Educational attainment and fertility: Do genetic endowments matter?

Nicola Barban[1], Elisabetta De Cao[2], Marco Francesconi[3], and Selin Köksal[4]

[1]*University of Bologna, Department of Statistical Sciences*

[2]*University of Bologna, Department of Economics*

[3]*University of Essex, Department of Economics*

[4]*University of Essex, ISER*

October 2023

Introduction

A rich body of demographic research has investigated the relationship between educational attainment and fertility behavior. Yet, the existing studies provide mixed findings. On the one hand, individuals (mostly women) with higher level of education are more likely to pursue their careers, postpone parenthood (Mills et al. 2011) and eventually have fewer children (Martín-García & Baizán 2006, Brand & Davis 2011). On the other hand, highly educated individuals can also mate with other highly educated people (Oppenheimer 1994) which, in return, allow them to form families with higher resources with more equal division of domestic work and possibility to outsource housework and childcare (Behrman & Rosenzweig 2002). Following this approach, research has shown that highly-educated couples have overall higher fertility (Mencarini & Tanturri 2006) or the fertility gap between low and high-SES disappears at later ages (Sobotka 2004).

Within this ample literature, the genetic propensity of reproductive behavior remains an overlooked factor. Fertility is a complex process that has been mutually shaped by social and biological mechanisms and their interactions (Almeling 2015, Harris & McDade 2018). Indeed, (Tropf & Mandemakers 2017) shows that genetic, familial and social factors act jointly in influencing the relationship between education and age at first birth. From a theoretical standpoint, leaving genetic mechanisms out of picture limits our understanding the way education influences fertility. From an empirical viewpoint, given

that both educational attainment and fertility behavior has a significant genetic component (Mills & Tropf 2015), omitting genetic factors would bias the estimates and could overstate the effect of environmental factors (Harris & McDade 2018).

This study aims at addressing the aforementioned gaps through investigating gene-environment interplay for fertility behavior. We exploit RoSLA reform which increased compulsory schooling age in UK from 15 to 16 in 1972 as a natural experiment to uncover the causal impact of education on women's and men's fertility. Furthermore, we interact the effect of RoSLA with polygenic indices (PGIs - also termed as polygenic scores or polygenic risk scores) on educational attainment and fertility (age at first birth for women, number of children ever born for men), to examine whether the genetic make-up play a moderating role in the relationship between education and fertility.

We expect that an increase in compulsory schooling age can be particularly effective in reducing teenage fertility and out-of-union fertility through forcing young adults to stay in education for at least one year. Staying in education longer may also alter teenagers educational and career aspirations and may lead to continue higher levels of education, which in return can also shape their adulthood fertility trajectories.

Data and Methods

Sample

This study makes use of UK Biobank (UKB), a survey that interviewed 502,506 out of 9.2 million NHS registered individuals aged between 40 and 69. Along with socio-economic and demographic information, samples of blood, urine and saliva are collected. Each survey participant is genotyped which makes UKB one of the largest publicly available data source for genetics research (Barban et al. 2021). However, UKB is not a nationally-representative sample as the data collection was volunteering-based. This led UKB to be healthier and of higher socioeconomic status than the general UK population (Fry et al. 2017). Volunteering-based participation to data leads to bias in estimations, and the direction of the bias is not known *a priori* as it depends on the variables driving the selection into participation in the first place, and how these variables are linked to the exposure and the outcome variables of interest (van Alten et al. 2023). To correct for this bias, van Alten et al. (2023) constructed an inverse probability weights which reduces 87% of volunteer bias on average. We employ these weights throughout our analyses.

Variables

As reproductive outcomes, we rely on several indicators; age at first sexual intercourse, age at first birth (collected only for women), number of children ever-born and childlessness. We categorize participants as *childless*, if the participant haven't declared any children at the time of the interview. For women, we also examine at the probability of giving a teenage birth by using the probability of giving birth by age 15 to 20 separately as outcome variables.

We identify the exposure to RoSLA reform using the information on month and year of birth. RoSLA increases the minimum school-leaving age in England, Scotland, and Wales is from 15 to 16 for students born on or after September 1, 1957. The students who were born before this date could drop out at age 15. As a result, RoSLA reform generated a discontinuity in the relationship between education and date of birth. Exploiting this exogenous discontinuity, we categorize individuals who were born after September 1, 1957 as the *treatment* group, while the ones who were born before are classified as control group.

The PGIs we used in the main analysis are based on the Polygenic Index Repository established by Becker et al. (2021). The repository contains PGIs for 47 outcomes (phenotypes) from 11 datasets that collects genetic information. From this repository we rely on PGIs for educational attainment (PGI-EA) and number of children ever born (PGI-NEB) for men. [1]

Empirical Strategy

Building on the specification suggested by Barcellos et al. (2018), we estimate the following model :

$$Fertility_i = \alpha + \beta_1 G_i + \beta_2 E_i + \beta_3 (G_i x E_i) + \beta_4 X_i' + \varepsilon_i \tag{1}$$

where $Fertility_i$ corresponds to reproductive outcomes such as age at first sexual intercourse, age at first birth (only for women), probability of giving first birth by age 15 to 20 (only for women) number of children ever born or childlessness, G_i is the genotype (educational attainment or fertility) of the individual i, E_i measures the effect of environment, in our case indicator of staying in school until the age 16, $G_i x E_i$ is the interplay between genetic influences and environment, X_i' is a set of control including age,

[1] For more information on PGI repository, see: https://www.thessgac.org/pgi-repository.

age squared, country of birth and month of birth and ε_i represents the idiosyncratic error term.

However, decision to stay in education until the age 16 may be correlated with other factors (e.g. physical health), which could be also correlated with that person's fertility decision. In order to address the omitted variable bias, we exploit the RoSLA reform that is enacted in 1972 as an instrument for E_i. Then, we estimate this model using a regression discontinuity (RD) design in which the date of birth constitutes the running variable.

We restrict the data to the participants born in England, Scotland, or Wales within 10 years of September 1957 – *who were born between September 1, 1947 and August 31, 1967*.

Results

First stage

In Figure 1 we present the fraction of female (left panel) and male (right panel) students who left the school by the age of 15. Right before the RoSLA reform, around 10 percent of the female students and around 15 percent of the male students were leaving the school by age 15. However, after the reform (i.e. for the students who were born after September 1, 1957), the same ratio declined significantly, until around 3 percent both for men and women. This discontinuity documents a convincing evidence on the effectiveness of RoSLA reform in increasing the school leaving age.

Reproductive outcomes

Table 1 shows the effect of being exposed to RoSLA reform on the age at first sex, number of children ever-born and probability of remamining childless both for women and men. Looking at the left panel, we see that women who are treated by RoSLA, have their first sexual intercourse later, have fewer children and they are more likely to remain childless. Yet, none of the coefficients are statistically meaningful. On the other hand, men who are exposed to RoSLA have their first sexual intercourse earlier, have fewer children and they are less likely to remain childless. However, similar to women's reproductive outcomes, coefficients for men are also not statistically significant.

Table 2 instead focuses only on women and shows the effect of being exposed to RoSLA on timing of fertility. Accordingly, women who were affected by the RoSLA reform give birth, on average, 1.4 years later ($p<0.05$) compared to their counterparts who were not exposed to the reform. Moreover, we find that women who are exposed to RoSLA have

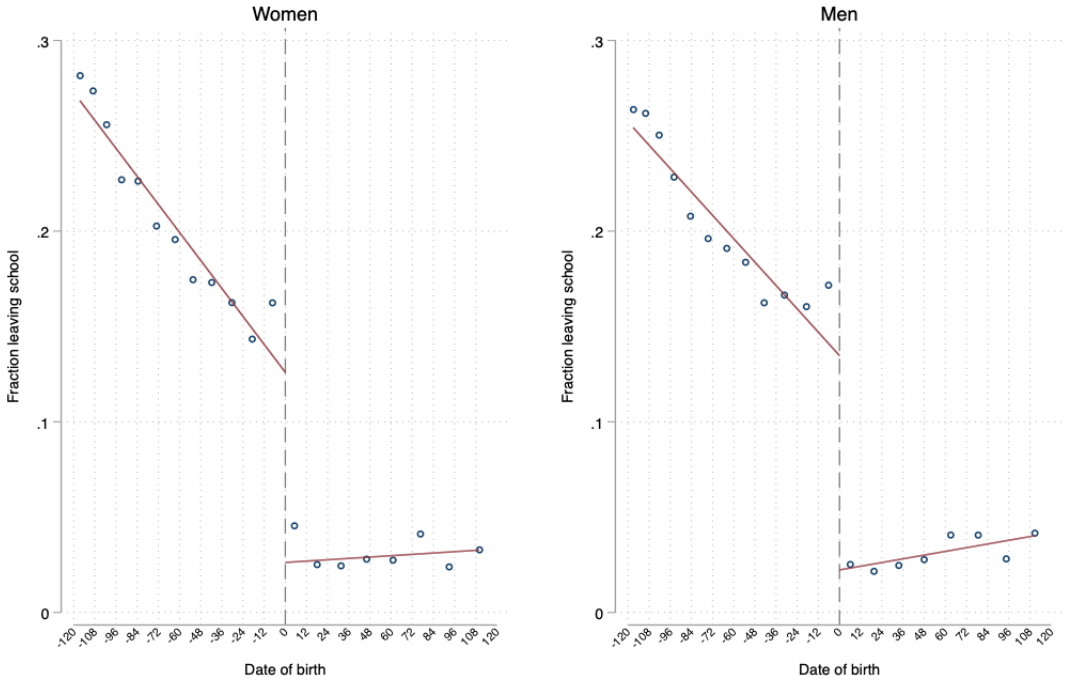

Figure 1: Proportion of women and men left school by the age 15 before and after the RoSLA

Table 1: The effect of RoSLA reform on women and men's reproductive outcomes

	Women			**Men**		
	N of children	Childlessness	Age at FS	N of children	Childlessness	Age at FS
Stayed in school until 16	-0.129	0.032	0.311	-0.093	-0.038	-0.224
	(0.146)	(0.045)	(0.384)	(0.203)	(0.064)	(0.653)
Observations	137,651	137,651	126,074	110,588	110,588	102,161
R-squared	0.012	0.009	0.023	0.007	0.001	0.001
Mean				1.63	0.27	18.45
Fstat	470.7	470.7	420.2	405	405	344.5

Notes: Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

significantly lower probability to give their first birth by age 18 ($p<0.05$). This postponement effect persists also for the age 19 and 20 ($p<0.1$).

Table 2: The effect of RoSLA reform on women's age at first birth

	Age at FB	FB by 15	FB by 16	FB by 17	FB by 18	FB by 19	FB by 20
Stayed in school until 16	1.406*	-0.009	-0.008	-0.068+	-0.109*	-0.102+	-0.100+
	(0.658)	(0.008)	(0.022)	(0.037)	(0.047)	(0.055)	(0.059)
Observations	88,896	88,896	88,896	88,896	88,896	88,896	88,896
R-squared	0.067	0.003	0.006	0.024	0.040	0.047	0.050
Mean		0	0.01	0.05	0.09	0.14	
Fstat	317	317	317	317	317	317	317

Notes: Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The interplay between genes and environment

Table 3: Gene-environment interplay - fertility

	Women			**Men**		
	N of child	Childlessness	Age at FS	N of children	Childlessness	Age at FS
LowPGI * treated	-0.852	0.172	0.254	-0.896	-0.327	-5.862+
	(0.683)	(0.226)	(2.097)	(1.175)	(0.326)	(3.354)
MidPGI * treated	-0.801	0.233	-0.214	-1.546	-0.214	-3.488
	(0.727)	(0.240)	(2.308)	(1.130)	(0.311)	(3.065)
HighPGI * treated	-1.161	0.263	-1.573	-1.321	-0.181	-5.418+
	(1.005)	(0.345)	(3.444)	(1.110)	(0.292)	(2.896)
LowPGI * control	0.084	-0.045	-3.182	-0.627+	0.206+	1.442
	(0.593)	(0.204)	(2.018)	(0.341)	(0.114)	(1.211)
MidPGI * control	-0.152	-0.038	-2.102	0.095	0.053	-1.092
	(0.523)	(0.182)	(1.740)	(0.258)	(0.080)	(0.767)
Observations	43,263	43,263	39,506	35,533	35,533	32,679

Notes: Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

As next step, we estimate whether the effect of RoSLA on reproductive outcomes depends on the genetic propensity for fertility. For women, we used the polygenic indices for

age at first birth, while for men we used the polygenic indices for number of children ever born. According to Table 3 the effect of RoSLA on women's reproductive outcomes, such as age at first sexual intercourse, number of children and the probability of remaining childless does not vary by genetic propensity for fertility. However, as compared to high PGI men who were not treated by RosLA, low PGI and high PGI men who are exposed to fertility have their first sex earlier. Even though the effect is significant at 10 percent, the magnitude is considerably sizeable (5.5 to 6 years). Thus, it requires further investigation.

Table 4 instead focuses solely on women and reports the gene-environment interaction coefficients for the timing of fertility. Accordingly, the effect of RoSLA on the timing of fertility is not moderated by genetic propensity to later or earlier age at first birth.

Table 4: Gene-environment interplay - fertility timing

VARIABLES	Age at FB	FB by 15	FB by 16	FB by 17	FB by 18	FB by 19	FB by 20
LowAFB * treated	4.398	-0.009	0.012	-0.281+	-0.265	-0.153	-0.167
	(2.903)	(0.031)	(0.079)	(0.160)	(0.192)	(0.237)	(0.262)
MidAFB * treated	4.409	-0.028	0.018	-0.261	-0.238	-0.095	-0.094
	(2.956)	(0.031)	(0.085)	(0.165)	(0.203)	(0.247)	(0.270)
HighAFB * treated	6.772	-0.009	0.076	-0.196	-0.071	0.236	0.263
	(4.266)	(0.046)	(0.116)	(0.215)	(0.270)	(0.332)	(0.369)
LowAFB * control	-1.021	0.002	0.077	0.120	0.268	0.511*	0.598**
	(2.674)	(0.037)	(0.070)	(0.129)	(0.166)	(0.201)	(0.225)
MidAFB * control	0.419	0.017	0.063	0.079	0.191	0.383*	0.425*
	(2.395)	(0.032)	(0.059)	(0.109)	(0.141)	(0.174)	(0.198)
Observations	27,836	27,836	27,836	27,836	27,836	27,836	27,836

Notes: Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

References

Almeling, R. (2015), 'Reproduction', *Annual Review of Sociology* **41**(1), 423–442.

Barban, N., De Cao, E. & Francesconi, M. (2021), 'Gene-environment effects on female fertility'.

Barcellos, S. H., Carvalho, L. S. & Turley, P. (2018), 'Education can reduce health differences related to genetic risk of obesity', *Proceedings of the National Academy of Sciences* **115**(42), E9765–E9772.

Becker, J., Burik, C. A., Goldman, G., Wang, N., Jayashankar, H., Bennett, M., Belsky, D. W., Karlsson Linnér, R., Ahlskog, R., Kleinman, A. et al. (2021), 'Resource profile and user guide of the polygenic index repository', *Nature human behaviour* **5**(12), 1744–1758.

Behrman, J. R. & Rosenzweig, M. R. (2002), 'Does increasing women's schooling raise the schooling of the next generation?', *American economic review* **92**(1), 323–334.

Brand, J. E. & Davis, D. (2011), 'The impact of college education on fertility: Evidence for heterogeneous effects', *Demography* **48**(3), 863–887.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R. & Allen, N. E. (2017), 'Comparison of sociodemographic and health-related characteristics of uk biobank participants with those of the general population', *American journal of epidemiology* **186**(9), 1026–1034.

Harris, K. M. & McDade, T. W. (2018), 'The biosocial approach to human development, behavior, and health across the life course', *RSF: The Russell Sage Foundation Journal of the Social Sciences* **4**(4), 2–26.

Martín-García, T. & Baizán, P. (2006), 'The impact of the type of education and of educational enrolment on first births', *European sociological review* **22**(3), 259–275.

Mencarini, L. & Tanturri, M. L. (2006), 'High fertility or childlessness: Micro-level determinants of reproductive behavior in italy', *Population* **61**(4), 389–415.

Mills, M. C. & Tropf, F. C. (2015), 'The biodemography of fertility: a review and future research frontiers', *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* **67**(1), 397–424.

Mills, M., Rindfuss, R. R., McDonald, P. & Te Velde, E. (2011), 'Why do people postpone parenthood? reasons and social policy incentives', *Human reproduction update* **17**(6), 848–860.

Oppenheimer, V. K. (1994), 'Women's rising employment and the future of the family in industrial societies', *Population and development review* pp. 293–342.

Sobotka, T. (2004), *Postponement of childbearing and low fertility in Europe*, Dutch University Press Amsterdam.

Tropf, F. C. & Mandemakers, J. J. (2017), 'Is the association between education and fertility postponement causal? the role of family background factors', *Demography* **54**(1), 71–91.

van Alten, S., Domingue, B. W., Faul, J., Galama, T. J. & Marees, A. T. (2023), 'Should representativeness be avoided? reweighting the uk biobank corrects for pervasive selection bias due to volunteering', *Reweighting the UK Biobank Corrects for Pervasive Selection Bias Due to Volunteering (June 12, 2023). CESR-Schaeffer Working Paper* (2023_002).