The Sociogenomics of Polygenic Scores of Reproductive Behavior and Their Relationship to Other Fertility Traits

MELINDA C. MILLS, NICOLA BARBAN, AND FELIX C. TROPF

Human reproductive behavior until relatively recently has been explained exclusively via individual and social characteristics. This article applies results from a recent Genome-Wide Association Study that combined sixty-two data sources to isolate twelve genetic loci associated with reproductive behavior. We create polygenic scores that allow us to include a summary variable of genetic factors into our statistical models. We use four datasets: the U.S. Health and Retirement Study, Dutch LifeLines, TwinsUK and the Swedish Twin register. First, we provide a brief overview of the dominant explanations of reproductive behavior. Second, we test the predictive power of polygenic scores. Third, we interrogate the robustness of our models using a series of sensitivity analyses to take into account possible confounders due to population stratification and selection.

Keywords: human reproduction, polygenic scores, genetics, educational attainment, age at first birth, number of children ever born, fertility

Human reproductive behavior—measured by age at first birth (AFB) and number of children ever born (NEB)—is a central topic of study within the social, medical, and biological sciences. AFB and NEB are complex behaviors not only related to biological fecundity, but also have a strong behavioral element in that they are driven by the reproductive choice of individuals and their partners. They are likewise influenced by the environment and social in-

Melinda C. Mills is Nuffield Professor of Sociology at the University of Oxford and Nuffield College and leads the Sociogenome project. **Nicola Barban** is senior research associate in the Department of Sociology at the University of Oxford and Nuffield College. **Felix C. Tropf** is a postdoctoral researcher in the Department of Sociology at the University of Oxford and Nuffield College.

© 2018 Russell Sage Foundation. Mills, Melinda C., Nicola Barban, and Felix C. Tropf. 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): 122–36. DOI: 10.7758/RSF.2018.4.4.07. The research leading to these results has received funding from the following awards to PI M.C. Mills, European Research Council (ERC) Consolidator Grant SOCIOGENOME (615603, www.sociogenome.com), and Economic and Social Research Council (ESRC) UK, National Centre for Research Methods (NCRM) grant SOCGEN, European Union's FP7 FamiliesAndSocieties project (no.320116) and the Wellcome Trust ISSF and John Fell Fund. Direct correspondence to: Melinda C. Mills at melinda.mills@nuffield.ox.ac.uk, Department of Sociology, Nuffield College, University of Oxford, OX1 3UQ United Kingdom; Nicola Barban at nicola.barban@sociology.ox.ac.uk; and Felix C. Tropf at felix.tropf@sociology.ox.ac.uk.

Open Access Policy: *RSF: The Russell Sage Foundation Journal of the Social Sciences* is an open access journal. This article is published under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.

stitutions, including multiple factors such as contraceptive legislation and availability, educational expansion, and social norms (Balbo, Billari, and Mills 2013). The past four decades have brought a rapid postponement of AFB by around four to five years in many advanced societies and a growth in childlessness (Mills et al. 2011). The biological ability to conceive starts to steeply decline for some women as early as age twenty-five, and almost 50 percent of women are sterile by the age of forty (Leridon 2008). This delay has been linked to an unprecedented growth in infertility (involuntary childlessness), which now affects around 10 to 15 percent of couples in Western societies, and forty-eight million couples worldwide are estimated as infertile (Boivin et al. 2007).

Relatively little is known about the specific genetic architecture of human reproductive behavior of AFB and NEB and the genetic relationship to other fertility traits that mark the reproductive window such as menarche and menopause or behaviorally relevant traits such as educational attainment (Okbay, Beauchamp, et al. 2016). The current study uses polygenic scores constructed from a recent large meta-GWAS (Genome-Wide Association Study) of AFB and NEB, which used data from sixty-two sources to isolate twelve loci linked to these traits (Barban et al. 2016). Some of the results reported here are briefly reported in the supplementary material of this study, but without detailed discussion, clarification or reflection.

CENTRAL EXPLANATIONS OF REPRODUCTIVE BEHAVIOR

Reproductive behavior has been largely explained by social scientists by focusing on individual and couple characteristics and social structural or institutional factors (Balbo, Billari, and Mills 2013). Core explanations, bolstered by a large body of empirical evidence, has related the timing and number of children to educational systems and the educational level of individuals (particularly women) (Bhrolcháin and Beaujouan 2012; Rindfuss, Morgan, and Offutt 1996; Tropf and Mandemakers 2017), gender equity (McDonald 2002; Mills et al. 2011), normative changes in preferences for children (de Kaa 1987), effective contraception (Murphy 1993), availability of childcare (Brewster and Rindfuss 2000), women's employment and occupation (Begall and Mills 2013; Brewster and Rindfuss 2000), social interactions (Balbo and Barban 2014) and economic uncertainty (Mills, Blossfeld, and Klijzing 2005).

The genetic basis of human reproduction has often been ignored or even actively resisted by social scientists. As a recent review of the biodemographic approach to fertility highlighted, the avoidance is largely attributed to the dark history related to eugenic policies, lack of proper interdisciplinary training, and appropriate genetic data that also contains social science behavioral measures (Mills and Tropf 2016). As noted by pioneers in this field (Kohler, Rodgers, and Christensen 1999; Rodgers et al. 2001), another reason this connection has been avoided is often attributed to an erroneously interpreted version of Ronald Fisher's (1930) Fundamental Theorem of Natural Selection. Some interpreted Fisher's theory to mean that since fertility is a fitness trait, this should theoretically entail that a genetic basis (referred to as *heritability*[1]), should be zero. Fisher actually argued that fitness is moderately heritable in human populations. A naïve interpretation has been that genes that reduce fitness should have been less frequently passed on, leading to the elimination of genetic variability in traits such as fertility (Courtiol, Tropf, and Mills 2016). Nevertheless, we find that fitness traits such as NEB and AFB have what is known as significant narrow-sense heritabilities.[2] It may be that new mutations restore any genetic variance lost to selection, that there are sexual an-

1. Heritability (H^2) is a statistical term used to denote the proportion of phenotypic (trait) variance due to variance in genotypes. It is important to note that it is specific to the population and environment of analysis and that it is a population and not an individual estimate. It is not a simple measure of the degree to which a trait or phenotype is genetic but rather the proportion of phenotypic variance that is the result of genetic factors.

2. Both broad- and narrow-sense heritability can be estimated. Broad-sense heritability is the ratio of the total genetic variance to total phenotypic (trait) variance or: $H^2 = V_G/V_P$. Narrow-sense heritability refers to ratio of the additive genetic component in contrast to the total (nonadditive) phenotypic variance or: $h^2 = V_A/V_P$.

tagonistic genetic effects (genes have opposite effects for the fertility of men and women), nonadditive genetic effects, environment and gene-environment interaction (Tropf et al. 2017; Verweij et al. 2017).

At least some genetic underpinnings of fertility behavior are plausible. In fact, a growing number of twin and family studies have shown a genetic component underlying AFB and NEB (Briley, Tropf, and Mills 2017; Tropf, Barban, et al. 2015; Zietsch et al. 2014). A recent meta-analysis of all twin studies conducted until 2012 shows average heritability of 0.45 (SE = 0.027, N = 50,265) among sixty-four reproductive disease traits of women and of 0.36 (SE= 0.054, N = 9,376) among twenty-five reproductive disease traits of men (Polderman et al. 2015). The advent of molecular genetic data and complementary analytical tools means that we are now able to go beyond twin models to examine for the first time the genetic relatedness of unrelated individuals (Mills and Tropf 2016; Yang et al. 2010). A recent study using whole-genome data of unrelated individuals shows that 10 percent of the variance in NEB and 15 percent in AFB are associated with common additive genetic variance (Tropf, Stulp, et al. 2015). These previous studies, however, were merely able to state that genetics contributed to fertility behavior only a certain proportion or amount. Until recently, we did not isolate any specific genes related to this behavior or whether they had a biological function. Our recent study isolated twelve genetic loci associated with AFB and NEB (Barban et al. 2016), which allows us for the first time to include a genetic variable or predictor of this behavior in our social science research. Given that human reproduction is a complex behavioral outcome, it is not simply one candidate gene that can be used to predict outcomes. Rather, the myriad of genetic loci are compiled into a comprehensive polygenic score (PGS). It is the relevance of these scores for AFB and NEB for research in the area of fertility and reproduction in the social sciences and beyond which that we explore in this article.

DATA

To examine these questions and avoid false positives from examining the associations in one limited dataset, we test our results using four datasets: the Health and Retirement Study (HRS), LifeLines, TwinsUK, and Swedish Twin Registry (STR).

HRS

The Health and Retirement Study is an ongoing cohort study of Americans, with interview data collected biennially on demographics, health behavior, health status, employment, income and wealth, and insurance status. The first cohort was interviewed in 1992 and subsequently every two years; five additional cohorts were added between 1994 and 2010. Between 2006 and 2008, the HRS genotyped 12,507 respondents who provided DNA samples and signed consent. DNA samples were genotyped using the Illumina Human Omni-2.5 Quad BeadChip, with coverage of approximately 2.5 million single nucleotide polymorphisms (SNPs). The full details of the study are described in (Juster and Suzman 1995).

LifeLines Cohort Study

The LifeLines Cohort Study is a multidisciplinary prospective population-based cohort study from the Netherlands, examining in a unique three-generation design the health and health-related behaviors of 167,729 persons living in the north of the Netherlands, including genotype information from more than thirteen thousand unrelated individuals (Klijs, Scholtens, and Mandemakers 2015). It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical, and psychological factors that contribute to the health and disease of the general population; its special focus is on multimorbidity and complex genetics.

TwinsUK

For the UK, we use data from TwinsUK, the largest adult twin registry in the country with more than twelve thousand respondents (Moayyeri et al. 2013). The TwinsUK Study recruited white monozygotic (MZ) and dizygotic (DZ) twin pairs from the TwinsUK adult twin registry, a group designed to study the heritability and genetics of age-related diseases. These twins were recruited from the general population through national media campaigns in the United King-

dom and shown to be comparable to age-matched population singletons in terms of clinical phenotype and lifestyle characteristics.

STR

The Swedish Twin Registry was first established in the late 1950s to study the importance of smoking and alcohol consumption on cancer and cardiovascular diseases while controlling for genetic propensity to disease. Between 1998 and 2002, the STR conducted telephone interview screening of all twins born in 1958 or earlier regardless of gender composition or vital status of the pair. This effort is known as Screening Across the Lifespan Twin study (SALT). A subsample of SALT (≈10,000) was genotyped as part of the TwinGene project (Lichtenstein et al. 2006) and we use this information in the current study.

CONSTRUCTING LINEAR POLYGENIC SCORES OF REPRODUCTION

Because we have direct access to genotypic data, we first performed out-of-sample prediction using cohorts for the four data sources. GWAS results are generally performed by a meta-analysis of the results from multiple datasets, which in our case was the combined results from sixty-two sources. Out-of-sample prediction refers to the fact that when we construct the PGS to use with a particular dataset, we first need to remove the contribution of the results from that dataset to avoid overfitting the model. In other words, the descriptive results from all of the four datasets were in the original meta-analysis that included sixty-two datasets, which we used to discover the genetic loci associated with AFB and NEB (Barban et al. 2016). To properly construct the PGS for use in the HRS, for example, we need to remove HRS from the results and re-run the meta-analysis without the HRS results to produce a new bespoke or tailored PGS, which can used for that dataset. We therefore calculated polygenic scores for AFB and NEB, based on GWA meta-analysis results and used regression models to predict the same phenotypes in each independent data source.

A polygenic score is a linear combination of the effects of genetic variants present in the entire genome and can be interpreted as a single quantitative measure of genetic predisposition. Just as a battery of multiple questions on personality types or attitudes toward immigration can make up a scale that is measured by one index, a PGS assumes that individuals fall somewhere on a continuum of genetic predisposition resulting from small contributions from many genetic variants. This is particularly relevant when single genetic variants have too small of an effect in explaining complex phenotypes, a common case for complex behavioral traits such as educational attainment (Okbay, Beauchamp, et al. 2016), well-being, neuroticism, depression (Okbay, Baselmans, et al. 2016), or fertility. A PGS for individual i can be calculated as the sum of the allele counts a_{ij} (0,1 *or* 2) from each SNP $j = 1,..M$, multiplied by a weight w_j:

$$PGS_i = \Sigma_{j=1}^{m} w_j a_{ij},$$

using as a choice of weights the association coefficients derived from our recent GWAS on fertility traits. To be clear, it is not the summary of the top genetic loci that were previously isolated, but a sum of the allele counts from all SNPs.

A pivotal question for social scientists is how relevant these PGSs are for applied research and for inclusion in our statistical models. In other words, do the genetic scores that we produce actually predict those observed outcomes? If so, what percentage of the variance do they explain? To determine this, we ran ordinary least-squares (OLS) regression models and report the R-squared as a measure of goodness-of-fit of the model. In addition, we tested how well our polygenic scores for NEB could predict childlessness at the end of the reproductive period (using age forty-five for women and fifty-five for men) and estimated a Cox model examining the impact of the PGS of AFB on observed AFB.

We then reran meta-analyses of the pooled AFB and NEB phenotypes, excluding each of the four independent cohorts. Using these summary statistics, we constructed linear polygenic scores using the effect sizes from the original meta-analysis. We constructed all scores using the software PLINK (Purcell et al. 2007) and PRSice (Euesden, Lewis, and O'Reilly 2014)

based on best call genotypes imputed to HapMap reference 3 panel. For each phenotype, we calculated nine scores using different *p*-value thresholds: 5e-08, 5e-07, 5e-06, 5e-05, 5e-04, 5e-03, 0.05, 0.5 and 1. Results are clumped using the genotypic data as a reference panel for linkage disequilibrium structure.

To control for cohort effects, we first regressed each phenotype on birth year, its square and cubic to control for nonlinear trends in fertility, and the first ten principal components. If the cohort included both men and women, we included sex as a covariate in the regression models. Next, we regressed the residuals from the previous regression on the polygenic score.

AFB and NEB

We now examine whether the polygenic scores predict AFB and NEB.

OLS and Goodness-of-Fit

To test the variance explained or statistical power of our PGS on predicting the actual observed AFB and NEB, we adopt two models. First, we performed a set of OLS regressions where we calculated the R-squared as an indicator of goodness-of-fit of the regression model. For the twin studies (STR and Twins-UK), we included only one MZ twin in the analysis and used clustered standard errors at the family level. Because MZ twins share the same genetics, their PGS is identical. At the same time, DZ twins share on average 50 percent of their genetic variants, leading to correlated PGSs in the sample. Removing a random MZ twin for each family from the sample and using robust clustered errors in the analysis allow us to control for correlated observations in the analysis. To obtain 95 percent confidence intervals around the incremental R-squareds, bootstrapping was performed with one thousand repetitions.

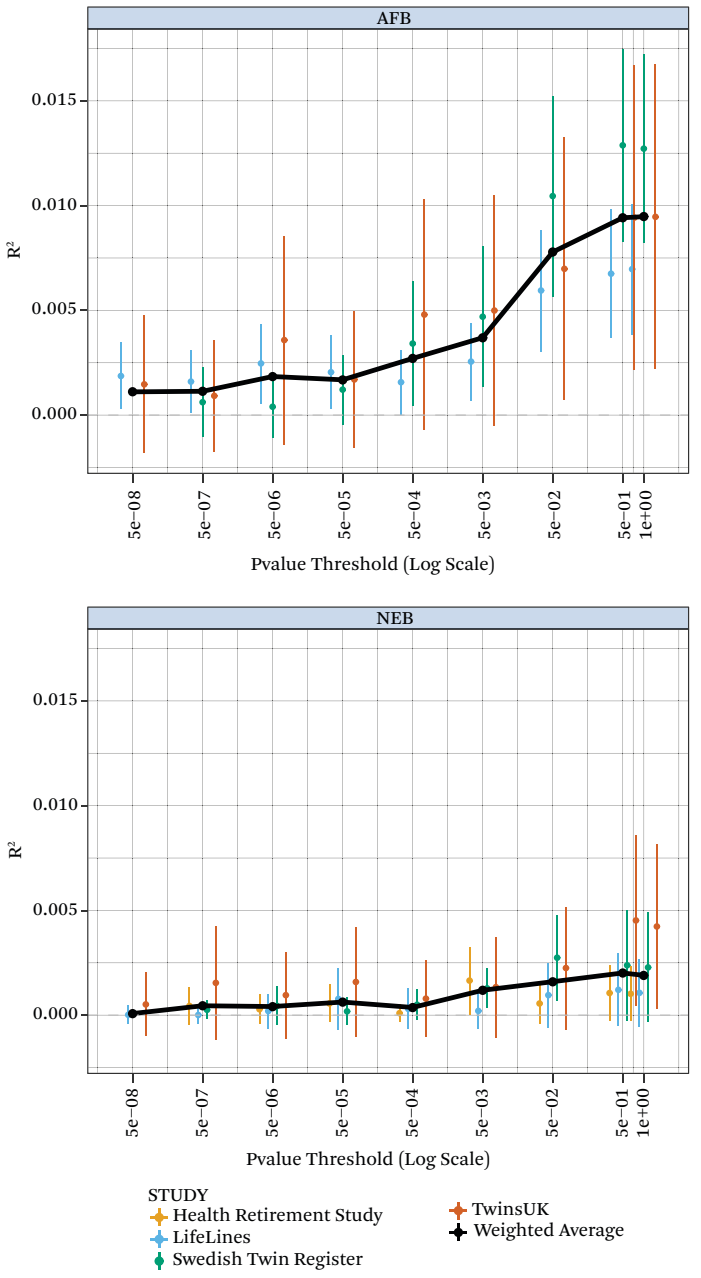
The results of the polygenic score analyses are depicted in figure 1. The sample-size-weighted mean predictive power of the AFB score constructed with all SNPs is 0.9 percent, and the NEB is 0.2 percent. On average, one standard deviation (SD) variation on the polygenic score for AFB is associated with 0.48 years (175 days) AFB for women and 0.33 years (120 days) for men. In other words, those who score higher on the genetic continuum are more prone to having their first child later and have an observed delay in first birth of almost six months for women and four months for men. The variation of one SD in the polygenic score for NEB is associated with an increase of 0.04 children on average. Although it is hard to think in terms of a "fraction" of a child for individuals, our results do indicate that those genetically prone to have more children indeed have more.

Cox Model of Age at First Birth

The previous OLS regression results for AFB include only those in the analysis who have a reported AFB. Logically, AFB is assessed only for men and women who ever became parents and does not take into account that a proportion of respondents are still at risk of having a child (that is, did not have a child yet by the date of the interview) or will remain childless. This problem is commonly referred to in the statistical literature as *right censoring* because the outcome is not observed for all respondents, even though some respondents may still experience the birth of their first child (Mills 2011).

As touched on previously, to model age at first birth more appropriately, it is important to account for right-censored data. The previous OLS models included information for only those who had actually experienced a first birth. Many individuals, however, either did not have a first child by the time of the interview due to their age or are childless. These are referred to as right-censored cases. In an event history framework such as a Cox model, it is possible to include these cases by including the information about these individuals up to the date of the last observation (Mills 2011). For this reason, we estimated a second model in the form of a semi-parametric Cox regression model (Cox 1972) for the effect of the polygenic score on increasing the hazard of having a child conditional on age. This class of models takes into account censoring and is widely used to study fertility timing. Our results show that the calculated PGS for AFB based on all SNPs is associated with an increased risk of childbearing at any age (see tables 1 and 2). The median AFB

Figure 1. Variance Explained by AFB and NEB Polygenic Scores



Source: Authors' calculations from the Health and Retirement Survey, LifeLines Cohort Study, Twins-UK, and STR.

Note: Calculated with the inclusion of SNPs at different levels of significance. Polygenic scores were calculated from the meta-analysis results excluding the validation cohort. The y-axis is the variance explained (R-squared from OLS regression with polygenic score as sole predictor). The x-axis represents the p-value inclusion threshold used in the construction of the polygenic score. The black line is the sample-size-weighted mean R-squared. Cohort-specific estimates and 95 percent confidence intervals obtained with one thousand bootstrap samples. Results are adjusted for birth cohort, first ten principal components, and sex. Clustered standard errors have been used for family-based studies.

Table 1. Logit Regression of Childlessness on NEB Polygenic Score

	Pooled Sample		LifeLines		STR		TwinsUK
	Women	Men	Women	Men	Women	Men	Women
Childless							
PGS NEB (all SNPs)	0.911***	0.979	0.904*	0.988	0.913	0.945	0.909*
SE	(0.022)	(0.037)	(0.038)	(0.043)	(0.043)	(0.081)	(0.037)
N	17,465	10,126	5,411	4,536	7,795	5,590	4,259

Source: Authors' calculations from LifeLines Cohort Study, TwinsUK, and STR.
Note: Age forty-five for women, fifty-five for men, using all SNPs on score. PGS = polygenic score; NEB = number of children ever born, SNPs = single nucleotide polymorphisms, exponentiated coefficients. Standard errors in parentheses.
*$p < .05$; **$p < .01$; ***$p < .001$

Table 2. Within Families Regressions

Regression coefficients	AFB OLS	AFB WF	NEB OLS	NEB WF
Beta (all SNPs)	0.508***	0.317**	0.0623***	0.0601*
	(0.0450)	(0.115)	(0.00989)	(0.0256)
Constant	−0.00202	−0.00367	−0.00402	−0.00401
	(0.0455)	(0.0371)	(0.0104)	(0.00907)
N	11,613	11,613	14,206	14,206
R^2	0.0118	0.00207	0.00280	0.00108
R^2 within		0.00207		0.00108
R^2 between		0.0131		0.00347
N groups		7,944		9,090

Source: Authors' calculations from TwinsUK and STR.
Note: AFB = Age at first birth; NEB = Number of children ever born, SNPs = single nucleotide polymorphisms; OLS = ordinary least-squares regression; WF= within-family.

for men in the pooled sample is twenty-eight and twenty-six for women. The hazard ratio of the PGS for AFB is 0.92 for women and 0.97 for men. This means that an increase of one standard deviation in the PGS is associated with a decrease of 8 percent in the probability of having a child at any age for women and 3 percent for men. Results for different cohorts and sex are presented in table 1.

Childlessness

We used the score for NEB in an additional analysis to predict the probability of remaining childless at the end of the reproductive period. Despite its limited predictive power in the previous OLS model of NEB, our analysis shows that an increase of one SD of the polygenic score is associated with a decrease of around 9 percent in the probability to remain childless for women, and that no significant differences among men are discernable (see table 3). The results are consistent across cohorts.

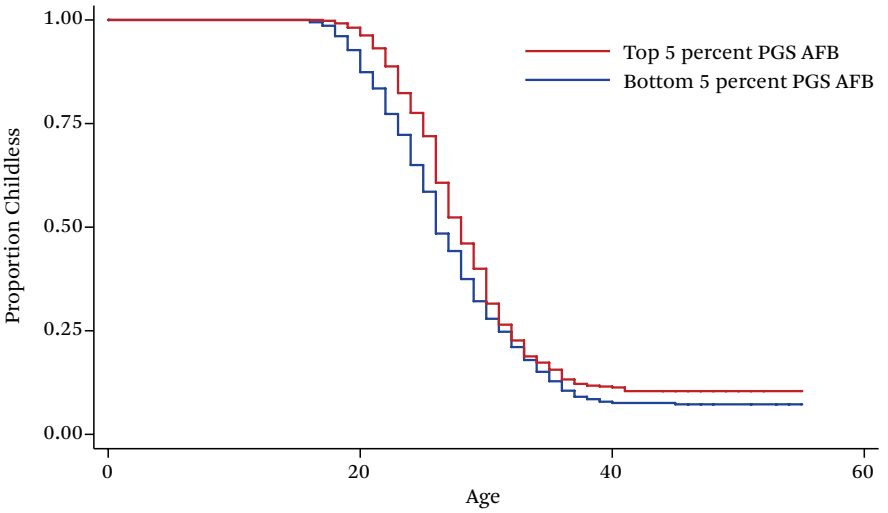
To illustrate differences in childlessness by genetic predisposition, we estimated the proportion of individuals without children by polygenic score, comparing individuals with extreme polygenic scores. Figure 2 shows that men and women with a PGS lower than the 5th percentile are more likely to ever have had a child at any age compared to individuals in the 95th percentile. This underscores the relevance of our genetic measures for fertility research.

Table 3. Cox Regression Model, Age at First Birth on AFB Polygenic Score (all SNPs)

	Pooled Sample		LifeLines		STR		TwinsUK
	Women	Men	Women	Men	Women	Men	Women
PGS AFB (all SNPs)	0.917***	0.968**	0.944***	0.978	0.871***	0.960*	0.938***
	(0.008)	(0.010)	(0.011)	(0.014)	(0.013)	(0.015)	(0.018)
N	16,132	9,136	7,154	4,611	5,409	4,525	3,569

Source: Authors' calculations from LifeLines Cohort Study, TwinsUK, and STR.
Note: PGS = Polygenic Score; AFB = age at first birth, SNPs = single nucleotide polymorphisms. Exponentiated coefficients. Standard errors in parentheses.
*$p < .05$, **$p < .01$, ***$p < .001$

Figure 2. Kaplan-Meier Estimation of Childlessness by Age and by Polygenic Score



Source: Authors' calculations from LifeLines Cohort Study, TwinsUK, and STR.

OTHER FERTILITY-RELATED TRAITS

We also wanted to know to what extent the linear PGS for AFB and NEB can predict related fertility traits, namely age at menarche and completed by age at menopause.

We used TwinsUK to model age at menarche. Age at menarche (AAM) has been assessed for 6,838 women using the following question: "How old were you when you had your first menstrual period?" The average age at AAM in the sample is thirteen (SD = 1.59 years). To examine menopause, we used the age at menopause measurement included in the Dutch LifeLines cohort. Age at menopause is measured with the question: "At what age did you have your last menstrual period?" We excluded women from the sample who reported having had their last menstruation before age thirty or after age sixty. The median age at natural menopause (ANM) in the sample is forty-five.

Our results in table 4 indicate that those with a genetic propensity to have a later AFB also show a shift of the entire reproductive window, to both later onset of menarche and menopause. Table 4 shows that an increase of one standard deviation on the PGS of AFB is associated with an increase of 0.06 years, or just under one month (twenty-two days), on age at menarche. The PGS for AFB is likewise associated with a later ANM. Because a substantive proportion of the sample of women in LifeLines is still in the pre-menopausal period, we

Table 4. Linear Prediction of Age at Menarche and Menopause Using AFB PGS Linear Score

	Age at Menarche OLSTwinsUK	Age at Menopause Cox Survival Model LifeLines
PGS AFB	0.0560*	0.971*
	(0.028)	(0.012)
_cons	13.00***	
	(0.029)	
N	3,424	6,923
R^2	0.00694	

Source: Authors' calculations from LifeLines Cohort Study and TwinsUK.
Note: PGS = polygenic score. Standard errors in parentheses.
*p < .05; **p < .01; ***p < .001

estimated a proportional hazard model (Cox regression) in which we estimate ANM as a function of PGS for AFB. Our estimates indicate that having higher predisposition for AFB is associated with a later ANM. The hazard ratio estimate 0.97 indicates that an increase of one standard deviation of the PGS for AFB is associated with a decrease of the occurrence of menopause at any age of about 3 percent.

SENSITIVITY TESTS

To test the robustness of our all-SNP polygenic scores, we estimated within family (WF) regressions of AFB and NEB on polygenic scores. These regressions control for possible bias due to population stratification. Population stratification refers to a systematic relationship between the allele frequency and the outcome of

Table 5. OLS Regressions and Heckman Two-Stage Regression Models of AFB on Polygenic Score (Using All SNPs)

	Pooled Sample				LifeLines			
	OLS Women	Heckman Women	OLS Men	Heckman Men	OLS Women	Heckman Women	OLS Men	Heckman Men
Main								
PGS AFB	0.483***	0.450***	0.331***	0.321***	0.392***	0.388***	0.267***	0.273***
	(0.037)	(0.036)	(0.050)	(0.049)	(0.050)	(0.049)	(0.065)	(0.065)
_cons	0.0058	−1.056***	−0.00942	−0.767***	0.00625	−0.506***	−0.00697	−0.268***
	(0.037)	(0.042)	(0.050)	(0.053)	(0.048)	(0.052)	(0.064)	(0.065)
Ever had children								
PGS NEB		0.0368***		0.0355*		0.0315		0.0627
		(0.010)		(0.016)		(0.020)		(0.033)
_cons		1.126***		1.323***		1.454***		1.833***
		(0.013)		(0.019)		(0.022)		(0.036)
Athrho								
_cons		1.481***		1.412***		1.163***		1.211***
		(0.039)		(0.050)		(0.059)		(0.083)
Insigma								
_cons		1.605***		1.611***		1.458***		1.500***
		−0.00839		−0.0113		(0.012)		(0.014)
N	14,389	16,405	8,341	9,136	6,646		4,471	4,611
r^2	0.012		0.006		0.010		0.004	
rho		0.902		0.888		0.822		0.837

Source: Authors' calculations from LifeLines Cohort Study, TwinsUK, and STR.
Note: First stage selection models based on NEB polygenic scores (all SNPs). AFB = age at first birth; NEB = number of children ever born, SNPs = single nucleotide polymorphisms; OLS = ordinary least-squares regression; WF = within-family. Standard errors in parentheses.
*p < .05; **p < .01; ***p < .001

interest in different subgroups of the population. Genetic similarity is often correlated with geographical proximity because human genetic diversity is the result of the history of population migration, ethnic admixture, and residential segregation. In such studies, it is essential to clarify whether results are a true signal of similarities or merely attributable to the presence of two or more population subgroups having different genetic or allele frequencies that are a result of a coincidence of being correlated with different levels of a particular trait. A common example is the chopsticks gene finding. In this fictional scenario, a geneticist wanted to discover why some people eat with chopsticks and others do not and found a considerable correlation to account for about half of the variance. The finding, however, was attributed to the different genetic or allele frequencies in Asians and Caucasians, who use cutlery and chopsticks to different extents—that is, cultural rather than genetic reasons.

By examining differences in the polygenic scores between DZ twins, WF regressions cancel out possible confounders due to the population structure of the sample. Because DZ twins have the same family environment, results from a family fixed-effect regression model are net of any family-specific confounder, including ancestry. As we did in the standard model, we standardized NEB and AFB on birth year, birth year squared, birth year cubic and sex. Our regressions are based on 7,944 twin couples for AFB and 9,220 twin couples for NEB. Table 5 reports the results of standard OLS and WF statistical models.

STR				TwinsUK	
OLS Women	Heckman Women	OLS Men	Heckman Men	OLS Women	Heckman Women
0.632***	0.458***	0.404***	0.342***	0.456***	0.467***
(0.069)	(0.064)	(0.077)	(0.073)	(0.091)	(0.086)
0.0269	−1.179***	−0.0126	−1.359***	−0.0217	−2.100***
(0.070)	(0.078)	(0.079)	(0.087)	(0.090)	(0.120)
	0.0459**		0.0231		0.0293
	(0.015)		(0.018)		(0.019)
	1.111***		1.027***		0.733***
	(0.021)		(0.023)		(0.025)
	1.941***		1.617***		1.519***
	(0.110)		(0.070)		(0.072)
	1.679***		1.719***		1.749***
	(0.014)		(0.017)		(0.019)
4,726	5,409	3,870	4,525	3,017	3,842
0.019		0.007		0.009	
	0.96		0.924		0.909

The regression analyses show that within family regression coefficients for both AFB and NEB are statistically different from zero when PGS are based on all SNPs. Both coefficients for AFB and NEB are larger than zero in within family analyses, confirming that the PGS uncover true polygenic signals. Overall, these results indicate a limited effect of population stratification and the existence of true polygenic signals.

A second potential problem is statistical selection; that is, individuals with a measurement of AFB may be genetically distinct from those who remain childless. If childless individuals are different from the general population, the association results on AFB may be biased by selection problems. To understand whether and how these issues would influence our results, we estimated bivariate Heckman selection models in which we estimate the probability of eligibility for AFB in a two-step procedure (Heckman 1974). Because we are interested in possible genetic differences among men and women who have had children rather than childless individuals, we used the PGS for NEB to model the probability of being at risk or eligible for AFB. The results from the Heckman selection models indicate slightly lower coefficients than OLS regression models but no substantial differences (for details, see table 5). We can therefore conclude that statistical selection due to genetic distinctiveness between those who have had a child (for which we have a measure for AFB) and those who have not does not influence our results.

DISCUSSION AND CONCLUSION

The aim of this article is to demonstrate the power of polygenic scores of age at first birth and number of children in predicting the actual observed outcomes and related fertility traits and to ensure that these results were robust. Using an OLS regression model to estimate the overall variance explained or R-squared goodness-of-fit, we show that the predictive power of the AFB PGS was around 1 percent and of the NEB PGS was 0.2 percent. We also see that one SD increase in the AFB PGS is associated with an 8 percent reduction of the hazard ratio of having the first child for women and with a 3 percent reduction for men. The NEB PGS can also be used to study childlessness, a one SD increase in the PGS decreasing the probability of remaining childless by 9 percent in women. It is essential to distinguish clearly between, on the one hand, the predictive power or R-squared that looks at the proportion of variance explained with the OLS models and, on the other, our coefficients from the Cox regression models. We need to think of the interpretation of PGS as changes in one standard deviation of the PGS and how they are related to an increase or reduction in the hazard ratio of reproduction. It is likewise incorrect to state that a one SD of the PGS for AFB is associated with an 8 percent increase of AFB. Rather, our results are presented as relative risk ratios. One SD of the PGS for AFB is associated with an increase of 0.5 years in AFB (and 0.3 years with a fixed-effect model, table 2). We acknowledge that it remains awkward and not immediately intuitive to interpret PGS in terms of SD changes and survival models in terms of hazard and relative risk ratios. For the time being this remains the prominent manner to interpret these findings.

Our results also demonstrate a fascinating underlying genetic link with the shifting of the entire reproductive window for certain individuals. The AFB PGS is clearly linked to development and the reproductive window, those having a genetic propensity to later AFB also having a later genetic propensity for the onset of menarche and ANM. Detailed LD-Score regression analyses have indeed shown a strong genetic association between human development and AFB, including age at voice-breaking for boys and age at first sex (Bulik-Sullivan et al. 2015; Barban et al. 2016). A recent study also found that our AFB PGS is linked to longevity (Mostafavi et al. 2017).

Several conclusions are indicated. First, the predictive power of our polygenic scores when entered alone in the model remains considerably lower than previous research would indicate. Recall that the R-squared goodness-of-fit tests show a predictive power of the linear AFB PGS of around 1 percent and 0.2 percent for NEB. This is a fraction of what previous twin and family studies have found, which predicted these outcomes to be between 25 percent and 45 percent heritable (Mills and Tropf

2016). It is also much lower than recent SNP-based GREML whole-genome methods that predicted that 15 percent for AFB and 10 percent for NEB of the variance was attributed to genetic factors (Tropf, Stulp, et al. 2015). In other words, the ceiling of SNP heritability should likely be more in the range of 10 to 15 percent than 1 percent. Missing heritability can be explained several ways, including nonadditive genetic effects, epistatic effects, and inflated estimates from twin studies due to shared environmental factors (missing heritability, Manolio et al. 2009; nonadditive effects, Zhu et al. 2015; epistatic effects, Zuk et al. 2012; inflated estimates, Felson 2014). Empirical studies, however, find no evidence for any of these reasons.

Jian Yang and his colleagues argue that most genetic effects are too small to be reliably detected in GWAS of current sample sizes, which is why they propose the whole-genome restricted maximum likelihood estimation performed by GCTA software (Yang et al. 2010, 2011). Studies applying these whole-genome methods typically yield estimates with predictive power between twin studies and polygenic scores. A recent investigation also demonstrates that including rare genetic variants can strongly increase the predictive power of genes for body mass index (BMI) and height (Yang, Bakshi, Zhu, Hemani, Vinkhuyzen, Lee, et al. 2015). Similarly, the first meta-GWAS on educational attainment produced three significant hits with small effect size and a total predictive power based on all SNPs of 2 percent (Rietveld et al. 2013). Meanwhile, the most recent meta-GWAS, which finds seventy-four significant hits, explains around 3.2 percent of the observed variance (Okbay, Beauchamp, et al. 2016). This refers to the predictive power of SNPs, though not all SNPs, which is the same approach used in our study. The main differences between the studies are the increased sample size in the latter study as well as the inclusion of more genetic variants. Current predictions are that these differences will only continue to increase with the release of larger datasets such as the UKBiobank. We therefore anticipate that in future GWAS studies, as sample sizes grow and including more detailed genetic information becomes possible, these traits will be more in the range of 10 to 15 percent. Our PGS scores as they stand, however, still had a notable predictive power for the timing of first birth and childlessness.

Another explanation is possible. A recent study on fertility suggests that next to rare variants and insufficient sample size, GWAS discoveries might be limited by heterogeneity across cohorts and birth cohorts under study (Tropf et al. 2017). Heterogeneity can arise on the phenotypic level if the phenotypic measurement differs across cohorts and birth cohorts, on the genotypic level if linkage disequilibrium differs across populations under study, and by gene-environment interaction. They find that the predictive power of the whole-genome methods increases as much as fivefold when heterogeneity across cohorts and birth cohorts is taken into account. Investigations on height and BMI find barely evidence for genome-wide heterogeneity across countries and sexes (Yang, Bakshi, Zhu, Hemani, Vinkhuyzen, Nolte, et al. 2015). Fertility is in large part environmentally determined and modified (Mills et al. 2011). It is therefore highly likely that gene-environment interaction across the more than sixty cohorts, as well as across birth cohorts within cohorts, limited genetic discovery in the most recent GWAS and leads to comparably small predictive power of the polygenic scores.

It is also not surprising that genetic factors are not especially strong in predicting reproductive behavior. A large body of social science research has consistently demonstrated that socio-environmental conditions are key factors shaping human reproduction. We know that women's higher educational attainment and presence in the labor market has resulted in postponed entry into parenthood (Balbo, Billari, and Mills 2013). Another obvious point is that the models presented in this article are not multivariate models. When the gold standard social science variables that predict AFB and NEB such as age at entry into a union or marriage and educational attainment are entered alone in a model, they also have low predictive power (from 6 to 15 percent). It is therefore artificial and unusual within the social sciences to enter only one predictor in a model and to not consider confounders or interactions. The purpose of this article, however, is

to introduce and demonstrate the polygenic scores in the hope that others will include and interrogate these further in multivariate models.

Genetics is likewise only one piece of the puzzle and in this study, we examine only one type of genetic variants (SNPs) and consider only one of the many possible biological and genetic ways in which individuals may vary. Other sources of molecular genetic variation remain to be discovered. We plan to examine our work further with denser genotyping platforms. Other GWAS studies for complex traits such as diseases have also consistently identified common variants with small effects, which explain only a small proportion of the trait of interest. This does not affect the biological importance of the findings, however, because many follow-up studies that have isolated particular genetic loci have the potential to substantially improve our understanding of human biology. In the context of human disease, for example, variants identified by GWAS for diabetes and cardiovascular diseases "tag" genes that encode well-known drug targets for the treatment of such diseases. This implies that a further understanding of the genes underlying the associations we identified for reproductive behavior may result in new reproductive strategies such as those for assisted reproductive technology treatment.

For social scientists who study reproductive behavior, we offer and provide an entirely new variable and way of theoretically thinking about and measuring human reproductive behavior. These polygenic scores for AFB and NEB will also be easily usable in publicly available datasets, which will allow researchers to include these predictors in their research. These PGS scores show that a genetic component underlies AFB and NEB and is related to other fertility traits such as childlessness, menarche, and menopause. This may force us to rethink existing behavioral theories that rarely included biology and genetics in their largely choice and preference-based theoretical models, such as the Theory of Planned Behavior, often used in social science fertility research (Ajzen 1991).

REFERENCES

Ajzen, Icek. 1991. "The Theory of Planned Behavior." *Organizational Behavior and Human Decision Processes* 50(2): 179–211.

Balbo, Nicoletta, and Nicola Barban. 2014. "Does Fertility Behavior Spread Among Friends?" *American Sociological Review* 79(3): 412–31.

Balbo, Nicoletta, Francesco C. Billari, and Melinda C. Mills. 2013. "Fertility in Advanced Societies: A Review of Research." *European Journal of Population/Revue Européenne de Démographie* 29(1): 1–38.

Barban, Nicola, Rick Jansen, Ronald de Vlaming, Ahmad Vaez, et al. 2016. "Genome-Wide Analysis Identifies 12 Loci Influencing Human Reproductive Behavior." *Nature Genetics* 48(12): 1462–72.

Begall, Katia, and Melinda Mills. 2013. "The Influence of Educational Field, Occupation, and Occupational Sex Segregation on Fertility in the Netherlands." *European Sociological Review* 29(4): 720–42.

Bhrolcháin, Maire Ni, and Éva Beaujouan. 2012. "Fertility Postponement Is Largely Due to Rising Educational Enrolment." *Population Studies* 66(3): 311–27.

Boivin, Jacky, Laura Bunting, John A. Collins, and Karl G. Nygren. 2007. "International Estimates of Infertility Prevalence and Treatment-Seeking: Potential Need and Demand for Infertility Medical Care." *Human Reproduction* 22(6): 1506–12.

Brewster, Karin L., and Ronald R. Rindfuss. 2000. "Fertility and Women's Employment in Industrialized Nations." *Annual Review of Sociology* 26(1): 271–96.

Briley, Daniel A., Felix C. Tropf, and Melinda C. Mills. 2017. "What Explains the Heritability of Completed Fertility? Evidence from Two Large Twin Studies." *Behavior Genetics* 47(1): 36–51.

Bulik-Sullivan, Brendan K., Po-Ru Loh, Hilary K. Finucane, Stephan Ripke, et al. 2015. "LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies." *Nature Genetics* 47(3): 291–95. DOI: 10.1038/ng.3211.

Courtiol, Alexandre, Felix C. Tropf, and Melinda C. Mills. 2016. "When Genes and Environment Disagree: Making Sense of Trends in Recent Human Evolution." *Proceedings of the National Academy of Sciences* 113(28): 7693–95.

Cox, David R. 1972. "Regression Models and Life-

Tables." *Journal of the Royal Statistical Society. Series B (Methodological)* 34(2): 187–220.

Euesden, Jack, Cathryn M. Lewis, and Paul F. O'Reilly. 2014. "PRSice: Polygenic Risk Score Software." *Bioinformatics* 31(9): btu848–1468. Accessed August 21, 2017. http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btu848.

Felson, Jacob. 2014. "What Can We Learn from Twin Studies? A Comprehensive Evaluation of the Equal Environments Assumption." *Social Science Research* 43 (January): 184–99. Accessed May 11, 2017. http://www.sciencedirect.com/science/article/pii/S0049089X13001397.

Fisher, Ronald A. 1930. *The Genetical Theory of Natural Selection.* Oxford: Oxford University Press.

Heckman, James J. 1974. "Effects of Child-Care Programs on Women's Work Effort." *Journal of Political Economy* 82(2): 136–63.

Juster, F. Thomas, and Richard Suzman. 1995. "An Overview of the Health and Retirement Study." Special Issue, *Journal of Human Resources* 30: S7–56.

de Kaa, Dirk J. Van. 1987. "Europe's Second Demographic Transition." *Population Bulletin* 42(1): 1–59.

Klijs, Bart, Salome Scholtens, and Jornt J. Mandemakers. 2015. "Representativeness of the LifeLines Cohort Study." *PloS one* 10(9): e0137203.

Kohler, Hans-Peter, Joseph L. Rodgers, and Kaare Christensen. 1999. "Is Fertility Behavior in Our Genes? Findings from a Danish Twin Study." *Population and Development Review* 25(2): 253–88.

Leridon, Henri. 2008. "A New Estimate of Permanent Sterility by Age: Sterility Defined as the Inability to Conceive." *Population Studies* 62(1): 15–24.

Lichtenstein, Paul, Patrick F. Sullivan, Sven Cnattingius, Margaret Gatz, et al. 2006. "The Swedish Twin Registry in the Third Millennium: An Update." *Twin Research and Human Genetics* 9(6): 875–82.

Manolio, Teri A., Francis S. Collins, Nancy J. Cox, David B. Goldstein, et al. 2009. "Finding the Missing Heritability of Complex Diseases." *Nature* 461(7265): 747–53.

McDonald, Peter. 2002. "Gender Equity in Theories of Fertility Transition." *Population and Development Review* 26(3): 427–39.

Mills, Melinda C. 2011. *Introducing Survival and Event History Analysis*. Menlo Park, Calif.: Sage Publications.

Mills, Melinda C., Hans-Peter Blossfeld, and Erik Klijzing. 2005. "Becoming an Adult in Uncertain Times." In *Globalization, Uncertainty and Youth in Society*, edited by H. P. Blossfeld, Erik Klijzing, Melinda Mills, and Karin Kurz. London: Routledge.

Mills, Melinda C., Ronald R. Rindfuss, Peter McDonald, and Egbert te Velde. 2011. "Why Do People Postpone Parenthood? Reasons and Social Policy Incentives." *Human Reproduction Update* 17(6): 848–60.

Mills, Melinda C., and Felix C. Tropf. 2016. "The Biodemography of Fertility: A Review and Future Research Frontiers." Special Issue, *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 55(Demography): 397–424.

Moayyeri, Alireza, Christopher J. Hammond, Anna M. Valdes, and Timothy D. Spector. 2013. "Cohort Profile: TwinsUK and Healthy Ageing Twin Study." *International Journal of Epidemiology* 42(1): 76–85.

Mostafavi, Hakhamanesh, Tomaz Berisa, Felix Day, John Perry, Molly Przeworski, and Joseph K. Pickrell. 2017. "Identifying Genetic Variants that Affect Viability in Large Cohorts." *PLoS Biology* 15(9): e2002458.

Murphy, Michael. 1993. "The Contraceptive Pill and Women's Employment as Factors in Fertility Change in Britain 1963–1980: A Challenge to the Conventional View." *Population Studies* 47(2): 221–43.

Okbay, Aysu, Bart M. L. Baselmans, Jan-Emmanuel De Neve, Patrick Turley, et al. 2016. "Genetic Variants Associated with Subjective Well-Being, Depressive Symptoms, and Neuroticism Identified through Genome-Wide Analyses." *Nature Genetics* 48(6): 624–33.

Okbay, Aysu, Jonathan P. Beauchamp, Mark A. Fontana, James J. Lee, et al. 2016. "Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment." *Nature* 533(7604): 539–42.

Polderman, Tinca J. C., Beben Benyamin, Christiaan A. de Leeuw, Patrick F. Sullivan, Arjen van Bochoven, Peter M. Visscher, and Danielle Posthuma. 2015. "Meta-Analysis of the Heritability of Human Traits Based on Fifty Years of Twin Studies." *Nature Genetics* 47(7): 702–09.

Purcell, Shaun, Benjamin M. Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, Paul I. W. de Bakker, Mark Daly, and Pak C. Sham. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81(3): 559–75.

Rietveld, Cornelius A., Sarah E. Medland, Jaime Derringer, Jian Yang, et al. 2013. "GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment." *Science* 340(6139): 1467–71.

Rindfuss, Ronald R., S. Philip Morgan, and Kate Offutt. 1996. "Education and Changing Age Pattern of American Fertility: 1963–1989." *Demography* 33(3): 277–90.

Rodgers, Joseph Lee, Kimberly Hughes, Hans-Peter Kohler, Kaare Christensen, Debby Doughty, David C. Rowe, and Warren B. Miller. 2001. "Genetic Influence Helps Explain Variation in Human Fertility: Evidence from Recent Behavioral and Molecular Genetic Studies." *Current Directions in Psychological Science* 10(5): 184–88.

Tropf, Felix C., Nicola Barban, Melinda C. Mills, Harold Snieder, and Jornt J. Mandemakers. 2015. "Genetic Influence on Age at First Birth of Female Twins Born in the UK, 1919–68." *Population Studies* 69(2): 129–45.

Tropf, Felix C., and Jornt J. Mandemakers. 2017. "Is the Association Between Education and Fertility Postponement Causal? The Role of Family Background Factors." *Demography* 54(1): 71–91. Accessed May 11, 2017. http://link.springer.com/10.1007/s13524-016-0531-5.

Tropf, Felix C., Gert Stulp, Nicola Barban, Peter M. Visscher, Jian Yang, Harold Snieder, and Melinda C. Mills. 2015. "Human Fertility, Molecular Genetics, and Natural Selection in Modern Societies." *PloS One* 10(6): e0126821.

Tropf, Felix C., Renske M. Verweija, Peter J. van der Most, Gert Stulp, et al. 2017. "Hidden Heritability Due to Heterogeneity Across Seven Populations." *Nature Human Behavior* 1(10): 757–65.

Verweij, Renske M., Melinda C. Mills, Felix C. Tropf, René Veenstra, Anastasia Nyman, and Harold Snieder. 2017. "Sexual Dimorphism in the Genetic Influence on Human Childlessness." *European Journal of Human Genetics* 25 (9): 1067–74.

Yang, Jian, Andrew Bakshi, Zhihong Zhu, Gibran Hemani, et al. 2015. "Genome-Wide Genetic Homogeneity Between Sexes and Populations for Human Height and Body Mass Index." *Human Molecular Genetics* 24(25): 7445–49.

Yang, Jian, Beben Benyamin, Brian P. McEvoy, Scott Gordon, et al. 2010. "Common SNPs Explain a Large Proportion of the Heritability for Human Height." *Nature Genetics* 42(7): 565–69.

Yang, Jian, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher. 2011. "GCTA: A Tool for Genome-Wide Complex Trait Analysis." *American Journal of Human Genetics* 88(1): 76–82.

Zhu, Zhihong, Andrew Bakshi, Anna A. E. Vinkhuyzen, Gibran Hemani, et al. 2015. "Dominance Genetic Variation Contributes Little to the Missing Heritability for Human Complex Traits." *American Journal of Human Genetics* 96(3): 377–85.

Zietsch, Brendan P., Ralf Kuja-Halkola, Hasse Walum, and Karin J. Verweij. 2014. "Perfect Genetic Correlation Between Number of Offspring and Grandoffspring in an Industrialized Human Population." *Proceedings of the National Academy of Sciences* 111(3): 1032–36.

Zuk, Or, Eliana Hechter, Shamil R. Sunyaev, and Eric S. Lander. 2012. "The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability." *Proceedings of the National Academy of Sciences* 109(4): 1193–98.