouses' genetic measures) do not have any shared confounder. The conditional independence assumption implies that U_2 does not confound the causal effects of X_1 on Y, and, similarly, U_1 does not confound the causal effects of X_2 on Y; take X_1 as an example, this sub-assumption means that $E(Y|x_1, x_2, \boldsymbol{C}, U_1, U_2) - E(Y|x_1', x_2, \boldsymbol{C}, U_1, U_2) = E(Y|x_1, x_2, \boldsymbol{C}, U_1) - E(Y|x_1', x_2, \boldsymbol{C}, U_1)$.

Next, Let $\delta_{bias|x_j,U_i} = E(Y|x_i, x_j, \boldsymbol{C}, U_i) - E(Y|x_i', x_j, \boldsymbol{C}, U_i) - [E(Y|x_i, x_j, \boldsymbol{C}) - E(Y|x_i', x_j, \boldsymbol{C})]$, where $(i, j) \in \{(1,2), (2,1)\}$. This effect refers to the bias in the estimation of the effect of X_i changing from x_i' to x_i on Y due to U_i, given that $X_j = x_j$. Similarly, we can define, $\delta_{bias|x_j',U_i} = E(Y|x_i, x_j', \boldsymbol{C}, U_j) - E(Y|x_i', x_j', \boldsymbol{C}, U_i) - [E(Y|x_i, x_j', \boldsymbol{C}) - E(Y|x_i', x_j', \boldsymbol{C})]$, which refers to the he bias in the estimation of the effect of X_i changing from x_i' to x_i on Y due to U_i, given that $X_j = x_j'$. Then, the second Sub-Assumption is,

Sub-Assumption (2): $\delta_{bias|x_j,U_i} = \delta_{bias|x_j',U_i}$ for all $x_j, x_j' \in X_j$

Sub-Assumption (A2) essentially states that the bias caused by U_1 is constant across levels of X_2, and the magnitude of bias caused by U_2 is constant across levels of X_1. In order words, neither X_1 nor X_2 has an interaction effect with the confounders to the other.

Based on the two assumptions, it can be shown that Estimator (3) can unbiasedly identify the causal interaction effects. To simply the proof, let $\delta_{true|x_2}$ and $\delta_{true|x_2'}$ denotes the true causal effects of X_1 on Y at different levels of X_2. Then, by Assumption (A1)

$$\delta_{true|x_2} = E\{Y(x_1, x_2) - Y(x_1', x_2)|\boldsymbol{C}, U_1, U_2\} = E(Y|x_1, x_2, \boldsymbol{C}, U_1) - E(Y|x_1', x_2, \boldsymbol{C}, U_1)$$

$$\delta_{true|x_2'} = E\{Y(x_1, x_2') - Y(x_1', x_2')|\boldsymbol{C}, U_1, U_2\} = E(Y|x_1, x_2', \boldsymbol{C}, U_1) - E(Y|x_1', x_2', \boldsymbol{C}, U_1)$$

Furthermore,

$E(Y|x_1, x_2, \boldsymbol{C}) - E(Y|x_1', x_2, \boldsymbol{C}) - [E(Y|x_1, x_2', \boldsymbol{C}) - E(Y|x_1', x_2', \boldsymbol{C})]$

$= \delta_{true|x_2} - \delta_{bias|x_2} - (\delta_{true|x_2'} - \delta_{bias|x_2'})$

$= \delta_{true|x_2} - \delta_{bias|x_2} - \delta_{true|x_2'} + \delta_{bias|x_2'}$

$= \delta_{true|x_2} - \delta_{bias|x_2} - \delta_{true|x_2'} + \delta_{bias|x_2}$, by Assumption (A2)

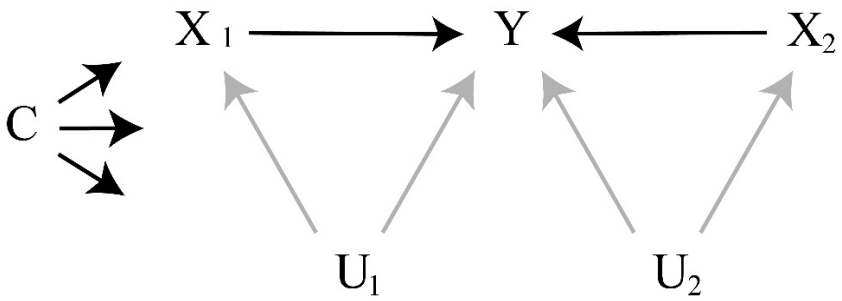
$= \delta_{true|x_2} - \delta_{true|x_2'}$

$= E\{Y(x_1, x_2) - Y(x_1', x_2)|\boldsymbol{C}, U_1, U_2\} - E\{Y(x_1, x_2') - Y(x_1', x_2')|\boldsymbol{C}, U_1, U_2\}$, which is

equivalent to the causal estimand specified by Formula (1).

Therefore, it has been proved that without controlling for unobserved confounders and without identifying causal effects of X_1 or X_2, the causal interaction effect of X_1 or X_2 can still be identified when assumptions (A1) and (A2) hold.

Figure C1. A Directed Acyclic Graph (DAG) of the Assumed Causal Structure.



The directed acyclic graph (DAG) in Figure 1 presents reflects the causal structure given Assumption (A2). The absence of arrows from U_2 to X_1 and U_1 to X_2 reflects Assumption (A2). Sub-Assumption (A2) cannot be reflected in a DAG.

When the causal structure can be properly estimated by linear models, Assumption (A2) can be understood as such: there are neither interaction effects between U_2 and X_1 on Y nor interaction effects between U_1 and X_2 on Y.

Finally, we would like to evaluate Assumption (A1) and (A2) in the scenario of our study. In the scenario of this paper, X_1 and X_2 refers to the couples NEB PGS (G_{0j}, G_{1j}). U_1 and U_2 are confounders to PGS variables, which can be parental genetics or, when there are any siblings, sibling genetics. Suppose that the unobserved confounders are parental genetics, which almost exist for every respondent, Assumption (A1) implies that the wife's parental genetics does not confound the husband's NEB PGS, and vice versa. We think that this assumption is very reasonable. Unless the spouses are close relatives, the genetic effects

for the spouses would be independent from each other. Assumption (A2) implies that the wife's parental genetics do not have any interaction effects with the husband's genetic dispositions, and vice versa. We also think that this is a reasonable assumption. It is hard to come up with substantive mechanisms that could explain such intergenerational cross-spousal interaction between genetic dispositions.

Appendix D. Extended Results Supplementary to the Results Section

Table D1. Summary statistics (mean and standard deviation) of the analytical sample.

Variable	Female		Male	
	Analytical Sample	HRS White/ European Ancestry Genetic Subsample	Analytical Sample	HRS White/ European Ancestry Genetic Subsample
Labor Market Outcome				
Ever worked	0.96 (0.19)	0.95 (0.21)	0.996 (0.07)	0.99 (0.10)
Total years of working	22.74 (11.71)	23.56 (12.65)	35.39 (9.47)	34.54 (10.74)
Number of jobs	1.74 (1.08)	1.73 (1.12)	2.12 (1.21)	2.11 (1.24)
Long-term household income	10.89 (12.83)	8.45 (10.33)	11.13 (12.83)	10.10 (11.42)
Long-term net household wealth	70.23 (117.82)	54.6 (99.46)	71.60 (123.06)	65.84 (120.06)
Long-term labor income	1.90 (2.91)	1.69 (2.68)	3.45 (5.58)	3.20 (5.16)
Non-Labor Market Outcome				
Neuroticism	2.10 (0.56)	2.08 (0.57)	1.95 (0.53)	1.96 (0.55)
Extraversion	3.23 (0.52)	3.21 (0.52)	3.13 (0.51)	3.12 (0.53)
Openness	2.94 (0.51)	2.91 (0.52)	2.94 (0.50)	2.94 (0.51)
Agreeableness	3.65 (0.36)	3.63 (0.38)	3.37 (0.45)	3.37 (0.47)
Conscientiousness	3.40 (0.37)	3.37 (0.39)	3.27 (0.40)	3.27 (0.41)
Depression	1.21 (1.34)	1.42 (1.45)	0.91 (1.10)	1.07 (1.28)
Life Satisfaction	5.04 (1.29)	4.85 (1.34)	4.99 (1.22)	4.86 (1.28)
Loneliness	1.31 (1.37)	1.52 (1.47)	1.14 (1.31)	1.31 (1.42)
Falsification Test				
Years of schooling	13.34 (2.27)	13.12 (2.36)	13.53 (2.69)	13.45 (3.41)
Treatment and Instruments				

Fertility (Number of children)	2.59 (1.49)	2.72 (1.54)	2.59 (1.49)	2.82 (1.59)
Covariates				
Year of Birth	1944 (10.92)	1940.99 (16.60)	1940 (10.89)	1940.55 (15.21)
N	3,282	6,894	3,282	5,196

Data source: The Health and Retirement Study, 2004–2016.

Note: † p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. "Insig." stands for insigificnat. Standard deviations are presented in parentheses. Population refers to all HRS respondents with non-missing values for any of the variables. To save some digits, long-term household income, long-term household wealth, and long-term labor income are reported in ten thousands. Year of birth is rounded to integers. PGS measures have been standardized. Fertility variable is constructed to be the same for males and females. Equal variance is assumed for T-tests. T-tests are not implemented on EA PGS, NEB PGS, and fertility, because they are constructed to have no group differences in means. Summary statistics of PCs are not reported.

Table D2.1 Coefficients of OLS regressions of selected labor market outcomes on realized fertility, males only

Gender	Ever Worked	Total years of working	Number of Jobs	Long-term Household Income	Long-term Net Household Wealth	Long-term Labor Income
Number of Children	-0.001	0.042	-0.015†	-0.030***	-0.215***	0.044
	(0.001)	(0.140)	(0.008)	(0.008)	(0.044)	(0.050)
Male's NEB PGS	0.001	0.496**	-0.007	0.000	-0.027	0.043
	(0.001)	(0.18)	(0.01)	(0.011)	(0.056)	(0.065)
Females' NEB PGS	0.000	-0.027	-0.002	-0.017	0.043	0.005
	(0.0002)	(0.216)	(0.013)	(0.013)	(0.068)	(0.078)
Male's Birthyear	0.000	0.496***	0.001	-0.009***	0.029*	-0.107***
	(0.0002)	(0.038)	(0.002)	(0.002)	(0.012)	(0.013)
Female's Birthyear	0.001	-0.075*	0.003	-0.009***	0.044***	-0.078***
	(0.001)	(0.038)	(0.002)	(0.002)	(0.012)	(0.013)
Male's EA PGS	0.002*	0.71***	-0.005	0.131***	0.35***	0.278***
	(0.001)	(0.198)	(0.012)	(0.012)	(0.062)	(0.071)
Females' EA PGS	0.000	0.141	-0.006	0.091***	0.176**	0.038
	(0.001)	(0.196)	(0.011)	(0.012)	(0.061)	(0.070)
Male's EA PGS	0.000	-0.099	0.011	0.003	-0.003	0.157*
X Female's EA PGS	(0.001)	(0.189)	(0.011)	(0.011)	(0.059)	(0.068)
N	3270	3270	3270	3282	3282	3282

Data source: Health and Retirement Study, 2004–2016.

Note: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control variables are adjusted in all models. Long-term income, wealth, and labor income are rescaled by hyperbolic sine function, and the respective coefficients can be interpreted as semi-elasticity.

Table D2.2 Coefficients of 2SLS regressions of selected labor market outcomes on realized fertility, males only

Gender	Ever Worked	Total years of working	Number of Jobs	Long-term Household Income	Long-term Net Household Wealth	Long-term Labor Income
Number of Children	-0.003	-0.758	-0.504†	-0.054	-0.568	0.217
	(0.010)	(1.663)	(0.295)	(0.100)	(0.522)	(0.594)
Male's NEB PGS	0.001	0.760	0.188†	0.008	0.088	-0.014
	(0.003)	(0.576)	(0.102)	(0.034)	(0.179)	(0.203)
Females' NEB PGS	0.001	0.326	0.228†	-0.007	0.199	-0.071
	(0.005)	(0.762)	(0.135)	(0.046)	(0.24)	(0.273)
Male's Birthyear	0.000	0.488***	-0.016*	-0.009***	0.025†	-0.105***
	(0.0002)	(0.041)	(0.007)	(0.002)	(0.013)	(0.015)
Female's Birthyear	0.000	-0.043	0.019	-0.008	0.058*	-0.085**
	(0.0004)	(0.075)	(0.013)	(0.005)	(0.024)	(0.027)
Male's EA PGS	0.002*	0.728***	0.048	0.131***	0.357***	0.275***
	(0.001)	(0.199)	(0.035)	(0.012)	(0.062)	(0.071)
Females' EA PGS	0.000	0.136	-0.006	0.091***	0.174**	0.039
	(0.001)	(0.193)	(0.034)	(0.012)	(0.061)	(0.069)
Male's EA PGS	0.000	-0.041	-0.024	0.003	-0.002	0.157*
X Female's EA PGS	(0.001)	(0.186)	(0.033)	(0.011)	(0.058)	(0.066)
N	3270	3270	3270	3270	3282	3282

Data source: Health and Retirement Study, 2004–2016.

Note: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control variables are adjusted in all models. Long-term income, wealth, and labor income are rescaled by hyperbolic sine function, and the respective coefficients can be interpreted as semi-elasticity. Coefficients for principal components are omitted.

Table D2.3 Coefficients of OLS regressions of selected labor market outcomes on realized fertility, females only

Gender	Ever Worked	Total years of working	Number of Jobs	Long-term Household Income	Long-term Net Household Wealth	Long-term Labor Income
Number of Children	-0.006**	-1.648***	0.006	-0.029**	-0.210***	-0.124*
	(0.002)	(0.196)	(0.02)	(0.009)	(0.045)	(0.054)
Male's NEB PGS	-0.005†	-0.243	-0.004	0.003	-0.017	-0.020
	(0.003)	(0.251)	(0.026)	(0.011)	(0.057)	(0.07)
Females' NEB PGS	-0.001	-0.048	0.015	-0.020	0.033	0.066
	(0.003)	(0.303)	(0.031)	(0.013)	(0.069)	(0.084)
Male's Birthyear	-0.001	-0.089†	-0.005	-0.011***	0.019	-0.033*
	(0.001)	(0.053)	(0.005)	(0.002)	(0.012)	(0.014)
Female's Birthyear	0.000	0.034	0.036***	-0.008***	0.055***	-0.152***
	(0.001)	(0.055)	(0.006)	(0.002)	(0.012)	(0.015)
Male's EA PGS	0.002	-0.370	0.035	0.126***	0.358***	-0.051
	(0.003)	(0.276)	(0.028)	(0.012)	(0.063)	(0.077)
Females' EA PGS	0.009**	1.276***	-0.008	0.094***	0.184**	0.376***
	(0.003)	(0.273)	(0.028)	(0.012)	(0.062)	(0.076)
Male's EA PGS	-0.003	-0.108	-0.037	0.003	-0.009	-0.020
X Female's EA PGS	(0.003)	(0.264)	(0.027)	(0.011)	(0.06)	(0.073)
N	3231	3231	3231	3282	3282	3282

Data source: Health and Retirement Study, 2004–2016.

Note: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control variables are adjusted in all models. Long-term income, wealth, and labor income are rescaled by hyperbolic sine function, and the respective coefficients can be interpreted as semi-elasticity. Coefficients for principal components are omitted.

Table D2.4 Coefficients of 2SLS regressions of selected labor market outcomes on realized fertility, females only

Gender	Ever Worked	Total years of working	Number of Jobs	Long-term Household Income	Long-term Net Household Wealth	Long-term Labor Income
Number of Children	-0.043	-4.826*	-0.391	-0.009	-0.504	-0.009
	(0.027)	(2.309)	(0.250)	(0.101)	(0.53)	(0.643)
Male's NEB PGS	0.007	0.794	0.141	-0.004	0.079	-0.058
	(0.009)	(0.793)	(0.086)	(0.035)	(0.181)	(0.22)
Females' NEB PGS	0.016	1.388	0.189	-0.029	0.163	0.015
	(0.013)	(1.084)	(0.117)	(0.046)	(0.243)	(0.295)
Male's Birthyear	-0.002*	-0.125*	-0.004	-0.011***	0.015	-0.032*
	(0.001)	(0.06)	(0.007)	(0.002)	(0.013)	(0.016)
Female's Birthyear	0.001	0.162	-0.013	-0.009†	0.067**	-0.157***
	(0.001)	(0.108)	(0.012)	(0.005)	(0.024)	(0.029)
Male's EA PGS	0.003	-0.297	-0.033	0.126***	0.364***	-0.054
	(0.003)	(0.287)	(0.031)	(0.012)	(0.063)	(0.076)
Females' EA PGS	0.008**	1.266***	0.055†	0.094***	0.182**	0.376***
	(0.003)	(0.279)	(0.03)	(0.012)	(0.061)	(0.075)
Male's EA PGS	-0.003	-0.090	-0.037	0.003	-0.008	-0.021
X Female's EA PGS	(0.003)	(0.27)	(0.029)	(0.011)	(0.059)	(0.072)
N	3231	3231	3231	3282	3282	3282

Data source: Health and Retirement Study, 2004–2016.

Note: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control variables are adjusted in all models. Long-term income, wealth, and labor income are rescaled by hyperbolic sine function, and the respective coefficients can be interpreted as semi-elasticity. Coefficients for principal components are omitted.

Table D3.1 Coefficients of OLS regressions of selected non-labor market outcomes on realized fertility, males only

Gender	Neuroticism	Extraversion	Openness	Agreeable-ness	Conscientio-usness	Depression	Life Satisfaction	Loneliness
Number of Children	-0.002	0.010	-0.011	0.015*	-0.004	0.027†	0.013	0.009
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.015)	(0.017)	(0.018)
Male's NEB PGS	-0.007	0.011	0.001	0.014†	-0.003	-0.011	0.041†	-0.052*
	(0.009)	(0.009)	(0.009)	(0.008)	(0.007)	(0.019)	(0.022)	(0.023)
Females' NEB PGS	0.013	-0.019†	-0.005	-0.007	-0.004	0.014	0.006	0.012
	(0.011)	(0.011)	(0.01)	(0.01)	(0.009)	(0.022)	(0.026)	(0.028)
Male's Birthyear	-0.001	-0.005*	-0.007***	-0.007***	-0.003	0.004	-0.011*	0.006
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.005)	(0.005)
Female's Birthyear	-0.003	0.001	-0.001	0.006**	0.002	-0.009*	0.015**	-0.017***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.005)	(0.005)
Male's EA PGS	-0.023*	-0.014	0.043***	-0.019*	0.009	-0.099***	0.067**	-0.019
	(0.01)	(0.01)	(0.009)	(0.009)	(0.008)	(0.021)	(0.024)	(0.025)
Females' EA PGS	-0.017†	0.005	0.024**	0.006	0.014†	-0.090***	0.05*	-0.018
	(0.01)	(0.01)	(0.009)	(0.009)	(0.008)	(0.02)	(0.023)	(0.025)
Male's EA PGS	0.006	-0.010	-0.008	-0.002	0.006	0.052**	-0.023	0.024
X Female's EA PGS	(0.01)	(0.01)	(0.009)	(0.008)	(0.007)	(0.02)	(0.022)	(0.024)
N	3126	3127	3122	3129	3129	3269	3128	3117

Data source: Health and Retirement Study, 2004–2016.

Note: † p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Control variables are adjusted in all models. Coefficients for principal components are omitted.

Table D3.2 Coefficients of 2SLS regressions of selected non-labor market outcomes on realized fertility, males only

Gender	Neuroticism	Extraversion	Openness	Agreeable-ness	Conscientio-usness	Depression	Life Satisfaction	Loneliness
Number of Children	0.033	-0.087	0.075	-0.045	-0.013	0.152	0.216	0.046
	(0.081)	(0.081)	(0.076)	(0.071)	(0.062)	(0.174)	(0.19)	(0.201)
Male's NEB PGS	-0.019	0.044	-0.028	0.034	0.000	-0.053	-0.029	-0.064
	(0.029)	(0.029)	(0.028)	(0.026)	(0.022)	(0.06)	(0.069)	(0.073)
Females' NEB PGS	-0.003	0.025	-0.044	0.020	0.000	-0.042	-0.085	-0.005
	(0.038)	(0.038)	(0.036)	(0.033)	(0.029)	(0.08)	(0.089)	(0.094)
Male's Birthyear	-0.001	-0.006**	-0.006**	-0.008***	-0.004*	0.005	-0.009†	0.006
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.005)	(0.005)
Female's Birthyear	-0.004	0.005	-0.004	0.008**	0.002	-0.014†	0.007	-0.019*
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.008)	(0.009)	(0.009)
Male's EA PGS	-0.024*	-0.012	0.041***	-0.017†	0.009	-0.102***	0.063**	-0.020
	(0.01)	(0.01)	(0.01)	(0.009)	(0.008)	(0.021)	(0.024)	(0.026)
Females' EA PGS	-0.017†	0.005	0.024**	0.006	0.014†	-0.089***	0.05*	-0.018
	(0.01)	(0.01)	(0.009)	(0.009)	(0.008)	(0.02)	(0.023)	(0.025)
Male's EA PGS X Female's EA PGS	0.006	-0.010	-0.008	-0.002	0.006	0.051**	-0.023	0.024
	(0.01)	(0.01)	(0.009)	(0.008)	(0.007)	(0.019)	(0.023)	(0.024)
N	3126	3127	3122	3129	3129	3269	3128	3117

Data source: Health and Retirement Study, 2004–2016.

Note: † p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Control variables are adjusted in all models. Coefficients for principal components are omitted.

Table D3.3 Coefficients of OLS regressions of selected non-labor market outcomes on realized fertility, females only

Gender	Neuroticism	Extraversion	Openness	Agreeable-ness	Conscientio-usness	Depression	Life Satisfaction	Loneliness
Number of Children	-0.010	-0.005	-0.015*	0.013**	-0.006	0.039*	-0.013	0.034†
	(0.008)	(0.007)	(0.007)	(0.005)	(0.005)	(0.018)	(0.018)	(0.019)
Male's NEB PGS	0.006	0.011	0.009	0.002	0.001	-0.019	0.059*	-0.026
	(0.01)	(0.009)	(0.009)	(0.006)	(0.007)	(0.023)	(0.023)	(0.024)
Females' NEB PGS	0.003	0.007	0.004	-0.011	-0.003	-0.037	-0.007	-0.009
	(0.012)	(0.011)	(0.011)	(0.008)	(0.008)	(0.028)	(0.027)	(0.029)
Male's Birthyear	0.001	0.000	-0.002	-0.001	-0.001	-0.001	-0.006	0.001
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.005)	(0.005)	(0.005)
Female's Birthyear	-0.006**	0.000	-0.002	0.000	0.000	-0.005	0.011*	-0.011*
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.005)	(0.005)	(0.005)
Male's EA PGS	-0.022*	0.003	0.022*	-0.011	0.001	-0.114***	0.063*	-0.072**
	(0.011)	(0.01)	(0.01)	(0.007)	(0.007)	(0.025)	(0.025)	(0.027)
Females' EA PGS	-0.027*	0.007	0.033**	-0.015*	0.007	-0.133***	0.079**	-0.058*
	(0.011)	(0.01)	(0.01)	(0.007)	(0.007)	(0.025)	(0.025)	(0.027)
Male's EA PGS	0.002	0.017†	0.016†	0.010	0.006	0.009	-0.025	-0.029
X Female's EA PGS	(0.01)	(0.01)	(0.009)	(0.007)	(0.007)	(0.024)	(0.024)	(0.026)
N	3093	3092	3088	3093	3092	3230	3095	3096

Data source: Health and Retirement Study, 2004–2016.

Note: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control variables are adjusted in all models. Coefficients for principal components are omitted.

Table D3.4 Coefficients of 2SLS regressions of selected non-labor market outcomes on realized fertility, females only

Gender	Neuroticism	Extraversion	Openness	Agreeable-ness	Conscientio-usness	Depression	Life Satisfaction	Loneliness
Number of Children	-0.078	-0.194*	-0.022	-0.105†	-0.052	-0.221	0.091	0.184
	(0.088)	(0.091)	(0.079)	(0.063)	(0.059)	(0.209)	(0.208)	(0.217)
Male's NEB PGS	0.029	0.075*	0.012	0.041†	0.017	0.066	0.024	-0.077
	(0.031)	(0.032)	(0.028)	(0.022)	(0.021)	(0.072)	(0.074)	(0.076)
Females' NEB PGS	0.035	0.096*	0.008	0.044	0.019	0.080	-0.055	-0.079
	(0.042)	(0.044)	(0.038)	(0.03)	(0.028)	(0.098)	(0.101)	(0.105)
Male's Birthyear	0.000	-0.002	-0.002	-0.003	-0.002	-0.004	-0.005	0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.005)	(0.005)	(0.006)
Female's Birthyear	-0.003	0.008*	-0.002	0.006*	0.002	0.005	0.007	-0.017
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.01)	(0.01)	(0.011)
Male's EA PGS	-0.021†	0.007	0.023*	-0.009	0.001	-0.108***	0.061*	-0.075**
	(0.011)	(0.011)	(0.01)	(0.008)	(0.007)	(0.026)	(0.025)	(0.027)
Females' EA PGS	-0.027*	0.007	0.03***	-0.015†	0.007	-0.134***	0.079***	-0.058*
	(0.011)	(0.011)	(0.009)	(0.008)	(0.007)	(0.025)	(0.024)	(0.026)
Male's EA PGS	0.002	0.017	0.016†	0.010	0.006	0.010	-0.025	-0.030
X Female's EA PGS	(0.010)	(0.011)	(0.009)	(0.007)	(0.007)	(0.024)	(0.024)	(0.025)
N	3093	3092	3088	3093	3092	3230	3095	3096

Data source: Health and Retirement Study, 2004–2016.

Note: † p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001. Control variables are adjusted in all models. Coefficients for principal components are omitted.

Appendix E: Robustness Analysis Results Concerning Potential Pleiotropy Bias

Table E1. Selected coefficients of 2SLS regressions of selected non-labor market outcomes on realized fertility, with or without additional adjustments.

Gender	Neuroticism	Extraversion	Depression	Life Satisfaction
		No Additional Adjustments		
Male	0.033	-0.087	0.152	0.216
	(0.081)	(0.081)	(0.174)	(0.19)
Female	-0.078	-0.194*	-0.221	0.091
	(0.088)	(0.091)	(0.209)	(0.208)
		With Additional Adjustments		
Male	0.034	-0.085	0.136	0.231
	(0.081)	(0.082)	(0.174)	(0.192)
Female	-0.084	-0.192*	-0.239	0.091
	(0.088)	(0.090)	(0.209)	(0.208)

Data source: Health and Retirement Study, 2004–2016.

Note: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Cluster standard errors at the spousal level are presented in parentheses. Control variables are adjusted in all models. Neuroticism and Extraversion are measured by Big-5 personality scale and range from 1 to 4. Depression is measured by the short CESD scale provided and ranges from 0 to 8. Life-satisfaction is measured by Diener's measure of life satisfaction and ranges from 1 to 7. "No Additional Adjustments" refer to the results reported in the main text (Table 5). "With Additional Adjustments" denotes the 2SLS results when the spousal PGS variables for the respective outcome and their interaction term have been controlled (e.g., spousal PGS variables for neuroticism and the interaction between the spousal neuroticism PGS variables).