PAA 2014 Annual Meeting, Boston, MA, May 1-3, 2014

# Using genetic markers as instrumental variables to the link between education and fertility

Nicola Barban[1], Melinda Mills[1], Jornt Mandemakers[1], Harold Snieder[2]

[1]University of Groningen, The Netherlands. Department of Sociology

[2]University of Groningen, University medical center Groningen, The Netherlands. Department of Epidemiology

Corresponding authors: n.barban@rug.nl, m.c.mills@rug.nl

**Draft prepared for the PAA 2014 Annual Meeting. Please do not quote.**

## SHORT ABSTRACT

The relationship between education and fertility has been a central focus within demography and related social sciences. Higher education is often associated with higher age at first birth and lower number of children, especially among women. The goal of this paper is to dig deeper into the relationship between education and fertility and explore the causal relationship by using genetic markers as instrumental variables. Specifically, by using the genetic markers for educational attainment from a recent GWAS (genome-wide association study), we attempt to unravel the causal relationship between education and age at first birth (AFB), number of children ever born (NEB) and childlessness. Our results using data from three large samples in contemporaneous western populations (LifeLines, TwinsUk and HRS) show that education is causally linked to higher age at first birth and decrease in childlessness but not to fertility. We suggest that the observed association between education and fertility is mainly affected by unobservable factors.

## INTRODUCTION

The relationship between education and fertility has been a central focus within demography and related social sciences (Rindfuss et al. 1980; 1984; 1996; Rodgers et al. 2008; Skiberekk et al. 2012, Begall & Mills 2013). There has been a massive delay in the age at first birth across many Western societies since 1970, which is now between 28-29 years (Mills et al. 2011). The majority of research has focused on three primary mechanisms linking education and fertility. The mechanism that is examined most often is how achieving higher education (particularly of women) operates to postpone the timing of fertility and particularly the age at first birth (Bulatao & Casterline 2001; Balbo et al. 2013). Longer educational enrolment can also limit the quantum of fertility by leaving a shorter reproductive period to have more children. A second mechanism is one of reverse causation where early fertility may impede higher educational attainment (Upchurch et al. 2002). Last, fertility and education may be influenced by common unobservable factors, such as personality, fertility preferences and attitude towards family.

There are two shortcomings within existing literature, which the current study will address. First, the majority of existing research has generally only examined associations not causality, with only a few exceptions (e.g., Rodgers et al. 2008, Leon, 2004, Monstad et al. 2008, Fort et al. 2011, Berhman and Kohler 2012, ). Since a majority of the reported associations between education and fertility have not been subject to deeper empirical scrutiny, it may be that the causal relationships we assume are either tenuous or artifacts. Second, there may be endogeneity problems, since it is not always apparent whether fertility or education is the cause or effect or whether other unobserved factors impact both variables.

The goal of this paper is to dig deeper into the endogeneity in the relationship between education and fertility and explore an additional mechanism and causal relationship by using genetic markers as instrumental variables. Specifically, by using the genetic markers for educational attainment from a recent GWAS (genome-wide association search) (Rietveld et al. 2013), we will attempt to unravel the causal relationship between education, age at first birth (AFB) and number of children ever born (NEB).

We first summarize previous research that has examined the relationship between education and fertility, followed by discussion of potential theoretical mechanisms that explain the relationship between education and fertility for women and men. After describing the three datasets that are employed in this paper we turn to a more detailed description of the use of genetic variants as instruments by adopting an instrumental variable (IV) or Mendelian Randomization (MR) approach. This is followed by a discussion of how the statistical assumptions for the IV approach are met or potentially violated. The preliminary results are then discussed and interpreted followed by a discussion of strengths and limitations of this approach, new knowledge that we have gained and future promising lines of research based on this approach.

## PREVIOUS RESEARCH AND THEORETICAL MECHANISMS

**The relationship between education and fertility**

Prior to the second demographic transition and overall fertility decline in Western countries, higher education and status was related to having higher numbers of children (e.g., Betzig 1986). Skirbekk (2008: 157-8) demonstrated that since early in the 20th century, the relationship between education and fertility has always been negative and that higher education depressed fertility by around 26% for all periods worldwide. This study likewise demonstrated that the effect of education on fertility was much stronger for women than men for recent historical periods: "in the period 1990-2006 for the whole world, highly educated women have 29.9% fewer children than women with low education, while highly educated men have 11.6% fewer children than low educated women" (Skirbekk 2008: 158). When discussing the impact of education on fertility, it is essential therefore to divide the discussion by gender due to the higher differential mechanisms underlying the impact of education on women versus men.

In historical research, male fertility is positively associated with hierarchy and status with higher status males have better access to more sexual partners and status resources (Skirbekk 2008). The relationship between having high status and preference for smaller family sizes has been described by some as evolutionary maladaptive behavior (Dieckmann & Ferriere 2004). There are several factors that contribute to fertility limitation among the higher educated that relate to both men and women. First, the highly educated have been shown to have a better knowledge and practice of contraceptive use (Cleland 2001; Kanazawa 2003). Second, as described in more detail shortly, women who are higher educated are more likely to participate in the labour market which in turn postpones childbirth. Education increases perspective earnings and therefore the opportunity costs of leaving the labor market to have and raise children (Becker, 1965). Third, the highly educated are more likely to be more secularized and individualized, which would lower fertility considering the strong body of research that demonstrates a relationship between religion and higher fertility (e.g., Sacerdote & Glaeser 2001; Schellekens & van Poppel 2006). Fourth, those with higher social status, such as having higher levels of education have also been shown to have higher consumption aspirations, meaning that children themselves would imply high opportunity costs that would infringe upon their own lifestyle (Becker and Lewis 1973, Lutz, Skirbekk & Testa 2006).

Finally, we anticipate a gendered difference in the relationship between education and number of children ever born (NEB) and particularly childlessness. Childlessness has been shown to be higher in men than women (Hakim 2005), but this appears to vary across different national contexts. Higher educated men in professional occupations are more likely to be childless in the U.K. (Kiernan 1989; Hakim 2005) where the opposite is the case in Australia and Italy (Parr 2007;). Although higher education increases the prevalence of childlessness in women in the U.S., it is not a significant predictor for men (Waren & Pals 2013). It appears that in many contexts, however, childlessness among men is associated with low education and worse social status and health whereas for women the opposite is true (Barthold, Myrskylä, & Jones, 2012; Nettle & Pollet, 2008; Kravdal & Rindfuss, 2008). It appears therefore, when we examine childlessness (i.e., NEB = 0) it may go in a different direction for each sex. In other words, women with a poorer social status and lower education will be less likely to be childless

(and have more children), with an opposite effect for men. Men with a lower education would be more likely to be childless. Voluntary childlessness thus has a very different gendered social status component influencing highly educated career women and lower educated comparatively less successful men. We now turn to sex-specific differences that we anticipate between women and men in relation to education and fertility.

**Education and women's fertility**

There is considerable documentation and research that demonstrates the association between female education and age at first birth (AFB) and number of children ever born (NEB). Early research in the United States identified the strong inverse relationship between education and fertility, with education postponing AFB (Rindfuss et al. 1980; Martin 2000). This is in particular related to women's increased levels of education, especially in gains to obtaining higher college and University degrees (Rindfuss, Morgan & Swicegood 1988; Rindfuss, Morgan & Offutt 1996; Martin 2000).

There are various reasons why the increased educational attainment of women leads to childbearing delay. First, it is generally untenable to balance the student and mother roles since both are time-intensive. This results in women opting to remain in education and fulfill their educational aspirations and postponement of parenthood (Mills et al. 2011). A second mechanism is that higher educated women are more likely to pursue occupational careers that involve considerably more responsibility, higher wages and greater authority and autonomy (Amuedo-Dorantes and Kimmel 2005). There is considerable evidence that those who opt to postpone childbearing do so in order to first establish their career (Happel et al. 1984). Furthermore, women in higher occupational positions have higher opportunity costs to step out of the labor market and are cognizant of the 'motherhood penalty' on their careers (Kravdal & Rindfuss 2007; Skirbekk, Kohler & Prskawetz 2004). Early childbearing has been consistently linked to a motherhood wage penalty, particularly for higher-educated women and those in professional occupations experiencing a substantial increase in earnings if they postpone parenthood (Gustafsson 2001; Amuedo-Dorantes & Kimmel 2005; Gustafsson & Kalwij 2006; Miller 2010; O'Donoghue et al. 2011). In a detailed econometric analysis, for instance, Gustafsson (2001) demonstrated that women's career planning was the main explanation for postponement, a finding replicated in more recent studies in Ireland (O'Donoghue et al. 2011) and the U.S. (Miller 2010). Researchers have estimated a 7% motherhood wage penalty per child (Budig and England 2001) and how a year of delayed motherhood increased women's earnings by 9%, work experience by 6% and average wage rates by 3% (Miller 2010). Dropping out of the labor market results in the loss of wages, but also the missed opportunity for additional training and specifically the depreciation of job-specific human capital (see for the U.S. Sweden Albrecht et al. 1999). This relates to a third aspect which is the high costs of children. Those with higher education may be more likely to delay childbearing until they can afford children, since prospective parents with higher educational levels have steeper age-income profiles and higher consumer aspirations (Becker 1991; Happel et al. 1984). Hence,

those with higher education levels might delay childbearing until they feel they can 'afford' them. All of these arguments predict a later age at first birth for women.

Previous research has demonstrated that a later age at first birth is associated with lower levels of completed childbearing, although results remain mixed. Not only is the biological window for childbearing shortened by postponement, but the longer an individual remains childless the more likely they might acquire interests that compete with the time required for the parental role (Kohler et al. 2002). There are suggestions, however, that this inverse relationship between age at first birth and cumulative childbearing is weakening. Martin (2000) demonstrated widening educational differentials in the timing of births in the U.S. In comparison to lower educated women, both the rate of first and second births after the age of 30 increased during the 1970s-1990s for women with a 4-year college degree. Sobotka (2004) also demonstrated that the lowest-low fertility rates observed in Europe were likely temporary, since higher educated women would eventually recuperate and have children at a later age. In other words, although the higher educated might postpone having a first child, they might be more likely to 'compress' childbirth and have additional children more rapidly. A recent study of Norwegian men and women by Kravdal and Rindfuss (2008) found that higher educated women postpone first births but that the cumulative impact of late motherhood on higher-order birth rates (i.e., second or third births) has disappeared in Norway. They attributed this to family-friendly ideologies, which translate into policies such as better access to high-quality and convenient daycare.

**Education and men's fertility**

The majority of research on education and fertility has overwhelmingly focused almost exclusively on women. Or as Presser (1997: 303) noted "women are given special (and usually exclusive) attention in fertility research." There are several reasons for why this is the case, which are interrelated with the fact that different underlying mechanisms predict the impact of obtaining higher education on fertility outcomes for men versus women.

The first factor relates to the strong biological differences in male and female fertility, which is related to postponing, having and caring for children. Women's more primary role in fertility is largely related to the focus on their ability to control fertility via contraceptive use, a primary driver of the second demographic transition (van de Kaa 1987; Lesthaege 1995). Although contraceptive use, decision making about fertility and care for children is clearly a couple process, predominant explanations focus on women. Some have argued that this focus on women is not related to differences in biology, but rather the social context and norms of demographic assumptions in Western countries when this discipline emerged (Greene & Biddlecom 2000). The fact that women are more biologically impacted by childbearing by actually bearing children and needing to leave the labor market, combined with social norms and accompanying institutional policies (e.g., maternity leave, childcare) means that the majority of policies, but also the impact of children has been on women. A second related aspect of biological differences is that men having a longer reproductive span than women. They

therefore not only have the possibility to postpone having children, but also have a longer reproductive window to do so.

A second primary difference is the disparate relationship and consequences that education and fertility has for women versus men. The consequence of having children on men's careers is comparatively minimal. The primary reason for this is that – with the exception of some Scandinavian countries – men do not exit the labor market to care for children for extended periods of time to the extent that women do. They therefore do not face the same opportunity costs and fatherhood wage penalties associated with children that women face, which were outlined previously. Having children may in fact have the opposite effect for men's career advancement. In contrast to women, research has shown that having children in a martial union may positively impact career advancement for men. This may be due to favoritism in promotion to higher positions by managers for these breadwinners .

A third related factor is that results have rather consistently shown that when we look at aggregated comparisons, men have an overall lower number of children ever born (NEB) than women. The central reason for lower male fertility appears, however, to be related to fertility reporting and a discrepancy in the number of children that is recognized by male respondents. A primary reason that men's fertility in relation to education has been understudied may therefore also be a methodological issue. It is common that men are simply not asked fertility questions in many medical and social science surveys that demographers used. Another issue is that when these questions are included, they are often discounted or not examined due to incomplete reporting. A now seminal study by Rendall et al. (1999) evaluated the reporting of men's retrospective fertility histories and updated panel histories in the Panel Study of Income Dynamics (PSID) in the U.S. and the UK's British Household Panel Survey (BHPS). They found that in the retrospective histories, between one third to one half of men's nonmarital births and births from previous marriages were missing from their retrospectively reported histories. This is also related to the fact that although men may bear and raise more children than women, children generally live longer with their mother than father. Additional research has likewise demonstrated that men often have less or no contact with their own children that live in another household.

A fourth aspect key to understanding men's fertility is the differential function of nuptuality behavior and the role that assortative mating plays. First, in comparison to women, certain groups of men may have a larger number of children due to second 'nests' or marriages where they establish a second family. Due to the shorter reproductive age span of women, this ability to start an additional nest is not feasible. However, as previous results have demonstrated (Rendall et al. 1999), men that have more complicated relationship histories with nonmarital children, children from the first partner, own children from a second union and stepchildren may underreport their fertility. A second aspect is that due to educational assortative mating (REF) men are more often married to women that are either of equal educational level or slightly below. For this reason, the postponement of women's fertility, particularly in the higher educational groups, would also operate to postpone men's timing of AFB. This indeed appears to the be case with vital statistics and survey data showing us that

mean age of father's AFB is increasing over time. It is likewise important to note that men on average generally partner with women that are around 2 years younger, which automatically means that they will have a slightly higher age at first birth. This means that in many Western societies, men have later reproduction than women and the bulk of male fertility takes place between age 30 to 35.

## DATA AND ANALYTICAL METHODS

### Data

A central problem in this type of research that includes genetic markers has been the lack of replicability of results (Duncan & Keller 2011; Ioannidis 2005). For this reason, it is essential to not only undertake this analysis in one dataset, but replicate it across a series of different data. For this article we replicate the analysis across three different datasets: the Dutch LifeLines Biobank and Cohort, TwinsUK registry and the US-based Health and Retirement Survey (HRS). For the preliminary version of this paper we currently only include the preliminary analyses for the first two datasets.

The Dutch *LifeLines Biobank and Cohort* is a three-generation longitudinal family design of 165,000 individuals from the Northern provinces of the Netherlands of where around 13,000 individuals have been genotyped (Stolk et al. 2008). By 2013, two waves will be collected and available, with additional waves each year. LifeLines has the advantage of a large sample, ability to separate non-genetic and genetic familial transmission, single and multiple SNPs and direct haplotype assessment.

The *TwinsUK registry* includes both monozygotic and dizygotic twins who voluntarily participated in surveys of the TwinsUK registry. The project involves more than 12,000 individuals and 60,000 observations since 1992 and represents the largest adult twin registry in the UK. It has primarily been conducted to answer questions of aetiology and epidemiology, but also contains demographic information. The sample contains genotypic information for 4905 individuals, 381 men and 4524 women. Since men are underrepresented in this sample, we restrict our analysis only on female.

The *Health and Retirement Survey (HRS)* is a longitudinal study with a representative sample of more than 26,000 Americans over the age of 50. Genotypic data are available for 12,500 respondents. HRS includes detailed information on childbearing history and educational achievements and can be linked to the genotypic data after approval from the National Institute of Health. Genotypic data are available at the National Institute of Health GWAS repository (dbGap). Since we are currently waiting from approval from the dbGaP, this current version of the paper does not include any results from HRS.

### Analytical methods: Genetic variables as instrumental variables

The previous studies linking education and fertility in demography and the social sciences often aim to make causal inferences using standard observational survey or register data. This is highly problematic since it is rarely possible to identify which variable of education or fertility is the cause or the effect or whether another third unobserved factor affects both variables. Various techniques have been introduced to deal with the direction or relationship of causation, confounding variables and endogeneity through twin designs (e.g., Rodgers et al. 2008), reforms in the educational system (e.g. Fort et al. 2011) or experimental data on schooling costs (Duflo et al., 2011).

The adoption of a behavioral genetics approach of twin analysis such as Rodgers et al. (2008) focusing on the heritability of cross-trait analysis (level of education, intelligence, fertility) comes closer to identifying causal factors. In the study by Rodgers et al. (2008), cognitive ability and education were used as predictors and AFB as the outcome variable. The underlying assumption is that these relationships may mediate each other or be mediated by other relationships and that these relationships can be the result of either genetic or (social) environmental causal factors. A core statistic is heritability, defined as the proportion of the total variance of a particular trait that is explained by genetic factors and is generally calculated from twin studies by comparing intra-pair monozygotic (identical) versus non-zygotic twins. Rodgers et al. (2008) reported the heritability for education as 0.48 while they did not find any heritability for age at first birth.

In this study, we propose to extend the field further by directly addressing the issues of reverse causation and unobserved factors confounding. This is achieved by adopting what is termed the Instrumental Variable (IV) approach from econometrics (Angrist et al. 1996) or Mendelian randomization in genetics (MR) (Davey Smith & Ebrahim 2003). Bi-directional Mendelian Randomization (MR) exploits the random assignment of an individual's genotype that occurs at conception, making the IV application with genetic instruments equivalent to MR (Davey et al. 2003; Wehby et al. 2008). Since individuals have an equal probability that either parental allele has been transmitted to them, individuals with different genotypes should not different systematically in any other respect.

MR is associated with Randomized Control Trails (RCT) familiar to social scientists (Lawlor et al. 2008). The central underlying problem of existing research is that the observational data that demographers primarily use is not randomized. If individuals could be randomly assigned into a treatment group (i.e., many years of education) and a control group (i.e., fewer years of education) such as in a randomized control trial (RCT), establishing causality would be easier. Random assignment would ensure that individuals in the treatment group who are 'exposed' to more years in education would be matched to an equal number of similar individuals in the control group with the only exception being that they have had fewer years of education. Since RCT is obviously not feasible with education, we propose to adopt an instrumental variable (IV) approach. Since we exploit the random assignment of an individual's genotype that occurs at conception (Davey et al. 2003). MR is closely linked to RCT and the quasi-experimental approaches sometimes used in the social sciences, making the IV application with genetic instruments equivalent to MR (Wehby et al. 2008). MR can therefore be

used to make causal inferences about the effects of modifiable (non-genetic) risk factors on different outcomes (Lawlor et al. 2008).

Specifically, the approach entails that we introduce a third variable (i.e., the instrument) of the genetic markers that have been recently identified for educational attainment for the two outcomes of years of education and college education (Rietveld et al. 2013). These genetic variant instruments are assumed to (partly) determine the level of the 'treatment', which is observed educational attainment (years, college educated) but does not have a direct or indirect effect on our outcomes of age at first birth (AF) and number of children ever born (NEB) other than through its effect on the observed educational level. This instrument is thus exploited to enable us to make causal inferences about the effect of the level of education on fertility. In demography and economics, this is often referred to as endogeneous treatment or selection into treatment. In other words, it may be that certain individuals 'select' or are 'selected into' their educational level (i.e., treatment) via other mechanisms such as intelligence, gendered field of education choices that lead to certain occupations amenable to combing work and family (Begall & Mills 2013), personality (Jokela 2013), or other unobserved factors. It may be that even after we adjust these baseline characteristics in our statistical models, these choices are also related to fertility outcomes. For this reason we cannot ignore the impact of these factors.

**Mathematical specification of the model**

To formally represent the model we will simplify the explanation by describing only one genetic marker instrument, one educational attainment outcome (e.g., years of education) and one fertility outcome (e.g., age at first birth). Let $S$ denote the random variable representing the fertility outcome, $A$ the random variable for educational level and $Z$ the genetic marker as instrumental variable for educational level. Specifically, Z is the number of risk allele previously associated with the educational outcome. As individuals have two alleles, randomly assigned from each parent, $Z_i$ takes value 0,1 or 2. Thus, $Z_i$= 0 means that individual $i$ does not carry any risk allele; $Z_i$ = 1 implies that she/he does carry one risk allele (or is heterozygous with respect to the risk allele); $Z_i$ = 2 means that individual $i$ does carry two risk allele (or is homozygous with respect to the risk allele. $A_i(z)$ is the educational level for individual $i$ when the instrument is set to $z$ since only one of the treatment assignments is ever observed for any one individual. It thus follows that $A_i(0)$ when we observe $i$'s educational level when $i$ does not carry any risk allele for educational attainment, $A_i(1)$ when $i$ does carry one risk allele for educational attainment and $A_i(2)$ when $i$ does carry two risk alleles for educational attainment. Turning to the treatment variable, let $S_i(a,z)$ be the fertility outcome for individual $i$ that is observed if the educational level treatment variable is set to $a$ and the instrument is set to $z$ by external intervention. This means that $A_i(z)$ is the potential treatments and and $S_i(a,z)$ is the potential outcome. The causal effect – or in other words the individual treatment effect – can be written as $S_i(a\dot{z}) - S_i(a,z)$, where $a$ is a certain baseline value. In light of the exclusion restriction, which is discussed in more detail shortly, we thus write $S_i(a\dot{z}) - S_i(a)$ with the causal estimand of interest specified as:

$$E[S_i(a) - S_i(a)]$$

The assumptions under which this causal effect is identified and potential violations of these conditions are discussed in the part of this section. We first, however, turn to a description of the measurement of variables used in this analysis.

**Polygenic allele score**

A major limitation of using common genetic variants as instrumental variables is the limited effect size and scarce predictive accuracy of single SNPs. Rietveld et al. (2013) show that the independent SNPs identified in a large collaborative consortium (sample size of over 100,000 individuals) explain approximately 0.02% of the variance in educational attainment. This problem is also known as missing heritability (Manolio et al. 2009) since individual genes account only for a small part of the heritability of diseases, behavior and other phenotypes. Using single genes as instrumental variables may lead to the problem of "weak instrument bias". This bias is in the direction of the observational confounded association, and its magnitude depends on the strength of association between genetic instrument and phenotype (Burgess and Thompson, 2011). A possible solution to this problem is to calculate an allele score and use the constructed variable as an instrument. An allele score is a single variable summarizing multiple genetic variants associated with a risk factor. It is calculated as the total number of risk factor-increasing alleles for an individual (unweighted score), or the sum of weights for each allele corresponding to estimated genetic effect sizes (weighted score). An allele score can be used in a Mendelian randomization analysis to estimate the causal effect of the risk factor on an outcome reducing the weak instrument bias. Burgess and Thompson (2013) show that the estimates obtained using allele scores as instrumental variables are generally robust to misspecification of the allele score, but not to instrumental variable violations**.** Allele scores enable valid causal estimates with large numbers of genetic variants. In this study, we adopt both strategies. First, we present an estimation of the causal effect of education using four SNPs as instrumental variables. Second, we construct a weighted allele score based on the complete association results published by the educational attainment consortium. In particular, we select, genetic variants in independent loci that have a statistical association of p-value$<10^{-06}$. The list of complete association results is publicly available on the website of the Social Science Genetic Association Consortium (www.ssgca.org). The independent loci are selected using a stepwise selection procedure in GCTA. Weights are proportional to the meta-analytical effects size.

All the models are estimated via OLS and two-stages least-squares (2SLS). Fertility outcomes are analyzed as continuous measures. In order to have similar diagnostic tests across the different outcomes, childlessness is estimated via a linear probability model (LPM).

**Measurement of variables**

*Fertility outcomes*

In this paper we consider two measures of fertility. The first outcome is the number of children ever born (NEB) during the fertile period. To avoid right censoring problems, we restrict our sample to women 45 or older at the moment of the last observation and men 55 or older. We consider the number of live births and not the number of pregnancies. We exclude, therefore, stillbirths and miscarriages. For the same reason, we consider twins as multiple births. The second outcome is the age at first birth. The analysis includes all the respondents that ever

became parent. Childless individuals are thus excluded from the analysis. The third outcome analyzed is childlessness (NEB=0) at age 45 for women and age 55 for men.

*Measure of educational attainment*

We use the same measures of educational attainment used in the GWAS educational attainment: years of education and college attainment (Rietveld et al, 2013). To have a consistent measure of education across the different studies, years of education is expressed in US-equivalent years of education.

*Genetic variants*

We instrument education with a set of (single nucleotide polymorphisms) SNPs that have been shown to be associated with educational attainment in a recent GWAS meta-analysis. Specifically, we consider those SNPs that have been genome-wide significant associated (p-value $<5x10^{-8}$) to education in the discovery phase of Rietveld et al. (2013). For years of education we consider the following four SNPs: rs9320913; rs3783006; rs8049439; rs13188378. All these SNPs are located in different chromosomes and are not in linkage disequilibrium. All SNPs are located on different chromosomes. As additional instrumental variables, we constructed genetic scores calculated using the number of "risk" alleles (by "risk", we mean alleles positively associated with educational attainment). For a description of the genetic markers used as instrumental variables, please refer to Table 1.

*Allele scores*

Allele scores are calculated as the total number of risk factor-increasing alleles for an individual weighted for the effect sizes of the single independent genetic variants. We used a p-value< of $10^{-06}$ as a threshold for selecting the genetic variants included in the allele score. The selection procedure is based on a step-wise procedure based on the complete meta-analysis results. Linkage disequilibrium is based on HapMap. Table 2 reports the list of SNPs included in the polygenic allele score. In the analysis, we use Z-standardized scores.

*Covariates*

As additional covariates we include in our analysis the year of birth of the respondents. To account for historical changes in fertility, we included in our analysis a quadratic trend calculated on year of birth. Both the sample analyzed are comparable in terms of educational level and birth year (see Table 3)

Table 1: SNPs description and allele frequency in the LifeLines and TwinsUK sample.

| SNP rs number | Chr | nearest gene | ref. allele* | other allele* | Homozygous for ref. allele | Heterozygous | Homozygous for other allele | Homozygous for ref. allele | Heterozygous | Homozygous for other allele |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | LIFELINES | | | TwinsUK | |
| rs9320913 | 6 | LOC100129158 | **A** | C | 28.04 | 49.44 | 22.52 | 28.57 | 50.51 | 20.92 |
| rs3783006 | 13 | STK24 | **C** | G | 24.09 | 52.13 | 23.78 | 31.17 | 48.12 | 20.71 |
| rs8049439 | 16 | ATXN2L | **T** | C | 19.60 | 49.15 | 31.25 | 16.51 | 46.56 | 36.93 |
| rs13188378 | 5 | SLCO6A1 | A | **G** | 0.00 | 0.17 | 99.83 | 0.56 | 11.27 | 88.17 |

*allele with positive effect on education are in bold and underlined

Table 2: Descriptive statistics, educational outcomes and covariates. LifeLines and TwinsUK.

| | LIFELINES | | | | | | | | | TwinsUK | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Men | | | Women | | | Pooled | | | Women | | |
| | mean | sd | N | mean | sd | N | mean | sd | N | mean | sd | N |
| Years of education | 13.60 | 3.89 | 5463 | 13.12 | 3.70 | 7554 | 13.32 | 3.79 | 13017 | 12.48 | 2.84 | 3811 |
| % college education | 0.31 | | 5590 | 0.26 | | 7795 | 0.28 | | 13385 | 21.07 | | 3811 |
| Number of children ever born (NEB) | 2.42 | 1.25 | 1484 | 2.29 | 1.21 | 4734 | 2.32 | 1.22 | 6218 | 2.08 | 1.23 | 3170 |
| Age at First Birth (AFB) | 29.01 | 4.53 | 4472 | 26.62 | 4.25 | 6641 | 27.59 | 4.52 | 11113 | 25.58 | 4.67 | 3093 |
| % Childless (age 45 women, 55 men) | 9.18 | | 1525 | 10.53 | | 4822 | 10.20 | | 6347 | 9.37 | | 4524 |
| Year of Birth | 1959.9 | 1.13 | 5590 | 1960.5 | 1.09 | 7795 | 1960.2 | 1.11 | 13385 | 1951.4 | 1.29 | 4524 |

## Tests of violation of statistical assumptions

To assess our models, it is essential to briefly examine the statistical assumptions in MR and determine if and how these conditions are met in order to clarify that the genetic instrument we have used in an appropriate one. We are cognizant of the implicit statistical assumptions and issues related to the validity of the instrument (Didelez & Sheehan 2007).

First, a broader critique is related to the general *validity of the use of genetic markers* as instruments, which is based assumption that there is an equal probability that either parental allele is randomly transmitted to offspring. It has been shown that at the population level, genetic variants are generally unrelated to most socioeconomic and behavioral outcomes, which are generally closely associated with one another as confounders (e.g., Davey Smith et al. 2008; Lawlor et al. 2008). Based on this research we can therefore be relatively confident that since genes are randomly assigned at meiosis, that individuals with different genotypes will not systematically differ in any other respect.

A second statistical condition for a valid causal interpretation of the IV estimand is the problem of *independence*, which is related to population stratification. A problem of population stratification would refer to the situation where the relationship between a genetic variant for education and the fertility outcome is observed differently over different sub-populations. The most common type of population stratification is ethnicity, since allele frequencies can differ across ethnic groups. If this ignored it may be that we are actually observing differences in genetic makeup across different subgroups and not an actual causal relationship, thus violating the independence assumption. One way to counter this problem is to only examine a more homogeneous population. The genetic markers identified by Rietveld et al. (2013) include only individuals of European ancestry in our model, which is achieved by adjusting for what is referred to as the principal components from genome wide association searches. Within our three samples, we will also test for this violation by examining the outcomes both within those of European ancestry and then separately analyzing these results by different ethnic sub-populations that we are able to identify. In addition, to control for population stratification, we include in our regression models the first ten principal components. Principal components capture population stratification and geographical differences in the sample (Price et al. 2006).

A third potential problem is whether our models met the *exclusion restriction* assumption and if not, whether this invalidates the instrument. The exclusion restriction may be violated by four central situations. First, biological processes could bias causal mechanisms due to canalization. Genes associated with educational attainment might have a biological function in regulating fertility. Rietveld et al. (2013) do, however, point to various potential biological pathways that link the genetic variants to educational attainment. Since these mechanisms only appear to be related to educational attainment, intelligence and other cognitive outcomes, there does not appear to be any clear association with fertility outcomes. For this reasons since the aforementioned biological mechanisms only appear to impact education and not our fertility outcomes, we assume that this assumption will not be violated for that reason. Second, genes associated with educational attainment might be in linkage disequilibrium (LD) with genetic variants that are associated with fertility. Although Mendel's second law states that the inheritance of one trait is independent of the inheritance of another, it appears that some variants may in fact be co-inherited, referred to as Linkage Disequilibrium (LD). LD could therefore potentially bias our estimates if our genetic marker instrument ($Z$) is in LD with another locus that only directly affects our fertility outcome ($S$). To limit the potential violation due to genes in LD with the causal variants, we included in the allele score only a limited set of independent variants (p-value$<10^{-06}$). A third and related potential problem is pleiotropy, which

refers to the situation where one genetic variant has multiple functions and is problematic for our model if our genetic markers for educational attainment are related to other genetic variants that affect fertility. Our approach will thus be incorrect if the pleiotropic effect influences the fertility outcomes directly. If however, our genetic marker instrument (*Z*) is in LD only with another locus that only affects the modifiable risk factor of educational attainment (*A*), this assumption is not violated. To test for pleiotropy, we engage in a bivariate association analysis with GCTA and test if the phenotypic correlation between education and fertility is due to genetic correlation. The analysis is restricted to the genes included as instruments in the analysis.

A final concern related to the violation of the exclusion restriction is that certain behaviors or conditions related to the parents may impact the genotype. This relates to epigenetic effects, which examines the more dynamic nature of genes and how chemical bases of gene expression are influenced by for example, DNA methylation. It may be that certain socio-environmental exposures such as maternal behavior during pregnancy (e.g., intrauterine effects of alcohol consumption) or later parental behavior might impact our fertility outcomes via the influence that they have on their offspring's behavior. Cameron et al. (2005) demonstrated that the maternal behavior of rats (grooming, nursing) affected gene expression among offspring in the brain regions that control defensive and reproductive behaviors. Another recent study on rates demonstrated epigenetic silencing or the inhibition of DNA methylation as a mechanism underlying the neuroendocrine control of female puberty (Lomniczi et al. 2013). It could be, for instance, that parents who carry the lower education alleles experience poorer labor market outcomes and lower perceived benefits of obtaining education. Particularly for women, this may mean that they opt to pull out of the labor market and remain at home and have more children. This may affect parent's own behavior and the preferences for their children's education. If this is the case, the exclusion restriction would be violated. However, the actual extent or seriousness of this violation would be highly dependent upon the actual effect sizes of the genetic variants. Given our study, we do not anticipate that the parental responses will be strong enough to actually impact our results enough to impact the outcome.

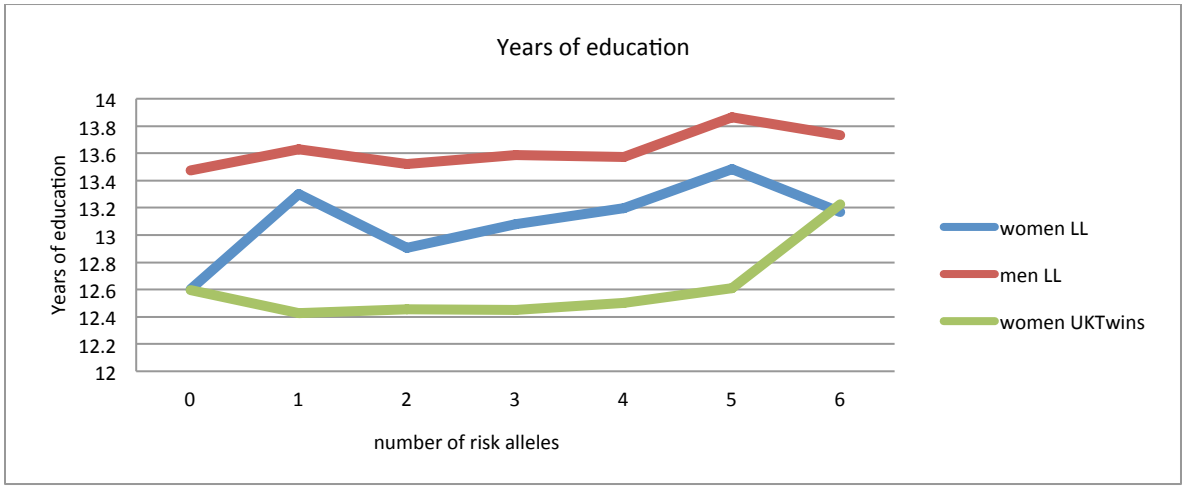
A fourth potential violation is referred to as the *weak instrument problem*, which is often a general problem within IV studies (Angrist & Krueger 1991). MR can only be employed with genetic variants that have been shown affect the risk factor of interest. In other words, we first engage in an analysis that demonstrates that the genetic markers for educational attainment actually predict educational level in our samples. If this is not the case, we are faced with the problem of a weak association, which in turn results in a biased IV estimate. It may be, however, that the instrument itself is not weak since we know that many complex and distal outcomes in the social sciences require very large sample sizes since the average causal effect of the risk factor on the fertility outcome might be very small (Angrist & Krueger 1991; Rietveld et al. 2013). We adopted two strategies to avoid the *weak instrument problem* due to small effect size of genetic variants. First, we use allele scores as instrumental variables. The combination of multiple genetic variants in an allele score has the function of increasing the association in the first stage of two-stage least-square model (2SLS). As discussed in the previous section, allele scores are robust to allele misspecification and represent an efficient solution for causal inference (Burgess and Thompson 2013). Second, we replicate the analysis in different

independent samples. Moreover, we will run an additional analysis in the pooled sample in order to increase the sample size and the statistical power of the first stage association[2].

**PRELIMINARY RESULTS**

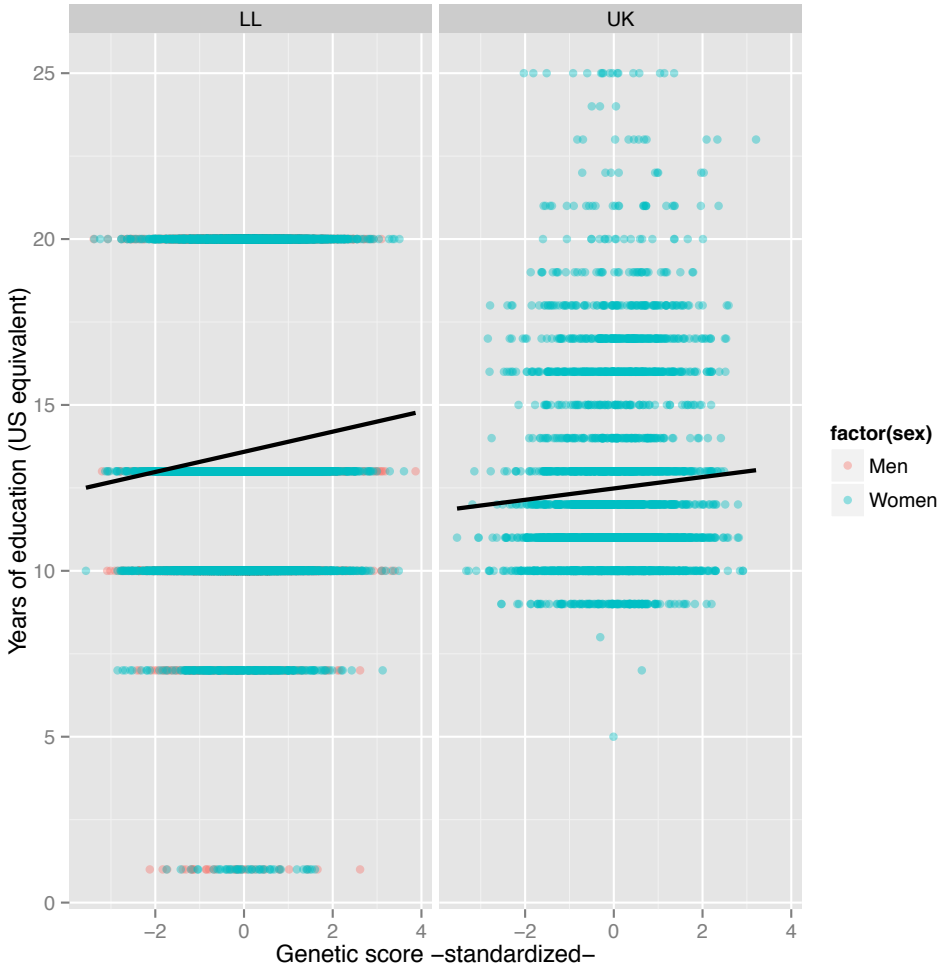
A common limitation of candidate genes studies is the lack of a robust association between the genetic markers and the trait of interest. To solve this problem, we use those genetic markers that were found be associated with educational attainment in a recent large GWAS meta-analysis (Table 1). Although the effect of single SNPs is relatively small, the association has been validated in independent samples and has biological foundation. To check if this association is consistent in LifeLines and TwinsUK, we calculated a simple genetic score, counting the number of alleles that were previously found associated with educational outcomes. As shown in Figure 1, the number of "risk" alleles is positively associated with both years of education and college attainment. The relationship is positive both in men and women. Figure 2, shows the association between the weighted polygenic allele score and education in the two samples.

*Figure 1: Years of education (US-equivalent) by the number of "risk" alleles. Men and Women in Lifelines TwinsUK.*




[2] Pooled analysis is not present in this version of the paper since HRS data are not yet available.

*Figure 2: Years of education (US-equivalent) by genetic score (Z-standardized). Men and Women in Lifelines and TwinsUK.*




**OLS and IV analysis**

We begin our analysis by examining the OLS association between years of education, college attainment and number of children and age at first birth.

Our results show that years of education have a different impact on men and women. Men who stay longer in education have in average a higher number of children (Table 3). The relationship is reverse among women. Higher education is associated with lower number of children both in LifeLines and TwinsUK. The association is stronger among women in TwinsUK than in the Dutch sample (Table 6). The IV analysis using genetic variants (both single SNPs and allele score), does not confirm the previous analysis. Once we take into account the possible endogeneity of education, we do not find evidence that supports a significant effect of education on total fertility. The results for women and for the total sample are supported by a strong

association between the instrumental variables and education (see First stage statistics in table 3[3])

*Table 3: Summary results OLS and IV regression on number of children ever born (NEB). LifeLines.*

| | Women | | | Men | | | Pooled | | |
|---|---|---|---|---|---|---|---|---|---|
| | **OLS** | **IV genetic score** | **IV SNPs** | **OLS** | **IV SNPs** | **IV genetic score** | **OLS** | **IV SNPs** | **IV genetic score** |
| Years of education | -0.008~ | 0.064 | -0.036 | 0.020** | 0.104 | 0.028 | 0.001 | -0.188 | 0.050 |
| | (0.004) | (0.057) | (0.125) | (0.007) | (0.271) | (0.125) | (0.004) | (0.170) | (0.053) |
| Year of birth | 0.358~ | 0.233 | 0.407 | -1.082* | -1.218 | -1.094 | 0.319~ | 0.382 | 0.235 |
| | (0.213) | (0.303) | (0.366) | (0.490) | (0.826) | (0.685) | (0.174) | (0.278) | (0.229) |
| Year of birth$^2$ | -0.047* | -0.043 | -0.049~ | 0.123* | 0.133 | 0.124 | -0.043* | -0.029 | -0.041~ |
| | (0.021) | (0.028) | (0.029) | (0.057) | (0.085) | (0.075) | (0.018) | (0.028) | (0.023) |
| SEX (women) | | | | | | | -0.024 | -0.299 | 0.049 |
| | | | | | | | (0.041) | (0.254) | (0.087) |
| Constant | 1.887*** | 1.532~ | 2.028* | 4.503*** | 3.786 | 4.439* | 1.897*** | 4.099* | 1.509~ |
| | (0.524) | (0.797) | (0.978) | (1.046) | (2.663) | (1.837) | (0.432) | (1.994) | (0.832) |
| Observations | 4754 | 4754 | 4754 | 1468 | 1468 | 1468 | 6222 | 6222 | 6222 |
| Hansen p-value (over-identification) | | | 0.695 | | 0.149 | | | 0.980 | |
| p-value of under-identification statistic | | 0.000 | 0.102 | | 0.893 | 0.029 | | 0.322 | 0.000 |
| First stage F Statistic | | 30.554 | 39.668 | | 4.735 | 5.949 | | 30.905 | 39.690 |

Standard errors in parentheses
~ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[3] [3] As a rule of thumb, the F-statistic of a joint test should be bigger than 10 in case of a single endogenous regressor.

The relationship between education and age at first birth is always positive. In both women and men, higher education and college attainment is associated with higher age at first birth. The effect is higher among women, where an increase of one year in education is associated with 0.3 years in age at first birth in The Netherlands and 0.5 years in UK (Table 7). Although these associations are strongly significant in all the models, the IV analysis using single genetic variants, does not confirm the previous analysis. However, once we use a genetic score as instrumental variable, we find evidence that supports a significant effect of education on age at first birth. The result is stronger among men (0.5 years) than women (0.26 years). Our results suggest that there is a causal effect of education on fertility tempo in the The Netherlands.

*Table 4: Summary results OLS and IV regression on age at first birth (AFB). LifeLines*

| | Women | | | Men | | | Pooled | | |
|---|---|---|---|---|---|---|---|---|---|
| | **OLS** | **IV SNPs** | **IV genetic score** | **OLS** | **IV SNPs** | **IV genetic score** | **OLS** | **IV SNPs** | **IV genetic score** |
| Years of education - US equivalent - | 0.319*** | -0.481 | 0.260~ | 0.152*** | 1.734 | 0.594** | 0.247*** | 0.137 | 0.405** |
| | (0.013) | (0.411) | (0.154) | (0.015) | (1.363) | (0.223) | (0.010) | (0.342) | (0.126) |
| Year of birth | 1.230** | 2.970** | 1.360* | 2.682*** | 1.144 | 2.252*** | 1.851*** | 2.007*** | 1.615*** |
| | (0.395) | (1.043) | (0.548) | (0.534) | (1.616) | (0.605) | (0.318) | (0.599) | (0.384) |
| Year of birth - squared - | -0.031 | -0.118~ | -0.038 | -0.151** | -0.073 | -0.129* | -0.080** | -0.087* | -0.069* |
| | (0.034) | (0.065) | (0.041) | (0.047) | (0.107) | (0.051) | (0.028) | (0.037) | (0.030) |
| Sex (women) | | | | | | | -2.354*** | -2.446*** | -2.221*** |
| | | | | | | | (0.080) | (0.297) | (0.132) |
| Constant | 16.200*** | 19.578*** | 16.451*** | 16.576*** | 0.731 | 12.147*** | 19.952*** | 20.903*** | 18.608*** |
| | (1.126) | (2.324) | (1.430) | (1.487) | (13.999) | (2.722) | (0.907) | (3.073) | (1.446) |
| Observations | 6580 | 6580 | 6580 | 4378 | 4378 | 4378 | 10958 | 10958 | 10958 |
| Hansen p-value (overidentification) | | 0.829 | | | 0.950 | | | 0.254 | |
| p-value of underidentification LM statistic | | 0.022 | 0.000 | | 0.745 | 0.000 | | 0.065 | 0.000 |
| First stage F Statistic | | 37.736 | 48.884 | | 8.977 | 12.849 | | 43.973 | 57.648 |

Standard errors in parentheses. First 10 PCAs included in the models.
~ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The effect of education on childlessness differ with gender. Education is associated with greater proportion of childless women, while the opposite is observed among men. However, the IV

analysis shows that, once we take into account possible sources of endogeneity, the causal relationship has a reverse sign. Higher education women are less likely to remain childless. This counterintuitive results shows how important is to take into account possible sources of counfoundness in the analysis in order to investigate causal relationship between educational attainment and fertility.

*Table 6: Summary results OLS and IV regression on Childlessness. LifeLines*

| | Women | | | Men | | | Pooled | | |
|---|---|---|---|---|---|---|---|---|---|
| | **OLS** | **IV SNPs** | **IV genetic score** | **OLS** | **IV SNPs** | **IV genetic score** | **OLS** | **IV SNPs** | **IV genetic score** |
| Years of education - US equivalent - | 0.003** | -0.014 | -0.030* | -0.004* | 0.021 | -0.023 | 0.001 | 0.002 | -0.028* |
| | (0.001) | (0.035) | (0.015) | (0.002) | (0.061) | (0.032) | (0.001) | (0.037) | (0.014) |
| Year of birth | -0.484*** | -0.457*** | -0.429*** | -0.261* | -0.301 | -0.230 | -0.472*** | -0.469*** | -0.451*** |
| | (0.055) | (0.090) | (0.074) | (0.116) | (0.190) | (0.155) | (0.043) | (0.059) | (0.056) |
| Year of birth$^2$ | 0.047*** | 0.046*** | 0.045*** | 0.021 | 0.024 | 0.018 | 0.045*** | 0.045*** | 0.047*** |
| | (0.005) | (0.007) | (0.007) | (0.013) | (0.020) | (0.017) | (0.004) | (0.006) | (0.006) |
| Sex (female) | | | | | | | 0.019~ | 0.020 | -0.024 |
| | | | | | | | (0.010) | (0.054) | (0.023) |
| Constant | 1.290*** | 1.377*** | 1.456*** | 0.885*** | 0.672 | 1.046* | 1.244*** | 1.225** | 1.552*** |
| | (0.134) | (0.249) | (0.194) | (0.247) | (0.605) | (0.446) | (0.108) | (0.424) | (0.211) |
| Observations | 4754 | 4754 | 4754 | 1468 | 1468 | 1468 | 6222 | 6222 | 6222 |
| Hansen p-value (overidentification) | | 0.777 | | | 0.663 | | | 0.958 | |
| p-value of underidentification LM statistic | | 0.102 | 0.000 | | 0.893 | 0.029 | | 0.322 | 0.000 |
| First stage F Statistic | | 30.554 | 39.668 | | 4.735 | 5.949 | | 30.905 | 39.690 |

Standard errors in parentheses. First 10 PCAs included in the models.
~ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

One of the possible criticisms of this approach is the validity of using genetic marker as instrumental variable for the analysis. The genetic markers used in this analysis seem to be consistently associated to educational outcomes. The first stage F-statistics indicated in table 4 5 and 6 indicate the strength of the association between the instrumental variables and

education[4]. Table A2,A3 and A4 in the appendix report the first stage results showing the association between the different genetic markers and educational attainment. For a detailed analysis of the validity of the instruments, please refer to the complete tables reported in the appendix.

**Replication**

To validate the results obtained in the LifeLines sample, we replicated the analysis using the TwinsUK register. Our analysis confirms the link between education and age at first birth (p-value 0.08). High education seems to be causally associated to postponement of age at first birth. We cannot replicate, however, the effect of education on childlessness. A possible reason is lack of statistical power. For this reason, we aim in replicating the results in an additional sample (HRS) and examine the association between education and fertility in a combined sample.

*Table 7: Summary results OLS and IV regression. TwinsUK*

| | NEB | | | AFB | | | Childlessness | | |
|---|---|---|---|---|---|---|---|---|---|
| | **OLS** | **IV SNPs** | **IV genetic score** | **OLS** | **IV SNPs** | **IV genetic score** | **OLS** | **IV SNPs** | **IV genetic score** |
| Years of education - US equivalent - | -0.039*** | -0.018 | 0.054 | 0.509*** | 0.783 | 1.401~ | 0.015*** | 0.149 | 0.022 |
| | (0.010) | (0.215) | (0.148) | (0.040) | (0.910) | (0.799) | (0.002) | (0.126) | (0.025) |
| Year of birth | -0.166 | -0.153 | -0.137 | -2.847*** | -2.653** | -2.669*** | 0.097*** | 0.166* | 0.099*** |
| | (0.287) | (0.353) | (0.294) | (0.645) | (0.843) | (0.761) | (0.026) | (0.075) | (0.027) |
| Year of birth - squared - | -0.002 | -0.008 | -0.013 | 0.298*** | 0.255* | 0.221* | -0.013*** | -0.029~ | -0.014*** |
| | (0.032) | (0.046) | (0.036) | (0.066) | (0.124) | (0.101) | (0.002) | (0.015) | (0.003) |
| Constant | 3.321*** | 3.129 | 2.288 | 25.990*** | 22.757* | 16.256~ | -0.222** | -1.791 | -0.306 |
| | (0.647) | (2.578) | (1.758) | (1.616) | (10.560) | (8.869) | (0.076) | (1.446) | (0.283) |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Observations | 2734 | 2160 | 2734 | 2537 | 1998 | 2537 | 3811 | 3000 | 3811 |
| Hansen p-value (overidentification) | 1821 | 1518 | 1821 | 1824. | 1502 | 1824 | 2554 | 2123 | 2554 |
| p-value of underidentification LM statistic | | 0.353 | | | 0.760 | | | 0.383 | |
| First stage F Statistic | | 0.372 | 0.004 | | 0.509 | 0.017 | | 0.756 | 0.000 |
| Observations | | 4.970 | 8.137 | | 7.431 | 11.825 | | 21.865 | 34.877 |

Standard errors in parentheses. Standard errors corrected for clustering. First 10 PCAs included in the models.
$^{\sim}\ p < 0.10,\ ^{*}\ p < 0.05,\ ^{**}\ p < 0.01,\ ^{***}\ p < 0.001$

## Discussion

In this paper, we integrate genetic markers to analyze the causal relationship between education and fertility. A simple association analysis shows that higher education is associated to children at later age, smaller number of children among women and higher probability of childlessness. In this paper, we show that these results are misleading since the relationship between education and fertility is very likely to be affected by unobserved factors. The availability of novel genetic results on the biological basis of educational attainment make us possible to investigate to what extent this education has a causal effect on fertility. Mandelian randomization is very similar to a randomized control trial since it is based on the assumption that there is an equal probability that either parental allele is randomly transmitted to offspring. Our analysis shows that education does not influence significantly fertility. The association that is found through OLS regression is thus very likely to be the effect of other unobservable factors that influence both the outcomes. On the other hand, we show that education has a significant effect in delaying the age at first birth both in the Netherlands and in the UK. In addition, we observe that education has the unexpected effect of reducing childlessness in the Netherlands. Although this effect is not replicated in the UK, this shows that a simple association analysis is misleading. Education has no causal effect on reducing total fertility of increasing the number of childless couples but only on delaying age at first birth. In addition, men and women with higher educational attainment are less likely to remain childless at the end of their reproductive career. One possible explanation is that high educated men and women compress their childbearing after completing education and obtained an higher social status. Moreover, high education might influence the ability to succeed in the marriage market increasing assortative mating and probably reducing divorce or couple instability, that represents an important factor of decrease in fertility.

To our knowledge, this is the first study that use molecular genetics to identify the link between education and fertility. Previous studies on this topic that used genetic information were conducted with twins (Kohler et al. 2011). However, twin studies rely on assumptions that are

not needed on this study. An additional innovation of this study is that we extend the analysis to men. Most of the previous works, in fact, looked only at women neglecting the role of men education in fertility. Last, we believe that testing this relationship on multiple samples is also a key innovation of this study.

We are aware that this approach has several limitations. First, it is impossible to statistically test all the violations of the IV analysis. The same genes that affect education, in fact, can affect other mechanisms related to fertility (pleyotropy). This can violate the exclusion restriction of the IV estimator. However, for this study, we restrict the analysis to genetic markers that are robustly associated with education level. As shown by Rietveld et al. (2013), these markers have specific biological pathways connected to cognitive development. We do not have any evidence that the same SNPs are also involved on human reproduction. Second, not all the educational attainment genetic variants predicted educational level. This is not surprising given the fact that educational attainment is a complex trait that is mostly socially based. However, with a sufficiently large sample it is possible to increase the statistical power in order to have robust estimates. Third, educational level may not be the best predictor. Educational field has been shown to have consequences for the timing and number of children (Lappegård & Rønsen 2005; Hoem et al. 2006; Martin-Garcia & Baizan 2006; Begall & Mills 2010). Van Bavel (2010), for instance, demonstrated that four features of study disciplines were key to reproductive decision making: the expected starting wage, steepness of the earning profile, attitudes towards gendered family roles and gender composition. Across 21 European countries, the postponement of first birth was the most pronounced for women who had studied in male-dominated disciplines and least postponed by those in the more female-dominated fields. The starting wage and steepness of the earning profile were also associated with postponement. Last, this study does not take into account that the link between education and fertility may change in different context. For example, in historical population or in development countries, education can still have a strong effect in reducing fertility. A limitation of the IV approach is the lack of external validity. For this reasons we apply the analysis on different contexts, using three independent samples. This study applies to contemporaneous population and specifically to The Netherlands, United Kingdom and US (with the inclusion of the HRS).

As further analysis, we will replicate the analysis on the HRS, using a large longitudinal US nationally representative study. In addition, we will investigate the use of alternative instrumental variables using genetic predictive scores.

**REFERENCES**

Angrist, Joshua D., and Krueger, Alan B. (1991) "Does Compulsory School Attendance Affect Schooling and Earnings?" QJ.E. 106 : 979-1014.

Balbo, N., F. Billari & M. Mills (2013). Fertility in Advanced Societies: A review of research. *European Journal of Population*.

Barthold, J. A., Myrskylä, M., & Jones, O. R. (2012). Childlessness drives the sex difference in the association between income and reproductive success of modern Europeans. *Evolution and Human Behavior*, *33*(6), 628–638.

Becker, Gary S. 1991. A Treatise on the Family. Cambridge,MA: HarvardUniversity Press.

Mills, M., & Begall, K. (2010). Preferences for the sex-composition of children in Europe: A multilevel examination of its effect on progression to a third child. *Population Studies*, *64*(1), 77–95. doi:10.1080/00324720903497081

Begall, K., & Mills, M. (2013). The Influence of educational field, occupation and occupational sex segregation on fertility in the Netherlands. *European Sociological Review*.

Betzig L. 1986 Despotism and Differential Reproduction: *A Darwinian View of History*

England, P., & Budig, M. J. (2001). The wage penalty for motherhood. *American Sociological Review*, *66*(2), 204–225.

Bulatao, Rodolfo A. and John B. Casterline, eds. (2001). Global Fertility Transition. *Population and Development Review*, Supplement to Vol. 27. New York: Population Council.

Burgess, S., & Thompson, S. G. (2011). Avoiding bias from weak instruments in Mendelian randomization studies. *International Journal of Epidemiology*, *40*(3), 755–764.

Burgess, S., Butterworth, A., & Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology*, *37*(7), 658–665. doi:10.1002/gepi.21758

Burgess, S., & Thompson, S. G. (2013). Use of allele scores as instrumental variables for Mendelian randomization. *International Journal of Epidemiology*, *42*(4), 1134–1144. doi:10.1093/ije/dyt093

Cameron, et al. 2005. The programming of individual differences in defensive responses and reproductive strategies in the rat through variations in maternal care. *Neuroscience and Biobehavioral Reviews* 29:843–865.

Davey Smith, G. & S. Ebrahim. 2003. 'Mendelian Randomization': Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease? *Inter J of Epid* 32: 1-22.

Didelez, V. & N.S. Sheehan. 2007. Mendelian Randomization as an Instrumental Variable Approach to Causal Inference, *Statistical Methods in Medical Research*, 16: 309-330.

Dieckmann U, Ferriere R (2004) Adaptive dynamics and evolving biodiversity
In Evolutionary Conservation Biology, R. Ferriere, U. Dieckmann, D. Couvet (eds)
Cambridge University Press, Cambridge, UK, pp.188-224 (2004)

Amuedo-Dorantes, C., & Kimmel, J. (2005). The motherhood wage gap for women in the United States: The importance of college and fertility delay. *Review of Economics of the Household*, *3*(1), 17–48.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. (2011). "Education, HIV, and Early Fertility: Experimental Evidence from Kenya." Working Paper.

Duncan, L. E., & Keller, M. C. (2011). A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry. *American Journal of Psychiatry*, *168*(10), 1041–1049.

Fort, M, N. Schneeweis and R. Winter-Ebmer (2011), More Schooling, More Children: Compulsory Schooling Reforms and Fertility in Europe, Working Paper 1105, Department of Economics, Johannes Kepler University of Linz.

Greene, M. E., & Biddlecom, A. E. (2000). Absent and problematic men: Demographic accounts of male reproductive roles. *Population and Development Review, 26,* 81-115.

Gustafsson, S. 2001. "Optimal Age at Motherhood. Theoretical and Empirical Considerations on Postponement of Maternity in Europe." *Journal of Population Economics* 14:225–47.

Gustafsson S, Kalwij A. Education and Postponement of Maternity. Economic Analyses of Industrialized Countries. European Studies of Population, Volume 15, Dordrecht, The Netherlands: Kluwer Academic Publishers, Springer, 2006.

Hakim, C. (2005), `Sex differences in work-life balance goals', pp 55-79 in Work-Life Balance in the Twenty-First Century (ed) D Houston, London: Palgrave Macmillan

Happel, S.K., J.K. Hill and S.A. Low, 1984, An economic analysis of the timing of childbirth, *Population Studies* 38, 298-311.

Ioannidis, J.P.A. (2005). Why most published research findings are false, *PLoS Medicine*, 2*8*(8): e124.

Kiernan, K.E. (1989) Who remains Childless? *Journal of Biosocial Scien*ce, Vol. 20, No. 4, 387-398.

Kohler, H. P., Rodgers, J. L., & Christensen, K. (2002). Between nurture and nature: the shifting determinants of female fertility in Danish twin cohorts. *Biodemography and Social Biology*, *49*(3-4), 218–248.

Kohler, Hans-Peter, Jere R. Behrman and Jason Schnittker (2011). Social science methods for twins data:Integrating causality, endowments and heritability. *Journal of Biodemography and Social Biology*, 57(1)

Kravdal, Ø., R. R. Rindfuss. 2007 "Changing relationships between education and fertility – a study of women and men born 1940-64." *Mimeo, Economics Dept. Univ of Oslo*.

Lappegård, T., & Rønsen, M. (2005). The Multifaceted Impact of Education on Entry into Motherhood. *European Journal of Population / Revue européenne de Démographie*, *21*(1), 31–49.

Lawlor, D.A., et al. 2008. Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology, *Statistics in Medicine* 27: 1133-63.

Leridon, H. (2004) Can assisted reproduction technology compensate for the natural decline in fertility with age? A model assessment. *Human Reproduction* **19**, 1548–1553 (2004).

Lutz,W. Skirbekk,V. Testa,M.R. (2006) The Low Fertility Trap Hypothesis: Forces that May Lead to Further Postponement and Fewer Births in Europe IIASA Interim Report IR-06-017

MacKinnon, David P., Ghulam Warsi, and James H. Dwyer. 1995. "A Simulation Study of Mediated Effect Measures." *Multivariate Behavioral Research* 30:41–62.

Martin SP. Diverging fertility among U.S. women who delay childbearing past age 30. (2000) *Demography*; 37:523–533.

Warren B Miller, David E Bard, David J Pasta, & Joseph Lee Rodgers. (2010). Biodemographic Modeling of the Links Between Fertility Motivation and Fertility Outcomes in the NLSY79. *Demography*, *47*(2), 393–414.

Mills, M., Rindfuss, R.R., McDonald, P., & Velde, E. te (2011). Why do people postpone parenthood? Reasons and social policy incentives. *Human Reproduction Update*, 17(6), 848-860.

Nettle, D., & Pollet, T. V. (2008). Natural selection on male wealth in humans. *The American Naturalist*, *172*(5), 658–666.

O'Donoghue C, Meredith D, O'Shea E. (2011) Postponing maternity in Ireland. *Cam J Econ* 2011;35:59– 84.

Parr, N. (2010). Satisfaction with life as an antecedent of fertility. *Demographic Research*, *22*, 635–662.

Presser, P. B. (1987). Work shifts of full-time dual-earner couples: Patterns and contrasts by sex of spouse. *Demography*, 24, 99–112.

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.

Rendall, Michael S., Lynda Clarke, H. Elizabeth Peters, Nalini Ranjit, and Georgia Ver-ropoulou. 1999. "Incomplete Reporting of Men's Fertility in the United States and Britain: A Research Note." *Demography* 36 (1): 135–44.

Rindfuss, R. R., Bumpass, L., & St John, C. (1980). Education and fertility: Implications for the roles women occupy. *American Sociological Review*, 431–447.

Ronald R. Rindfuss, Craig St. John and Larry L. Bumpass (1984) Education and the Timing of Motherhood: Disentangling Causation *Journal of Marriage and Family* 46:4, 981-984

Rindfuss RR, Morgan SP, Swicegood G. First Births in America. Changes in the Timing of Parenthood. Berkeley, CA: University of California Press, 1988.

Rindfuss, R.R., et al. 1996. Education and Changing Age Pattern of American Fertility: 1963-1989, *Demography*, 33; 277-90.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., et al. (2013). GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science*, *340*(6139), 1467–1471.

Rodgers, J.L. et al. 2008. Education and cognitive ability as direct, mediating, or spurious influences on female age at first birth: behaviour genetic models fit to Danish Twin data, *American Journal of Sociology* 114(Supp): S202-32.

Schellekens, J. and F. van Poppel (2006), Religious differentials in marital fertility in The Hague (Netherlands) 1860-1909. *Population Studies*, 60(1), pp. 23-38.

Skirbekk, V., Kohler, H., & Prskawetz, A. (2004). Birth month, school graduation, and the timing of births and marriages. *Demography*, *41*(3), 547–568.

Skirbekk, V. (2008). Fertility trends by social status. *Demographic Research*, *18*(5), 145–180.

Skirbekk, V., & Samir, K. C. (2012). Fertility-Reducing Dynamics of Women's Social Status and Educational Attainment. *Asian Population Studies*, *8*(3), 251–264.

Sacerdote, B. and E.L. Glaeser (2001). "Education and Religion, National Bureau of Economic Research, working paper no. 8080,

Sobotka T. Is lowest-low fertility in Europe explained by the postponement of childbearing? *Popul Dev Rev* 2004a;30:195–220.

Stolk, R.P. et al. 2008. Universal risk factors for multifactorial diseases. *Eur J of Epid*, 23(1): 67-74.

Upchurch, D. M., Lillard, L. A., & Panis, C. W. (2002). Nonmarital childbearing: Influences of education, marriage and fertility. *Demography*, 39, 311 – 329.

Van Bavel J. Choice of study discipline and the postponement of motherhood in Europe: the impact of expected earnings, gender composition and family attitudes. *Demography* 2010;47:439–458.

Wray, N. R., Yang, J., Ben J Hayes, Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, *14*(7), 507–515. doi:doi:10.1038/nrg3457

Waren, W., & Pals, H. (2013). Comparing characteristics of voluntarily childless men and women. *Journal of Population Research*, 1–20.

Wehby, G.L., Ohsfeldt, R.L. & Murray, J.C. 2008. 'Medelian randomization' equals instrumental variable analysis with genetic instruments. *Statistics in Medicine*. 27(15): 2745-9.

## APPENDIX

Table A1: List of SNPs used in the polygenic risk score

| Chromosome | SNP name | Reference allele | Allele frequency HapMap | Beta | SE | joint p-value |
|---|---|---|---|---|---|---|
| 1 | rs12741781 | T | 0.69 | 0.02 | 0.004 | 5.75E-07 |
| 2 | rs1606974 | A | 0.08 | 0.026 | 0.005 | 9.50E-08 |
| 2 | rs984363 | A | 0.46 | -0.022 | 0.004 | 1.82E-08 |
| 2 | rs12615145 | T | 0.31 | 0.022 | 0.004 | 3.82E-08 |
| 2 | rs11884495 | A | 0.48 | 0.022 | 0.004 | 3.82E-08 |
| 2 | rs13008687 | T | 0.75 | -0.021 | 0.004 | 1.53E-07 |
| 2 | rs7600417 | A | 0.86 | -0.025 | 0.005 | 5.75E-07 |
| 2 | rs10929169 | T | 0.14 | -0.027 | 0.005 | 6.69E-08 |
| 3 | rs6809216 | A | 0.21 | 0.022 | 0.004 | 3.81E-08 |
| 4 | rs2955259 | A | 0.64 | 0.02 | 0.004 | 5.75E-07 |
| 5 | rs6882046 | A | 0.75 | -0.021 | 0.004 | 1.53E-07 |
| 6 | rs4461735 | A | 0.75 | -0.02 | 0.004 | 6.59E-07 |
| 6 | rs1056667 | T | 0.56 | 0.023 | 0.004 | 4.82E-09 |
| 6 | rs545787 | A | 0.46 | -0.019 | 0.004 | 9.51E-07 |
| 6 | rs1487441 | A | 0.46 | 0.025 | 0.004 | 4.14E-10 |
| 7 | rs7788657 | T | 0.06 | 0.035 | 0.007 | 5.75E-07 |
| 9 | rs1010587 | T | 0.22 | 0.02 | 0.004 | 5.75E-07 |
| 11 | rs11602566 | A | 0.29 | -0.021 | 0.004 | 1.53E-07 |
| 12 | rs247929 | C | 0.41 | -0.02 | 0.004 | 5.75E-07 |
| 12 | rs2066955 | A | 0.24 | 0.021 | 0.004 | 1.53E-07 |
| 13 | rs3783006 | C | 0.39 | 0.022 | 0.004 | 3.82E-08 |
| 14 | rs11620952 | A | 0.72 | 0.021 | 0.004 | 1.53E-07 |
| 15 | rs11855635 | A | 0.65 | 0.021 | 0.004 | 1.53E-07 |
| 15 | rs8034147 | T | 0.55 | 0.02 | 0.004 | 5.75E-07 |
| 16 | rs8049439 | T | 0.62 | 0.021 | 0.004 | 1.53E-07 |
| 18 | rs1187220 | T | 0.32 | -0.02 | 0.004 | 5.75E-07 |
| 21 | rs17646111 | A | 0.25 | -0.022 | 0.004 | 3.81E-08 |

Table A2. First stage regression NEB, LifeLines

| | Women | | Men | | Pooled | |
|---|---|---|---|---|---|---|
| | **IV SNPs** | **IV genetic score** | **IV SNPs** | **IV genetic score** | **IV SNPs** | **IV genetic score** |
| Year of birth | 1.734* | 1.851** | 1.721 | 1.487 | 0.617 | 0.666 |
| | (0.712) | (0.710) | (1.760) | (1.758) | (0.589) | (0.588) |
| Year of birth - squared - | -0.054 | -0.065 | -0.140 | -0.112 | 0.048 | 0.044 |
| | (0.071) | (0.071) | (0.204) | (0.204) | (0.060) | (0.060) |
| rs9320913_A | 0.105 | | -0.112 | | 0.049 | |
| | (0.079) | | (0.165) | | (0.072) | |
| rs3783006_C | 0.093 | | -0.048 | | 0.061 | |
| | (0.083) | | (0.163) | | (0.074) | |
| rs8049439_T | 0.138~ | | 0.126 | | 0.133~ | |
| | (0.082) | | (0.162) | | (0.074) | |
| rs13188378_A | 0.898~ | | -0.123 | | 0.629 | |
| | (0.540) | | (2.698) | | (1.116) | |
| Genetic score - standardized- | | 0.312*** | | 0.238* | | 0.288*** |
| | | (0.056) | | (0.109) | | (0.050) |
| SEX | | | | | -1.469*** | -1.475*** |
| | | | | | (0.142) | (0.142) |
| Constant | 2.851 | 4.685** | 8.649 | 8.864* | 9.394*** | 10.784*** |
| | (2.031) | (1.736) | (6.475) | (3.748) | (2.728) | (1.461) |
| Observations | 4754 | 4754 | 1468 | 1468 | 6222 | 6222 |
| r2_a | 0.080 | 0.085 | 0.035 | 0.039 | 0.067 | 0.072 |
| F | 30.554 | 39.668 | 4.735 | 5.949 | 30.905 | 39.690 |

Standard errors in parentheses. First 10 PCAs included in the models.
~ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3. First stage regression AFB, LifeLines

| | Women | | Men | | Pooled | |
|---|---|---|---|---|---|---|
| | **IV SNPs** | **IV genetic score** | **IV SNPs** | **IV genetic score** | **IV SNPs** | **IV genetic score** |
| Year of birth | 2.190*** | 2.272*** | 0.963~ | 0.913~ | 1.463*** | 1.500*** |
| | (0.401) | (0.398) | (0.516) | (0.514) | (0.316) | (0.315) |
| Year of birth - squared - | -0.110** | -0.118*** | -0.049 | -0.044 | -0.066* | -0.069* |
| | (0.035) | (0.035) | (0.045) | (0.045) | (0.028) | (0.027) |
| rs9320913_A | 0.131* | | 0.048 | | 0.099~ | |
| | (0.066) | | (0.090) | | (0.054) | |
| rs3783006_C | 0.014 | | 0.074 | | 0.040 | |
| | (0.069) | | (0.092) | | (0.056) | |
| rs8049439_T | 0.142* | | 0.091 | | 0.117* | |
| | (0.068) | | (0.091) | | (0.055) | |
| rs13188378_A | 1.018* | | 0.277 | | 0.633 | |
| | (0.480) | | (1.716) | | (0.916) | |
| Genetic score - standardized- | | 0.312*** | | 0.313*** | | 0.311*** |
| | | (0.047) | | (0.063) | | (0.038) |
| Sex (female) | | | | | -0.843*** | -0.851*** |
| | | | | | (0.080) | (0.080) |
| Constant | 1.839 | 3.944*** | 9.270* | 10.172*** | 7.027*** | 8.466*** |
| | (1.459) | (1.132) | (3.751) | (1.433) | (2.071) | (0.900) |
| Observations | 6580 | 6580 | 4378 | 4378 | 10958 | 10958 |
| $R^2$ | 0.077 | 0.082 | 0.027 | 0.033 | 0.059 | 0.064 |
| F | 37.736 | 48.884 | 8.977 | 12.849 | 43.973 | 57.648 |

Standard errors in parentheses. First 10 PCAs included in the models.
~ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4. First stage regression Childlessness, LifeLines

| | Women | | Men | | Pooled | |
|---|---|---|---|---|---|---|
| | **IV SNPs** | **IV genetic score** | **IV SNPs** | **IV genetic score** | **IV SNPs** | **IV genetic score** |
| Year of birth | 1.734* | 1.851** | 1.721 | 1.487 | 0.617 | 0.666 |
| | (0.712) | (0.710) | (1.760) | (1.758) | (0.589) | (0.588) |
| Year of birth - squared - | -0.054 | -0.065 | -0.140 | -0.112 | 0.048 | 0.044 |
| | (0.071) | (0.071) | (0.204) | (0.204) | (0.060) | (0.060) |
| rs9320913_A | 0.105 | | -0.112 | | 0.049 | |
| | (0.079) | | (0.165) | | (0.072) | |
| rs3783006_C | 0.093 | | -0.048 | | 0.061 | |
| | (0.083) | | (0.163) | | (0.074) | |
| rs8049439_T | 0.138~ | | 0.126 | | 0.133~ | |
| | (0.082) | | (0.162) | | (0.074) | |
| rs13188378_A | 0.898~ | | -0.123 | | 0.629 | |
| | (0.540) | | (2.698) | | (1.116) | |
| Genetic score - standardized- | | 0.312*** | | 0.238* | | 0.288*** |
| | | (0.056) | | (0.109) | | (0.050) |
| Sex (female) | | | | | -1.469*** | -1.475*** |
| | | | | | (0.142) | (0.142) |
| Constant | 2.851 | 4.685** | 8.649 | 8.864* | 9.394*** | 10.784*** |
| | (2.031) | (1.736) | (6.475) | (3.748) | (2.728) | (1.461) |
| Observations | 4754 | 4754 | 1468 | 1468 | 6222 | 6222 |
| $R^2$ | 0.080 | 0.085 | 0.035 | 0.039 | 0.067 | 0.072 |
| F | 30.554 | 39.668 | 4.735 | 5.949 | 30.905 | 39.690 |

Standard errors in parentheses. First 10 PCAs included in the models.
~ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5. First stage regression, TwinsUK

| | NEB | | AFB | | Childlessness | |
|---|---|---|---|---|---|---|
| | **IV SNPs** | **IV genetic score** | **IV SNPs** | **IV genetic score** | **IV SNPs** | **IV genetic score** |
| Year of birth | -0.551 | -0.301 | -0.430 | -0.210 | -0.417 | -0.222 |
| | (0.746) | (0.647) | (0.487) | (0.421) | (0.305) | (0.264) |
| Year of birth$^2$ | 0.135 | 0.110 | 0.110$^{*}$ | 0.088$^{*}$ | 0.115$^{***}$ | 0.096$^{***}$ |
| | (0.085) | (0.074) | (0.049) | (0.043) | (0.028) | (0.025) |
| rs9320913_A | 0.104 | | 0.081 | | 0.085 | |
| | (0.088) | | (0.088) | | (0.076) | |
| rs3783006_C | -0.062 | | 0.002 | | 0.032 | |
| | (0.084) | | (0.084) | | (0.075) | |
| rs8049439_T | -0.126 | | -0.091 | | 0.003 | |
| | (0.085) | | (0.081) | | (0.075) | |
| rs13188378_A | -0.107 | | -0.234 | | -0.115 | |
| | (0.195) | | (0.193) | | (0.168) | |
| Genetic score -standardized- | | 0.166$^{**}$ | | 0.131$^{*}$ | | 0.194$^{***}$ |
| | | (0.057) | | (0.054) | | (0.049) |
| Constant | 11.973$^{***}$ | 11.052$^{***}$ | 11.941$^{***}$ | 10.918$^{***}$ | 11.515$^{***}$ | 10.936$^{***}$ |
| | (1.677) | (1.383) | (1.285) | (1.003) | (0.896) | (0.681) |
| Observations | 2160 | 2734 | 1998 | 2537 | 3000 | 3811 |
| N_clust | 1518.000 | 1821.000 | 1502.000 | 1824.000 | 2123.000 | 2554.000 |
| r2_a | 0.046 | 0.052 | 0.079 | 0.078 | 0.148 | 0.148 |
| F | 4.970 | 8.137 | 7.431 | 11.825 | 21.865 | 34.877 |

Standard errors in parentheses. First 10 PCAs included in the models. Clustered standard errors.
$^{\sim}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$